บทที่ 2
วรรณกรรมและงานวิจัยที่เกี่ยวข้อง

การพัฒนาระบบคลังข้อมูลผู้ป่วยใน ผู้วิจัยได้ทำการศึกษาหลักการของทฤษฎีต่าง ๆ และเทคโนโลยีที่เกี่ยวข้องจาก ตำรา วารสาร บทความ และงานวิจัยที่เกี่ยวข้อง เพื่อเป็นแนวทางในการออกแบบและพัฒนาระบบ ซึ่งมีรายละเอียดดังต่อไปนี้

1. ระบบคลังข้อมูล (data warehouse) ดังนี้
   - 1.1 ฐานข้อมูล
   - 1.2 นิยามของคลังข้อมูล
   - 1.3 ประเภทของคลังข้อมูล
   - 1.4 วัตถุประสงค์ของการสร้างคลังข้อมูล
   - 1.5 ความสัมพันธ์ระหว่างระบบคลังข้อมูลกับระบบฐานข้อมูล
   - 1.6 ความแตกต่างระหว่างระบบคลังข้อมูลและระบบฐานข้อมูล
   - 1.7 ส่วนประกอบของคลังข้อมูล
   - 1.8 คุณลักษณะเฉพาะของคลังข้อมูล
   - 1.9 สถาปัตยกรรมของคลังข้อมูล (Data Warehouse Architecture--DWA)
   - 1.10 กระบวนการคลังข้อมูล (data warehousing)
   - 1.11 การวิเคราะห์ข้อมูลในคลังข้อมูล
   - 1.12 การดำเนินการวิเคราะห์ข้อมูลในคลังข้อมูล
2. การออกแบบและสร้างคลังข้อมูล ดังนี้
   - 2.1 สิ่งที่ควรพิจารณาก่อนสร้างคลังข้อมูล
   - 2.2 หลักเกณฑ์วิธีการออกแบบและพัฒนาคลังข้อมูล
   - 2.3 เทคนิคในการสร้างคลังข้อมูล
   - 2.4 กระบวนการโอนถ่ายข้อมูล (Extraction Transformation and Load--ETL)
   - 2.5 การดึงข้อมูลจากฐานข้อมูลประจำวัน

2.6 การส่งถ่ายข้อมูล Middle Ware

3. Metadata ดังนี้

3.1 Metadata หรือ Information Directory

3.2 ข้อมูลที่เก็บอยู่ใน Metadata

3.3 แนวทางการออกแบบ Metadata Repository

3.4 การเปลี่ยนแปลงแก้ไข Metadata

4. OLTP และ OLAP ดังนี้

4.1 Online Transaction Processing (OLTP)

4.2 Online Analytical Processing (OLAP)

5. การออกแบบโครงสร้างของ Dimensional Database ดังนี้

5.1 องค์ประกอบของ Dimensional Database

5.2 OLAP Cube

5.3 OLAP Storage

6. หลักการทำงานของคลังข้อมูล ดังนี้

6.1 การนำเอาคลังข้อมูลไปใช้งาน

6.2 ประโยชน์ของคลังข้อมูล

6.3 การออกแบบโครงสร้างในการจัดเก็บข้อมูลของคลังข้อมูล

6.4 การทำงานในการสร้างระบบคลังข้อมูล

6.5 การสร้างและการใช้งานระบบคลังข้อมูล

6.6 การวิเคราะห์และการออกรายงาน

7. ระบบสนับสนุนการตัดสินใจ ดังนี้

7.1 ประโยชน์ของระบบสนับสนุนการตัดสินใจ

8. ระบบสารสนเทศการบริหารโรงพยาบาล

9. งานวิจัยที่เกี่ยวข้อง ดังนี้

9.1 งานวิจัยในประเทศ

9.2 งานวิจัยในต่างประเทศ

ระบบคลังข้อมูล (data warehouse)

ฐานข้อมูล

ฐานข้อมูล (database) หมายถึง การเก็บรวบรวมข้อมูลที่มีความสัมพันธ์ไว้ด้วยกัน มีระบบในการจัดการฐานข้อมูลอย่างเป็นระบบ เพื่อการจัดเก็บและเรียกใช้งานข้อมูลที่ทันสมัยได้อย่างรวดเร็วและมีประสิทธิภาพ (สมจิตร อาจอินทร์ และงามนิจ อาจอินทร์ (พิสิษฐ์เจริญทัต), 2549, หน้า 11)

ฐานข้อมูล (database) หมายถึง ข้อมูลหรือแฟ้มข้อมูลจะเป็นแฟ้มข้อมูลที่ถูกจัดเก็บแบบกระจายไปตามหน่วยงานหรือแผนกต่าง ๆ ทั่วไป แต่ละแผนกต่างก็มีกระบวนการจัดเก็บแฟ้มข้อมูลเป็นของตนเอง โดยฐานข้อมูลจะเป็นแหล่งหรือศูนย์รวมของข้อมูลที่มีความสัมพันธ์กัน มีกระบวนการจัดหมวดหมู่ของข้อมูลที่มีแบบแผน ซึ่งก่อให้เกิดฐานข้อมูลที่เป็นแหล่งรวมของข้อมูลจากแผนกต่าง ๆ และจัดเก็บไว้อย่างเป็นระบบภายในฐานข้อมูลชุดเดียว ผู้ใช้งานต่าง ๆ ในแต่ละแผนกสามารถใช้ข้อมูลส่วนกลางนี้เพื่อนำไปประมวลผลร่วมกันได้และสนับสนุนการใช้ฐานข้อมูลร่วมกัน ทำให้ไม่เกิดความซ้ำซ้อนในข้อมูล เช่น ระบบแฟ้มข้อมูลและแนวคิดของฐานข้อมูลนั้นสามารถช่วยแก้ไขปัญหาต่าง ๆ ที่เกิดขึ้นจากการประมวลผลด้วยวิธีแฟ้มข้อมูลได้ (โอภาส เอี่ยมสิริวงศ์, 2548, หน้า 28)

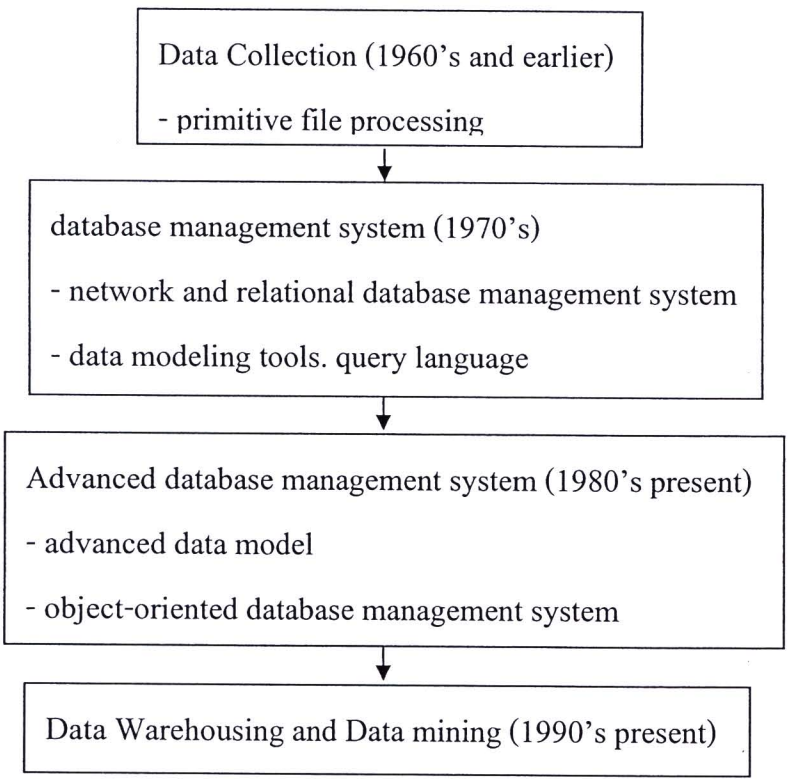
ฐานข้อมูล (database) หมายถึง ข้อมูลรวมถึงความสัมพันธ์ของข้อมูลที่จัดเก็บรวบรวมไว้เป็นกลุ่ม (วรรณวิภา ติตถะสิริ, 2545, หน้า 9)

ฐานข้อมูล (database) หมายถึง การร่วมใช้กลุ่มของข้อมูลที่มีความสัมพันธ์กันซึ่งได้ออกแบบ เพื่อให้กลุ่มข้อมูลนั้นสามารถให้สารสนเทศที่เพียงพอต่อความต้องการใช้งานสำหรับหน่วยงานต่าง ๆ ในองค์กรได้ (ทัดดาว ศีลคุณ, อัจฉรา ธารอุไร และภคินี อุปถัมภ์, 2541, หน้า 9)

ฐานข้อมูล (database) หมายถึง กลุ่มของข้อมูลที่เก็บรวบรวมไว้ มีความสัมพันธ์ซึ่งกันและกัน เพื่อลดความซ้ำซ้อนของข้อมูลและเก็บข้อมูลเหล่านี้ไว้ที่ศูนย์กลาง เพื่อนำข้อมูลที่ถูกเก็บไว้นำมาใช้ร่วมกัน (กระทรวงสาธารณสุข, ศูนย์เทคโนโลยีสารสนเทศ-และการสื่อสาร, ม.ป.ป.)

ระบบฐานข้อมูล เป็นการพัฒนาแฟ้มข้อมูลโดยการรวบรวมแฟ้มข้อมูลหลาย ๆ แฟ้มข้อมูลเข้าด้วยกัน มีการขจัดความซ้ำซ้อนของข้อมูลออกและเก็บแฟ้มข้อมูลไว้ที่ศูนย์กลาง เพื่อการใช้งานและควบคุมดูแลรักษาร่วมกันเมื่อต้องการใช้งาน และเป็นผู้มีสิทธิ์ที่จะใช้ข้อมูลเท่านั้นที่สามารถดึงข้อมูลที่ต้องการออกไปใช้ได้ ข้อมูลบางส่วนอาจใช้ร่วมกันกับผู้อื่นได้ แต่บางส่วนผู้มีสิทธิ์เท่านั้นจึงจะสามารถใช้ได้

วิวัฒนาการของเทคโนโลยีฐานข้อมูล จากอดีตถึงปัจจุบันอธิบายวิวัฒนาการดังนี้ (Silberschatz, Korth, & Sudarshan, 1996) ดังแสดงในภาพ 1



ภาพ 1 วิวัฒนาการของเทคโนโลยีฐานข้อมูล

ที่มา. จาก *Database System Concepts*, by A. Silberschatz, H. Korth, and S. Sudarshan, 1996, New York: McGraw-Hill.

ในปี ค.ศ. 1960 เทคโนโลยีฐานข้อมูลได้เริ่มพัฒนาจาก File Processing พื้นฐาน

ในปี ค.ศ. 1970 มีการนำไปสู่การพัฒนาระบบการเก็บข้อมูลในรูปแบบตาราง (relation database system) มีเครื่องมือจัดการ โมเดลข้อมูลและมีเทคนิคการใช้อินเด็กซ์ (index) และการบริหารข้อมูล นอกจากนี้ผู้ใช้ยังได้รับความสะดวกในการเข้าถึงข้อมูล โดยการใช้ภาษาในการเรียกข้อมูล (query language)

ในปี ค.ศ. 1980 เทคโนโลยีฐานข้อมูลได้เริ่มมีการปรับปรุงและการพัฒนาในระบบจัดการที่มีศักยภาพมากขึ้น ความก้าวหน้าในเทคโนโลยี Hardware ใน 30 ปีที่ผ่านมาได้นำไปสู่การจัดเก็บข้อมูลจำนวนมากที่มีความซับซ้อนได้อย่างมีประสิทธิภาพเพิ่มขึ้น

ในปี ค.ศ. 1990 ถึงปัจจุบัน สามารถจัดเก็บข้อมูลได้ในหลากหลายรูปแบบมากยิ่งขึ้น แตกต่างกันทั้งระบบปฏิบัติการ หรือการจัดเก็บฐานข้อมูลซึ่งการนำข้อมูลทั้งหมดมารวมและจัดเก็บไว้ในรูปแบบเดียวกัน เรียกว่า Data Warehouse

นิยามของคลังข้อมูล

ปัจจุบันข้อมูลที่เกิดจากการทำธุรกรรม (transaction) ต่าง ๆ มีจำนวนมากมาย ผู้บริหารที่มีวิสัยทัศน์ยาวไกลสามารถนำข้อมูลมาใช้ได้อย่างเหมาะสมและเป็นประโยชน์ต่อองค์กรเป็นอย่างมาก ในสภาพการแข่งขันที่มากขึ้นวิธีที่นิยมของผู้บริหารบางองค์กรในการนำข้อมูลจำนวนมากที่มีอยู่มาใช้ คือ การทำระบบคลังข้อมูล จากแนวคิดของระบบคลังข้อมูล (data warehouse) มีผู้ให้คำนิยาม ดังนี้

ระบบคลังข้อมูล (data warehouse) คือ เป็นที่เก็บรวบรวมข้อมูลสำคัญและจำเป็นจากแหล่งต่าง ๆ ซึ่งเป็นประโยชน์ต่อการตัดสินใจของผู้บริหาร เพื่อให้ผู้บริหารสามารถเรียกใช้ข้อมูลที่ต้องการได้อย่างรวดเร็วและมีประสิทธิภาพมากขึ้น ข้อมูลเชิงบริหารนี้จะสามารถช่วยลดปัญหาที่เกิดจากการฐานข้อมูลปฏิบัติการ (operational database) ซึ่งเป็นการเก็บข้อมูลในรูปแบบ Transaction System ได้ (กองบรรณาธิการไอทีซอฟต์, 2540, หน้า 120)

ระบบคลังข้อมูล (data warehouse) คือ ที่เก็บข้อมูลขององค์กรที่ได้ออกแบบเพื่อช่วยการตัดสินใจของฝ่ายบริหาร ในทางปฏิบัติสิ่งที่เก็บอยู่ในคลังไม่ได้มีแต่เพียงข้อมูลเท่านั้น หากยังเก็บเครื่องมือสำหรับดำเนินการกับข้อมูล กระบวนการทำงานกับข้อมูล

และทรัพยากรอื่น ๆ ระบบคลังข้อมูลเพื่อการบริหารได้ถูกออกแบบมาเพื่อใช้ในการเก็บข้อมูลขนาดใหญ่ โดยระบบข้อมูลเพื่อการบริหารนี้จะแยกข้อมูลออกจากฐานข้อมูลที่ใช้งานประจำวัน (operational database) ซึ่งข้อมูลสำหรับการบริหารโดยส่วนใหญ่จะเป็นข้อมูลสรุป (summary data) อาจจะเป็นข้อมูลในอดีต ข้อมูลอ้างอิงหรือข้อมูลปัจจุบัน ซึ่งอาจได้มาจากข้อมูล Operational Database หรือมีการประมวลผลข้อมูลใน Operational Database ให้เป็นข้อมูลสรุป หรืออาจนำมาจากที่อื่นภายนอกองค์กรและเพิ่มเติมลงไปก็ได้ ข้อมูลที่ถูกจัดเก็บอยู่ในคลังข้อมูลถือว่าเป็นข้อมูลในรูปแบบ Relational Database Management System (RDBMS) ที่มีประสิทธิภาพสูง ส่วนใหญ่เราจะเลือกเก็บเฉพาะข้อมูลที่จำเป็นสำหรับการตัดสินใจหรือหัวข้อของธุรกิจ (ทบวงมหาวิทยาลัย, โครงการ-เครือข่ายสารสนเทศเพื่อพัฒนาการศึกษา, 2544)

คลังข้อมูล (data warehouse) คือ ศูนย์รวมของข้อมูลที่จัดให้อยู่ในลักษณะที่สะดวกต่อการนำไปวิเคราะห์โดยง่าย โดยทำหน้าที่ในการรวบรวมข้อมูล คำนวณจัดเก็บตลอดจนสามารถบริหารและจัดการข้อมูลเพื่อให้ผู้บริหารสามารถรับข้อมูลที่ถูกต้องต่อการตัดสินใจในแนวทางของธุรกิจ (สมพร จิวรสกุล, 2545, หน้า 695)

คลังข้อมูลเป็นศูนย์รวมของข้อมูล ที่จัดให้อยู่ในลักษณะที่สะดวกต่อการนำไปวิเคราะห์โดยง่าย ซึ่งข้อมูลอาจมาจากหลาย ๆ ระบบหรือจากหลาย ๆ แหล่ง มักกำหนดให้อยู่ในลักษณะให้สามารถอ่านข้อมูลได้อย่างเดียว (บันทึกไม่ได้) เป็นระบบสนับสนุนข้อมูลที่ช่วยในการตัดสินใจ เก็บผลการคำนวณไว้ล่วงหน้า เพื่อช่วยให้สอบถามข้อมูลได้เร็วขึ้นและช่วยในการวิเคราะห์แนวทางของธุรกิจได้ตามระยะเวลาที่ผ่านมา (ทวีสิน ศิริวัฒนาวงศ์ชัย, 2547)

ระบบคลังข้อมูล (data warehouse) ตามคำจำกัดความที่ บิลล์ อินมอน บิดาแห่งคลังข้อมูลให้ไว้ หมายถึง แหล่งข้อมูลที่ใช้ในการตัดสินใจของระดับบริหาร โดยข้อมูลมีการเก็บในลักษณะที่รวบรวม (integrated) ถูกจัดเก็บไว้เป็นเรื่อง ๆ (subject oriented) แปรผันตามมิติของเวลา (time variant) และข้อมูลเหล่านี้จะไม่มีการเปลี่ยนแปลง (nonvolatile) (ภัทรชัย ลลิตโรจน์วงศ์, 2547, หน้า 21)

Inman (อ้างถึงใน พันทิพย์ คูอมรพัฒนะ, 2548) ได้นิยามไว้ว่า “Data Warehouse is a Subject-Oriented, Integrated, Time-Variant, Nonvolatile Collection of Data is

สำนักงานคณะกรรมการวิจัยแห่งชาติ
ห้องสมุด
วันที่ 24 ก.ค. 2550
เลขทะเบียน 247468
เลขเรียกหนังสือ

Support of Management's Decision Making Process" ระบบคลังข้อมูล หมายถึง ข้อมูลที่เป็นข้อมูลเฉพาะเรื่องที่สนใจในการนำไปวิเคราะห์ เช่น ข้อมูลการขาย ข้อมูลสินค้า และข้อมูลต่าง ๆ เหล่านี้ จะถูกรวบรวมมาจากระบบสารสนเทศ (management information system) ต่าง ๆ ขององค์กรมาไว้ที่เดียวกันเพื่อให้ข้อมูลลักษณะต่าง ๆ ถูกนำมาอยู่ในรูปแบบเดียวกัน เช่น ระบบสารสนเทศการผลิต ระบบสารสนเทศการบัญชี ระบบสารสนเทศการเงิน เป็นต้น และจากข้อมูลภายนอกองค์กร (external data) ข้อมูลเรื่องในคลังข้อมูลจะมีเวลาเข้ามาเกี่ยวข้องและการวิเคราะห์ข้อมูลส่วนใหญ่จะเป็นการเปรียบเทียบข้อมูลในช่วงเวลาที่แตกต่างกันและข้อมูลที่ถูกจัดเก็บในคลังข้อมูลจะต้องไม่มีการเปลี่ยนแปลงหรือแก้ไข

ระบบคลังข้อมูล (data warehouse) คือ การรวบรวมข้อมูลจากฐานข้อมูลของส่วนปฏิบัติงาน (operational database) หลาย ๆ รูปแบบหรืออาจจะมาจากแหล่งข้อมูลที่สำคัญและจำเป็นอื่น ๆ มาทำการเปลี่ยนแปลงหรือสรุปให้อยู่ในรูปของฐานข้อมูลที่มีรูปแบบเหมาะสมต่อการใช้ในการวิเคราะห์ การเก็บรวบรวมและการนำกลับมาใช้ ทำให้ได้เป็นแหล่งรวมข้อมูลที่อยู่ในความสนใจของผู้ใช้ เพื่อใช้ประกอบในการตัดสินใจใช้เป็นข้อมูลทางธุรกิจ การวางแผน หรือเป็นข้อมูลสำหรับผู้บริหาร ซึ่งจะช่วยให้สามารถทำการตัดสินใจได้อย่างถูกต้องโดยง่ายและรองรับข้อมูลจำนวนมากได้ เป็นการนำเสนอแนวทางในการเข้าถึงข้อมูลในองค์กรได้อย่างมีประสิทธิภาพ (กสิภพ ประศาสน์ศรีสุภาพ และเกียรติ ชุมพล สุทธิศิริกุล, 2552)

ระบบคลังข้อมูล (data warehouse) คือ ฐานข้อมูลขนาดใหญ่ขององค์กรหรือหน่วยงานหนึ่ง ๆ ซึ่งเก็บรวบรวมข้อมูลจากฐานข้อมูลระบบงานประจำวันหรือเรียกอีกอย่างว่า Operational Database และฐานข้อมูลอื่นภายนอกองค์กร หรือเรียกว่า External Database โดยข้อมูลที่ถูกจัดเก็บในคลังข้อมูลนั้น มีวัตถุประสงค์ในการนำมาใช้งานและมีลักษณะของการจัดเก็บแตกต่างไปจากข้อมูลในฐานข้อมูลระบบงานอื่น โดยข้อมูลในคลังข้อมูลจะถูกนำมาใช้เพื่อสนับสนุนการตัดสินใจบริหารงานของผู้บริหาร โดยเฉพาะการเป็นข้อมูลพื้นฐานให้กับระบบงานเพื่อการบริหารงานอื่น เช่น ระบบสนับสนุนการตัดสินใจ "Decision Support" (DSS) เป็นต้น (Mattison, 1996)

ระบบคลังข้อมูล (data warehouse) คือ ระบบฐานข้อมูลที่มีข้อมูล เพื่อใช้สำหรับองค์กรที่มีขนาดใหญ่ ซึ่งประกอบไปด้วยข้อมูลย้อนหลังหลาย ๆ ปี จนถึงข้อมูลปัจจุบัน นำมาเพื่อใช้ช่วยในการวิเคราะห์และตัดสินใจในงานองค์กร และสนับสนุนการใช้งานสำหรับผู้ใช้งานในหลายระดับ ขั้นตอนในการทำคลังข้อมูลนี้จะเริ่มตั้งแต่การเก็บข้อมูล ตรวจสอบ จัดเก็บ วิเคราะห์และแสดงออกมาเป็นรายงานต่าง ๆ โดยที่ข้อมูลเองไม่ได้มาจากแหล่งเดียว คือ อาจจะเก็บไว้ในหลาย ๆ ฐานข้อมูลที่แตกต่างกันและอาจจะต่างระบบปฏิบัติการกันได้ ลักษณะของข้อมูลในคลังข้อมูลจะเหมาะสำหรับงานวิเคราะห์ (Online Analytical Processing--OLAP) ที่มีการใช้ข้อมูลและเวลานานมาก ๆ เช่น การทำรายงานประจำเดือน เป็นต้น มากกว่างานที่เป็นการจัดการข้อมูลรายวัน (Online Transaction Process--OLTP) ที่มีการเพิ่ม ลบ และแก้ไขข้อมูลเกิดขึ้นพร้อม ๆ กันตลอดเวลา (Singh, 1996)

ระบบคลังข้อมูล (data warehouse) หมายถึง เป็นการเก็บข้อมูลในเชิงหัวข้อ (subject-oriented) ซึ่งทำการรวบรวมข้อมูลมาจากแหล่งต่าง ๆ (integrated) โดยเก็บข้อมูลเป็นระยะเวลานาน (time-variant) และไม่มีการเปลี่ยนแปลงของชุดข้อมูลได้โดยง่าย (non-volatile) ซึ่งเป็นข้อมูลสำหรับที่ใช้ช่วยการสนับสนุนในการดำเนินการตัดสินใจเพื่อการบริหาร (Roiger & Geatz, 2003, p. 184)

Subject-Oriented คือ ข้อมูลจะถูกสร้างและถูกเก็บในเชิงหัวข้อโดยจะเก็บเฉพาะข้อมูลที่จำเป็นต่อกระบวนการตัดสินใจ

Intergrated คือ เป็นการรวบรวมข้อมูลจากแหล่งต่าง ๆ ทั้งภายในและภายนอกที่เกี่ยวกับเซตข้อมูลที่เราต้องการให้มาอยู่ที่เดียวกัน (ในฐานข้อมูลเดียวกัน) พร้อมทั้งทำให้ข้อมูลที่มาจากต่างแหล่งกันนั้นมีความสอดคล้อง

Time Variant คือ ในการเก็บข้อมูลในคลังข้อมูลนั้นจะใช้ระยะเก็บข้อมูลนานเพื่อช่วยการวิเคราะห์ข้อมูล

Non-Volatile คือ ข้อมูลที่อยู่ในคลังข้อมูลจะไม่มีการเปลี่ยนแปลงหรือแก้ไขง่าย ๆ ผู้ใช้สามารถโหลดหรือเข้าถึงข้อมูลได้เท่านั้น (สุนีย์ พงษ์พินิจภิญโญ, ม.ป.ป.)

ประเภทของคลังข้อมูล

ประเภทของคลังข้อมูล แบ่งได้ดังนี้ (บริษัท เดอะเพาเวอร์ สเตชั่น จำกัด, ม.ป.ป.)

1. Enterprise Data Warehouse เป็นคลังข้อมูลที่ทำหน้าที่ในการวิเคราะห์และตัดสินใจในระดับองค์กร

2. DOS (Operation Data Store) เป็นคลังข้อมูลที่มีขอบเขตกว้างกว่าเรื่องที่สนใจ แต่ไม่ครอบคลุมในระดับทั้งองค์กรเหมือนกับใน Enterprise Data Warehouse และจะมีการปรับปรุงข้อมูลอยู่เสมอใกล้เคียงกับการทำงานแบบ Real Time

3. Data Mart มีลักษณะเป็นส่วนหนึ่งของคลังข้อมูล ออกแบบมาเพื่อวิเคราะห์เรื่องที่สนใจโดยใน Mart ข้อมูลที่เป็นอิสระจะสามารถรวบรวมข้อมูลได้โดยตรงจากแหล่งข้อมูลต่าง ๆ

วัตถุประสงค์ของการสร้างคลังข้อมูล

เป้าหมายของการสร้างคลังข้อมูล คือ การแยกกลุ่มข้อมูลสารสนเทศที่ใช้ในการวิเคราะห์ทางธุรกิจออกจากฐานข้อมูลที่ใช้งานประจำวัน (operation database) มาเก็บอยู่ใน Relation Database Management Systems (RDBMS) ประสิทธิภาพสูงและทำให้การเรียกใช้ข้อมูลชุดนี้ทำได่อย่างยืดหยุ่น จากเครื่องมือที่อยู่บนเครื่องเดสก์ทอปทั่วไป โดยลด Off-Loading เพิ่มกลไกการช่วยตัดสินใจ ปรับปรุงเวลาที่ตอบสนอง (response time) รวดเร็วยิ่งขึ้นอย่างมากและผู้บริหารสามารถเรียกข้อมูลรายละเอียดที่จำเป็นที่ถูกเก็บมาก่อนหน้านี้ (historical data) มาใช้ช่วยในการตัดสินใจทางธุรกิจแม่นยำยิ่งขึ้น (ทบวงมหาวิทยาลัย, โครงการเครือข่ายสารสนเทศเพื่อพัฒนาการศึกษา, 2544)

เป้าหมายของการสร้างคลังข้อมูลมีดังนี้

1. คลังข้อมูลทำให้สามารถเข้าถึงข้อมูลขององค์กร ผู้ใช้งานและนักวิเคราะห์ขององค์กรสามารถเชื่อมต่อเข้าไปยังข้อมูลจากเครื่องคอมพิวเตอร์ของตนได้ ซึ่งการเชื่อมต่อสามารถทำได้ทันทีตามความต้องการและด้วยประสิทธิภาพสูง เครื่องมือที่มีให้ผู้ใช้งานและนักวิเคราะห์ใช้งานง่าย สามารถออกรายงานได้ด้วยการคลิกปุ่มเดียว

2. ข้อมูลในคลังข้อมูลมีความถูกต้องตรงกันหมด คำถามเดียวกันต้องได้รับคำตอบที่เหมือนกันเสมอ ไม่ว่าผู้ถามจะเป็นใคร ถามเวลาใด

3. ข้อมูลในคลังข้อมูลสามารถถูกวิเคราะห์จากหัวข้อในธุรกิจประเภทนั้น โดยแบ่งข้อมูลหรือรวบรวมข้อมูลมาวิเคราะห์ตามความต้องการ

4. คลังข้อมูลเป็นส่วนที่ผลิตข้อมูลจากข้อมูลรายวัน (Online Transaction Process--OLTP) ข้อมูลไม่เพียงแต่ถูกรวบรวมมาไว้ที่ศูนย์กลางอย่างเดียวแต่ข้อมูลจะถูกรวบรวมอย่างระมัดระวังจากแหล่งข้อมูลหลาย ๆ แหล่งนอกองค์กรด้วยแล้วมาปรับปรุงให้เหมาะสมกับการใช้งานเท่านั้น ถ้าให้ข้อมูลเชื่อถือไม่ได้หรือไม่สมบูรณ์จะไม่ถูกอนุญาตให้นำไปใช้

5. คุณภาพของข้อมูลในคลังข้อมูลเป็นตัวผลักดันให้สามารถทำการปรับโครงสร้างใหม่ (reengineering) ธุรกิจได้

ความสัมพันธ์ระหว่างระบบคลังข้อมูลกับระบบฐานข้อมูลปฏิบัติการ

ข้อมูลจาก ทบวงมหาวิทยาลัย, โครงการเครือข่ายสารสนเทศเพื่อพัฒนาการศึกษา (2544) พบว่า ในปัจจุบันมีการใช้ฐานข้อมูลอย่างกว้างขวางในระบบงานทั่วไป จึงมีการวิจัยและพัฒนาวิธีเก็บข้อมูลจำนวนมาก รวมถึงการค้นหาและนำข้อมูลที่ต้องการออกมาจากระบบฐานข้อมูลด้วย แต่เนื่องจากระบบฐานข้อมูลทั่วไป (operational database) ที่นิยมใช้อยู่ในปัจจุบันมีหลักในการเก็บข้อมูลที่เน้นในเรื่องการลดความซ้ำซ้อน (redundancy) รักษาความถูกต้อง (integrity) ลดการสูญหายของข้อมูล (information lost) และลดความผิดพลาดที่เกิดขึ้นจากการแก้ไขข้อมูล (update anomalies)

เนื่องจากฐานข้อมูลทั่วไป (operational database) มีลักษณะดังได้กล่าวมาแล้ว จึงมีความสามารถเพียงแค่การเรียกใช้ข้อมูลที่มีอยู่ แต่ไม่สามารถจะนำมาช่วยในการสนับสนุนการตัดสินใจได้ เพราะเมื่อมีการเรียกใช้ข้อมูลจะต้องเรียกใช้ข้อมูลจากฐานข้อมูลขนาดใหญ่ ซึ่งมีข้อมูลจำนวนมหาศาลและมีการแตกตารางที่นอร์มัลไลซ์ (normalized table) แล้วออกเป็นหลายตาราง จึงไม่รองรับคำถามที่ต้องการจะนำมาใช้ช่วยในการสนับสนุนการตัดสินใจ (decision support queries) มีการรวม (join) กันของตารางต่าง ๆ ที่ซับซ้อน ซึ่งจะทำให้มีประสิทธิภาพของการค้นหาข้อมูลจากฐานข้อมูล

น้อยลงและทำงานช้าลง ไม่สามารถเรียกใช้ข้อมูลที่ต้องการได้ทั้งหมดเพราะมีรูทีนอัตโนมัติ (automate routine) จึงสามารถค้นหาข้อมูลแบบที่ไม่ซ้ำซ้อนเท่านั้น นอกจากนี้การเก็บข้อมูลในระบบฐานข้อมูลทั่วไป (operational database) ยังไม่มีการเก็บข้อมูลย้อนหลัง (historical data) เพื่อใช้ช่วยในการคาดคะเนแนวโน้มที่คาดว่าจะเป็นไปได้ในอนาคต

ความแตกต่างระหว่างคลังข้อมูลและฐานข้อมูลปฏิบัติการ

ระบบคลังข้อมูล เรียกได้ว่าเป็นระบบฐานข้อมูลที่ผ่านการวิเคราะห์แล้ว ผลลัพธ์ที่ได้จะถูกนำมาช่วยในการตัดสินใจได้ ในขณะที่ระบบปฏิบัติงานทั่วไปเป็นเพียงระบบฐานข้อมูล ที่ออกแบบมาเพื่อช่วยอำนวยความสะดวกในงานประจำวันเท่านั้น แตกต่างโดยสิ้นเชิงกับระบบคลังข้อมูล ซึ่งจะนำข้อมูลมาผ่านการวิเคราะห์ก่อน ซึ่งออกแบบมาให้ข้อมูลเหล่านั้น ช่วยให้ผู้ใช้ในการตัดสินใจได้ทันทีไม่ว่าจะโดยการนำข้อมูลมาวิเคราะห์ สรุปและเสนอเป็นรายงาน ซึ่งความแตกต่างระหว่างฐานข้อมูลปฏิบัติการกับคลังข้อมูล (Singh, 1996) ดังข้อมูลที่ปรากฏในตาราง 1

ตาราง 1

ความแตกต่างระหว่างคลังข้อมูลและฐานข้อมูลปฏิบัติการ

การใช้งาน	ฐานข้อมูลปฏิบัติการ	คลังข้อมูล
ลักษณะการจัดการข้อมูล	Application (application oriented)	ตามหัวข้อเรื่องที่ต้องการ (subject oriented)
โครงสร้างข้อมูล	ซับซ้อนแล้วแต่เครื่องมือและการคำนวณ แต่เป็นรูปแบบที่ชัดเจนประมวลผลซ้ำเรื่องเดิม	มีโครงสร้างไม่แน่นอน ประมวลผลแบบวิเคราะห์แต่ง่ายเหมาะกับองค์กร
เนื้อหาและช่วงเวลา	ปัจจุบัน	อดีตและปัจจุบัน
การปรับปรุงข้อมูล	เป็นเรื่อง ๆ ไป มีจำนวนน้อยและทำเป็นประจำ	แล้วแต่สถานการณ์และความต้องการ ไม่มีการปรับปรุงข้อมูลโดยตรง
การเคลื่อนไหวของข้อมูล	ตลอดเวลา	คงที่จนกว่าจะปรับปรุงใหม่

ตาราง 1 (ต่อ)

การใช้งาน	ฐานข้อมูลปฏิบัติการ	คลังข้อมูล
ความแน่นอนการใช้ข้อมูล	แน่นอน	ไม่แน่นอน
แหล่งข้อมูล	ภายในองค์กร	ภายในและภายนอกองค์กร
ขนาดของข้อมูล	กิกะไบต์	กิกะไบต์ถึงเทราไบต์

ที่มา. จาก *Data Warehouse: Concept Technology, Implementations and Management*, by H. Singh, 1996, Upper Saddle River, NJ: Prentice-Hall.

ส่วนประกอบของคลังข้อมูล

ระบบ Data Warehouse ไม่ใช่ระบบสำเร็จรูปที่ซื้อมาแกะกล่องก็ใช้งานได้ทันที จำเป็นต้องมีการติดตั้ง Project ขึ้นเพื่อทำการ Customize เพื่อให้ใช้งานได้ ซึ่งการ Customize ที่ว่าก็ต้องการขั้นตอนจำนวนหนึ่งรวมถึงผลิตภัณฑ์อื่น ๆ มาช่วยด้วย ดังนี้ (เอื้อน ปิ่นเงิน, 2541)

1. เครื่องมือพัฒนาเพื่อช่วยออกแบบฐานข้อมูลของ Data Warehouse และโปรแกรมที่จะทำหน้าที่เก็บรวบรวม ข้อมูลจากระบบปฏิบัติงานหรือแหล่งข้อมูลอื่น ๆ

2. ส่วนที่ทำหน้าที่เป็น Directory ของข้อมูลเพื่ออำนวยความสะดวกแก่ผู้ดูแลระบบหรือผู้ใช้ทั่วไป ให้เข้าใจถึงข้อมูลแต่ละตัวและความหมาย

3. ตัวฐานข้อมูลของ Data Warehouse เอง

4. ส่วนที่ทำหน้าที่ Data Acquisition ซึ่งก็คือ ตัวที่ทำหน้าที่ดักจับเก็บรวบรวมข้อมูล รักษาความถูกต้อง โอนย้ายหรือแปลงข้อมูลจากแหล่งข้อมูลอื่นให้อยู่ในรูปแบบที่เหมาะสมก่อนจะเก็บเข้าสู่ Data Warehouse ต่อไป

5. ส่วนที่ทำหน้าที่ Data Management สำหรับจัดการและควบคุมการปฏิบัติงานของ Data Warehouse

6. ส่วนที่ทำหน้าที่เข้าถึงข้อมูล จะเป็นส่วนที่ให้บริการแก่ผู้ใช้ที่มีพื้นมาทางธุรกิจให้สามารถใช้เป็นเครื่องมือช่วยการตัดสินใจได้ ซึ่งแน่นอนว่าต้องการเครื่องมือตัวนี้เพื่อช่วยเข้าถึงและวิเคราะห์ข้อมูล

7. ส่วนที่ทำหน้าที่โอนย้ายข้อมูลเป็นส่วนที่ทำหน้าที่ Export ข้อมูลภายใน Data Warehouse ออกสู่ภายนอก อาจจะเพื่อจะโอนย้ายไปสู่ระบบ Data Warehouse ระบบ Data Mart หรือระบบอื่น ๆ ทั่วไป

ระบบคลังข้อมูล จะต้องมีการทำ Customize เพื่อให้ใช้งานได้ โดยมีขั้นตอนและผลิตภัณฑ์อื่น ๆ ดังนี้ (กิตติพงษ์ กลมกล่อม, 2552, หน้า 7)

1. เครื่องมือพัฒนาเพื่อช่วยออกแบบฐานข้อมูลของ Data Warehouse และโปรแกรมที่จะทำหน้าที่เก็บรวบรวมข้อมูลจากระบบปฏิบัติงาน หรือแหล่งข้อมูลอื่น ๆ

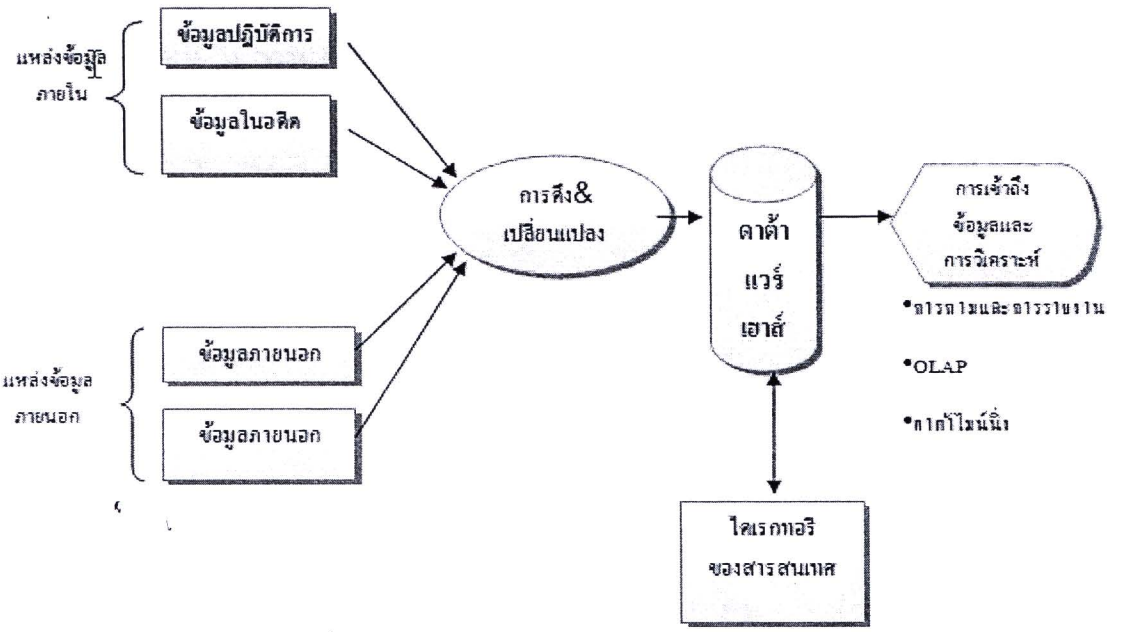
2. ส่วนที่ทำหน้าที่เป็น Directory ของข้อมูลเพื่ออำนวยความสะดวกแก่ผู้ดูแลระบบหรือผู้ใช้ทั่วไปให้เข้าใจถึงข้อมูลแต่ละตัวและความหมาย

3. ตัวฐานข้อมูลของ Data Warehouse

4. ส่วนที่ทำหน้าที่ Data Acquisition คือ ตัวที่ทำหน้าที่ดักจับเก็บรวบรวมข้อมูล รักษาความถูกต้องโอนย้ายหรือแปลงข้อมูลจากแหล่งข้อมูลอื่นให้อยู่ในรูปแบบที่เหมาะสมก่อนจะเก็บเข้าสู่ Data Warehouse ต่อไป

5. ส่วนที่ทำหน้าที่ Data Management สำหรับจัดการและควบคุมการปฏิบัติงานของ Data Warehouse

6. ส่วนที่ทำหน้าที่เข้าถึงข้อมูล จะเป็นส่วนที่ให้บริการแก่ผู้ใช้ที่มีพื้นมาทางธุรกิจให้สามารถใช้เป็นเครื่องมือช่วยการตัดสินใจ เพื่อช่วยในการเข้าถึงและวิเคราะห์ข้อมูล ส่วนประกอบของระบบคลังข้อมูล ดังแสดงในภาพ 2



ภาพ 2 ส่วนประกอบของ Data Warehouse

ที่มา. จาก *การออกแบบและพัฒนาคลังข้อมูล (data warehouse)* (หน้า 7), โดย กิตติพงษ์ กลมกล่อม, 2552, กรุงเทพมหานคร: โรงพิมพ์เคทีพี คอมพ์ แอนด์ คอนซัลท์.

คุณลักษณะเฉพาะของคลังข้อมูล

คุณลักษณะของคลังข้อมูล จากนิยามของคลังข้อมูลที่กล่าวถึงความแตกต่างกันระหว่างคลังข้อมูลกับฐานข้อมูลปฏิบัติการ ซึ่งสามารถสรุปคุณลักษณะเฉพาะของคลังข้อมูล ได้ดังนี้ (คมกริช ศิริแสงชัยกุล, 2542, หน้า 37)

1. Subject Oriented หรือการแบ่งโครงสร้างตามเนื้อหา คือ คลังข้อมูลถูกออกแบบมาเพื่อมุ่งเน้นไปในแต่ละเนื้อหาที่สนใจ ไม่ได้เน้นไปที่การทำงานหรือกระบวนการแต่ละอย่างโดยเฉพาะเหมือนอย่างฐานข้อมูลปฏิบัติการ ในส่วนของรายละเอียดข้อมูลที่จัดเก็บในระบบทั้งสองแบบ ก็จะแตกต่างกันไปตามความต้องการใช้งานด้วยเช่นกัน คลังข้อมูลจะไม่เก็บข้อมูลที่ไม่มีส่วนเกี่ยวข้องกับการประมวลผล เพื่อสนับสนุนการตัดสินใจในขณะที่ข้อมูลนั้นจะถูกเก็บไว้ในฐานข้อมูลปฏิบัติการ หากมีส่วนที่เกี่ยวข้องกับกระบวนการทำงาน

2. Integration หรือการรวมเป็นหนึ่ง ซึ่งถือได้ว่าเป็นคุณลักษณะที่สำคัญที่สุดของคลังข้อมูล คือ การรวบรวมข้อมูลจากหลายฐานข้อมูลปฏิบัติการเข้าด้วยกันและทำ

ให้ข้อมูลมีมาตรฐานเดียวกัน เช่น กำหนดให้มีค่าตัวแปรของข้อมูลในเนื้อหาเดียวกันให้เป็นแบบเดียวกันทั้งหมด เป็นต้น

3. Time Variance หรือความสัมพันธ์กับเวลา หมายถึง ข้อมูลในคลังข้อมูลจะต้องจัดเก็บโดยกำหนดช่วงเวลาเอาไว้ โดยจะสัมพันธ์กับการดำเนินธุรกิจของหน่วยธุรกิจนั้นเพราะในการตัดสินใจด้านการบริหาร จำเป็นต้องมีข้อมูลเปรียบเทียบในแต่ละช่วงเวลา แต่ละจุดของข้อมูลจะเกี่ยวข้องกับจุดของเวลา และข้อมูลแต่ละจุดสามารถเปรียบเทียบกันได้ตามแกนของเวลา

4. Nonvolatile หรือความเสถียรของข้อมูล หมายถึง ข้อมูลในคลังข้อมูลจะไม่มีการเปลี่ยนแปลงบ่อย ไม่ว่าจะเป็นการเพิ่มเติมข้อมูลใหม่หรือการปรับปรุงแก้ไขข้อมูลเดิมที่บรรจุอยู่แล้ว ผู้ใช้ทำได้เพียงการเข้าถึงข้อมูลเท่านั้น

สถาปัตยกรรมของคลังข้อมูล (Data Warehouse Architecture--DWA)

สถาปัตยกรรมของคลังข้อมูล เป็นโครงสร้างมาตรฐานที่ใช้อธิบายเพื่อให้เข้าใจแนวคิดและกระบวนการของคลังข้อมูลนั้น ๆ ซึ่งโดยทั่วไปแล้วคลังข้อมูลแต่ละระบบอาจจะมีรูปแบบที่เหมือนกันได้ เพื่อให้เหมาะสมกับองค์กรนั้น ๆ ทั้งนี้ส่วนประกอบต่าง ๆ ภายในสถาปัตยกรรมคลังข้อมูลที่สำคัญ ได้แก่ (เลิศ เลิศศิริโสภณ, 2541, หน้า 95)

1. Operational Database หรือ External Database Layer ทำหน้าที่จัดการกับข้อมูลในระบบงานปฏิบัติการหรือแหล่งข้อมูลภายนอกองค์กร

2. Information Access Layer เป็นส่วนที่ผู้ใช้ปลายทางติดต่อผ่านโดยตรง ประกอบด้วย Hardware และ Software ที่ใช้ในการแสดงผลเพื่อการวิเคราะห์โดยมีเครื่องมือช่วยเป็นตัวกลางที่ผู้ใช้ใช้ติดต่อกับคลังข้อมูล โดยในปัจจุบันเครื่องมือที่ได้รับความนิยมเพิ่มขึ้นอย่างรวดเร็ว คือ Online Analycal Processing Tool หรือ OLAP Tool ซึ่งเป็นเครื่องมือที่มีความสามารถในการวิเคราะห์ที่ซับซ้อน และแสดงข้อมูลในรูปแบบหลายมิติ

3. Data Access Layer เป็นส่วนต่อประสานระหว่าง Information Access Layer กับ Operational Layer

4. Data Director (metadata) Layer เพื่อให้เข้าถึงข้อมูลได้ง่ายขึ้นและเป็นการเพิ่มความเร็วในการเรียกและดึงข้อมูลของคลังข้อมูล

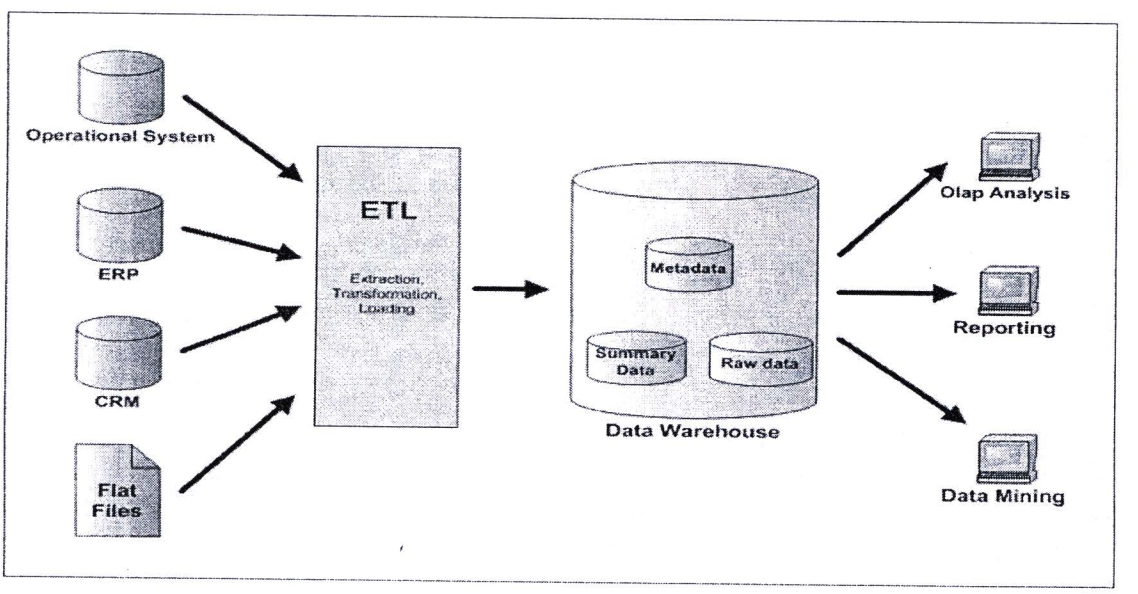
5. Process Management Layer ทำหน้าที่จัดการกระบวนการทำงานทั้งหมด

6. Application Messaging Layer เป็นมิดเดิลแวร์ทำหน้าที่ในการส่งข้อมูลภายในองค์กรผ่านทางเครือข่าย

7. Data Warehouse (physical) Layer เป็นแหล่งเก็บข้อมูลของทั้ง Information Data และ External Data ในรูปแบบที่ง่ายแก่การเข้าถึงและยืดหยุ่นได้

8. Data Staging Layer เป็นกระบวนการแก้ไขและดึงข้อมูลจาก External Database

สิริวัฒน์ ธนุรเวท (2543, หน้า 19) นำเสนอแนวคิดเกี่ยวกับสถาปัตยกรรมของคลังข้อมูลดังนี้ การเก็บข้อมูลได้จากแหล่งที่ต่าง ๆ ไม่ได้มาจากแหล่งเดียว อาจมีการเก็บไว้ในหลาย Database หรือต่างระบบปฏิบัติการ การตรวจสอบจะมีการตรวจสอบชนิดประเภทของข้อมูลว่าตรงกันหรือไม่ ข้อมูลที่มีอยู่ถูกต้องหรือไม่ การจัดเก็บข้อมูลทำการจัดเก็บลงคลังข้อมูลที่สร้างขึ้น การวิเคราะห์และแสดงผล ดังแสดงในภาพ 3



ภาพ 3 สถาปัตยกรรมของคลังข้อมูล

ที่มา. จาก *Data Warehouse*, โดย สิริวัฒน์ ธนุรเวท, 2543, ค้นเมื่อ 20 มกราคม 2549, จาก http://www.kku.ac.th/data warehouse.ppt

กระบวนการคลังข้อมูล (data warehousing)

คลังข้อมูลเป็นศูนย์รวมของหลักการและวิธีการที่หลากหลาย เช่น การออกแบบและสร้างโครงสร้างของข้อมูลใน Data Warehouse วิธีการเพื่อให้ได้มาซึ่งข้อมูล วิธีการสร้างผลลัพธ์จากข้อมูลที่มีนั้น รวมไปถึงวิธีดูแลรักษาและวิธีการปรับปรุงประสิทธิภาพเพื่อการจัดการข้อมูลและการใช้งานคลังข้อมูล ทำให้เกิดแนวคิด กระบวนการและวิธีการหลาย ๆ อย่างขึ้นในคลังข้อมูล อาทิเช่น (กิตติพงษ์ กลมกล่อม, 2552, หน้า 8-9)

1. กลุ่มกระบวนการจัดเก็บข้อมูล (data storing) กระบวนการจัดเก็บข้อมูลหมายถึง กระบวนการต่าง ๆ ที่ดำเนินการเพื่อให้การจัดเก็บข้อมูลเป็นไปอย่างถูกต้อง สามารถบอกเล่าถึงธุรกิจขององค์กรได้ถูกต้องแม่นยำ และมีประสิทธิภาพ กระบวนการเหล่านี้ ได้แก่ การสร้างแบบจำลองข้อมูล สร้างฐานข้อมูล การจัดแบ่งพื้นที่บนฐานข้อมูล (database partitioning) การปรับแต่งฐานข้อมูล (database tuning) เป็นต้น

กระบวนการในกลุ่มนี้สำคัญเป็นอย่างยิ่ง คือ การสร้างแบบจำลองข้อมูล (data modeling) เนื่องจากแบบจำลองข้อมูลจะถูกนำมาใช้กำหนดว่าข้อมูลใน Data Warehouse Database มีรายละเอียดอย่างไร และมีความสัมพันธ์กันอย่างไร และมีเรื่องราวทางธุรกิจอย่างไร ด้วยเหตุนี้จึงถือว่าแบบจำลองข้อมูล (data model) เปรียบได้กับหัวใจของคลังข้อมูล

2. กลุ่มกระบวนการเคลื่อนย้ายข้อมูล (data propagation) เมื่อได้ศึกษาต่อไป พบว่าคลังข้อมูลมีองค์ประกอบต่าง ๆ มากมาย และมีหลายองค์ประกอบของระบบทำหน้าที่จัดเก็บข้อมูล กระบวนการ Data Propagation เป็นกระบวนการเพื่อการเคลื่อนย้ายข้อมูลจากองค์ประกอบหนึ่งไปยังอีกองค์ประกอบหนึ่ง เทคนิคที่ใช้ในคลังข้อมูล เรียกว่า “กระบวนการ Extract Transform Load (ETL)”

3. กระบวนการเตรียมการเพื่อการใช้ข้อมูล (data provisioning) คือ กระบวนการเพื่อการออกแบบโครงสร้างข้อมูลใน Data Mart และการใช้ข้อมูลจาก Data Mart ในแบบต่าง ๆ

4. กระบวนการบริหารจัดการข้อมูล (housekeeping) ข้อมูลในคลังข้อมูลมักเป็นข้อมูลที่จัดเก็บไว้เป็นระยะเวลายาวนาน และไม่ถูกนำออกหรือลบทิ้งจากคลังข้อมูล (non-volatile data) แต่ในแง่ของการจัดการทรัพยากรระบบนั้น ฐานข้อมูลที่ใช้เพื่อจัดเก็บข้อมูลมีจำนวนและขนาดที่จำกัด ดังนั้น จึงจำเป็นต้องคาดการณ์การเติบโตของข้อมูล

เพื่อจัดเตรียมพื้นที่ให้เพียงพอต่อการใช้งานตามช่วงเวลาที่เหมาะสม (เรียกขั้นตอนการประมาณขนาดว่า “data sizing”) รวมถึงการจัดการกับข้อมูลที่ล้าสมัย ให้จัดเก็บอยู่ใน Data Warehouse Database ได้ต่อไป แต่มีขนาดเล็กลง (เรียกกระบวนการนี้ว่าการทำ “data packing”) และกระบวนการอื่น ๆ ที่จำเป็นต้องดำเนินการเพื่อให้คลังข้อมูลสามารถทำงานต่อไปได้ด้วยความราบรื่น

5. กระบวนการเพื่อการควบคุมการทำงานของระบบ (controlling) คลังข้อมูลก็เหมือนกับระบบทั่ว ๆ ไป ที่จำเป็นต้องมีการควบคุมการทำงาน ดังนั้น เราจึงใช้ข้อมูลประเภทพิเศษที่เรียกว่า “Metadata” เพื่อเก็บข้อมูลเกี่ยวกับคุณสมบัติของข้อมูลและของระบบ ซึ่งจะถูกใช้เพื่อการควบคุมการทำงานต่าง ๆ ในคลังข้อมูล

Metadata เป็นแนวคิดใหม่ที่มีใช้เป็นครั้งแรกในระบบคลังข้อมูล มีรายละเอียดปลีกย่อยมากมาย และมีความเฉพาะตัวแตกต่างกันไปในแต่ละคลังข้อมูล

การวิเคราะห์ข้อมูลในคลังข้อมูล

กิตติพงษ์ กลมกล่อม (2548, หน้า 11) อธิบายว่า โดยทั่วไปแล้วจะมีการดำเนินการที่นิยมทำอยู่ 3 รูปแบบ คือ รายงานและสอบถาม (report and query) วิเคราะห์แบบหลายมิติ (multidimensional data analysis) และการประมวลผลในเชิงวิเคราะห์แบบออนไลน์ (Online Analytic Processing--OLAP)

ข้อแตกต่างระหว่างรายงานและสอบถามกับ (Online Analytic Processing--OLAP) คือ รายงานและสอบถามจะทำการประมวลผลจากฐานข้อมูลทุกครั้งที่การเรียกใช้แต่ (Online Analytic Processing--OLAP) จะทำการประมวลผลไว้ก่อนในช่วงเวลาที่ข้อมูลไม่ได้ถูกใช้งาน หรือข้อมูลที่มีการสรุปไว้แล้วใน Cube เมื่อเรียกดูข้อมูลจะมีประสิทธิภาพและรวดเร็วกว่า

รายงานและสอบถามจะเป็นการใช้โปรแกรมหรือระบบที่เรียกว่า “ระบบสร้างรายงาน (report generatior)” เพื่อใช้ข้อมูลที่เกิดจากการปฏิบัติงานหรือในคลังข้อมูลมาประมวลผลเพื่อประโยชน์ในการตัดสินใจ โดยรายงานที่ได้ไม่มีความซับซ้อนมากนัก กล่าวคือ เป็นวิธีการที่ง่ายที่สุดในการวิเคราะห์ข้อมูลในคลังข้อมูล

วิเคราะห์แบบหลายมิติ (multidimensional data analysis) หรือการประมวลผลในเชิงวิเคราะห์แบบออนไลน์ (Online Analytic Processing--OLAP) ซึ่งต่อไปเราจะเรียกว่า OLAP ซึ่งเป็นวิธีการสอบถามที่สนับสนุนการวิเคราะห์ข้อมูลในสภาพแวดล้อมหลายมิติที่ประกอบไปด้วยมิติ (dimension) และตารางข้อเท็จจริง (fact table) ซึ่งตารางข้อเท็จจริง คือ ข้อเท็จจริงที่เราต้องการวิเคราะห์ เช่น ยอดขาย ผลกำไร จำนวนสินค้าที่มีอยู่ในปัจจุบัน เป็นต้น ในตารางข้อเท็จจริงจะประกอบไปด้วย Measure เป็นข้อมูลที่ต้องการวัดทั้งในเชิงปริมาณ (quantitative) และเชิงคุณภาพ (qualitative) กับข้อมูลที่ใช้เชื่อมความสัมพันธ์กับมิติ (dimension) โดยที่ข้อมูลจะถูกเก็บอยู่ในรูปของลูกบาศก์ที่มีหลายมิติ ซึ่งเรียกว่า Cube

การใช้ระบบ OLAP ช่วยตอบสนองความต้องการของผู้ใช้ได้รวดเร็วขึ้น สามารถใช้ได้กับปัญหาทางธุรกิจที่เกิดขึ้นจริง ช่วยในการตัดสินใจของผู้บริหารในองค์กร และยังใช้ทรัพยากรมนุษย์ได้อย่างมีประสิทธิภาพ จึงทำให้เพิ่มรายได้และผลกำไรมากขึ้น โดยลักษณะเด่นของ OLAP มีลักษณะดังนี้ (Youness, 2005, pp. 10-12)

1. สามารถมองข้อมูลได้หลายวิธี (multidimensional views of data หรือ data cubes) โดยปกติแล้วตัวแบบทางธุรกิจสามารถมองข้อมูลได้หลายมิติ ตัวอย่างเช่น การขาย เราสามารถมองยอดขายในมิติของเวลา สินค้า สถานที่ ลูกค้า เป็นต้น โดยมิติของเวลาเราอาจจะดูข้อมูลเป็นรายเดือน รายไตรมาส รายปี ในมิติสถานที่ สามารถดูข้อมูลตามลักษณะภูมิศาสตร์ มิติของสินค้าสามารถดูข้อมูลตามกลุ่มของสินค้า ในการมองมุมมองข้อมูหลายมิตินั้นจะอ้างถึง Data Cube ซึ่ง Cube ในตัวแบบทางธุรกิจ อาจจะมีมิติได้มากกว่า 3 มิติ

2. สามารถที่จะคำนวณข้อมูลได้มากขึ้น (calculation-intensive) เนื่องจาก OLAP จะทำการสรุปข้อมูลตามลำดับชั้น ทำให้สามารถทำการคำนวณที่ซับซ้อนได้มากขึ้น เช่น เปอร์เซ็นต์ของจำนวนรวมทั้งหมด ค่าเฉลี่ยแบบเคลื่อนที่ เปอร์เซ็นต์การเจริญเติบโตซึ่ง OLAP ถูกออกแบบมาให้รองรับการคำนวณที่ซับซ้อน ซึ่งจะเป็นผลประโยชน์ต่อการตัดสินใจมากขึ้น โดยปกติแล้วระบบ Online Transaction Processing (OLTP) เป็นระบบที่ใช้ในการรวบรวมและจัดการข้อมูลแต่ละระบบ OLAP จะเป็นระบบที่ใช้สร้างสารสนเทศจากข้อมูลที่ถูกรวบรวมซึ่งจะทำให้เกิดองค์ความรู้ใหม่

3. สามารถคำนวณข้อมูลตามเวลาได้ (time-intelligence) กล่าวคือ เวลาเป็นมิติที่นิยมใช้มากในระบบ OLAP มิติของเวลานั้นสามารถใช้ในการเปรียบเทียบประสิทธิภาพของการดำเนินการทางธุรกิจ ตัวอย่างเช่น ในการพิจารณาการปฏิบัติงานของเดือนปัจจุบันกับการปฏิบัติงานของเดือนที่ผ่านมา หรือการเปรียบเทียบผลกำไรขององค์กรในไตรมาสล่าสุดกับไตรมาสเดียวกันในปีที่ผ่านมา เป็นต้น

การดำเนินการกับ OLAP ในการวิเคราะห์ข้อมูล

การดำเนินการกับ OLAP ในการวิเคราะห์ข้อมูล มีดังนี้ (กิตติพงษ์ กลมกล่อม, 2548, หน้า 110)

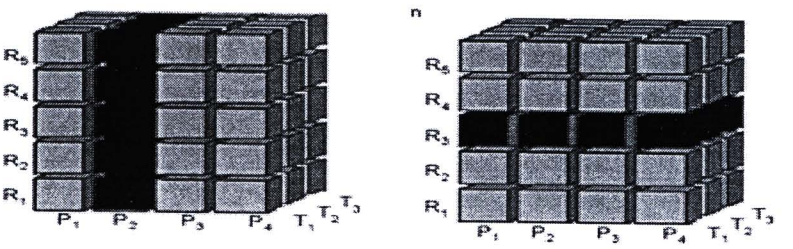
1. Roll Up และ Drill Down เนื่องจาก OLAP จะมีการสรุปข้อมูลตามมิติต่าง ๆ ที่เราสนใจและข้อมูลตามมิตินั้น จะมีการเก็บอยู่ในลักษณะของลำดับชั้น ดังนั้น Roll Up และ Drill Down คือ การเปลี่ยนแปลงระดับความละเอียดของลำดับชั้นของมิติในการพิจารณาข้อมูล โดยการดำเนินการนี้จะใช้กับ Snowflakes Schema เป็นส่วนใหญ่

โดยการ Roll Up จะเป็นการเปลี่ยนแปลงระดับความละเอียดจากระดับที่เล็กไปสู่ระดับที่ใหญ่กว่า และ Drill Down จะเป็นการเปลี่ยนแปลงในทางตรงกันข้าม กล่าวคือ จะเป็นการเปลี่ยนแปลงระดับความละเอียดจากระดับที่ใหญ่ไปสู่ระดับที่เล็กกว่า ตัวอย่างการทำ Roll Up และ Drill Down พิจารณาตารางข้อเท็จจริงของการขายแยกตามภูมิภาค (region) ร้านค้า (shop) ประเภทสินค้า (product type) และสินค้า (product)

2. Slice เป็นการพิจารณาเฉพาะบางส่วนของ Cube ที่เราสนใจ เพราะในบางครั้งถ้า Cube มีขนาดใหญ่มาก การพิจารณาข้อมูลทั้งหมดอาจจะทำให้ยากต่อการพิจารณา การวิเคราะห์โดยการแบ่ง Cube ออกมาเฉพาะบางส่วนจะช่วยให้ง่ายต่อการพิจารณา

3. Dice เป็นการพลิกแกนมุมมองในการดูข้อมูลให้ตรงตามความต้องการใช้ได้กับ Snowflakes Dimension Data Model และ Star Schema Dimension Data Model

การวิเคราะห์ข้อมูลให้อยู่ในรูปแบบหลายมิติ (multidimensional model) ซึ่งจะทำให้รายงานสามารถเจาะลึก (drill down) พลิกแพลง (slice dice) ได้ (กาญจนา หินเธาว์, 2552) ดังแสดงในภาพ 4



ภาพ 4 การวิเคราะห์ข้อมูลให้อยู่ในรูปแบบหลายมิติ

ที่มา. จาก *Business Intelligent*, โดย กาญจนา หินเธาว์, 2552, ค้นเมื่อ 15 ธันวาคม 2550, จาก http://www.google.co.th/search?hl=th&q=Business+Intelligent+%E0%B8%84%E0% B8% E0% B8%AD&meta=lr%3Dlang_th

การออกแบบและสร้างคลังข้อมูล

สิ่งที่ควรพิจารณาก่อนสร้างคลังข้อมูล

เนื่องจากการลงทุนสร้างคลังข้อมูลขึ้นมาใช้เพื่อสนับสนุนการทำงานขององค์กรนั้น จำเป็นต้องมีค่าใช้จ่ายในการลงทุนมหาศาลทั้งที่สามารถวัดออกมาเป็นตัวเงินได้ เช่น ค่าใช้จ่ายด้าน Hardware ด้าน Software และ Infrastructure อื่น ๆ ที่จำเป็นต้องใช้ ส่วนค่าใช้จ่ายที่ไม่เป็นตัวเงินแต่มีความสำคัญอย่างมาก ได้แก่ กำลังแรงงานที่เสียไปของทรัพยากรบุคคลขององค์กรและเวลาที่ใช้ในการพัฒนา ดังนั้น เมื่อองค์กรตัดสินใจสร้างคลังข้อมูลขึ้นแล้วควรจะประสบความสำเร็จด้วย ทั้งนี้ Poe ได้เสนอ The Big Eight หรือ 8 ประการที่ควรให้ความสนใจ โดยมีรายละเอียดดังนี้ (Poe, 1996, p. 5)

1. ควรมีเป้าหมายที่ชัดเจนร่วมของการสร้างระบบของทุกคนในองค์กร เหมือนการตอบคำถามว่าทำไมคุณถึงคิดจะสร้างคลังข้อมูล ซึ่งคำตอบขององค์กรที่จะได้ คือ เป้าหมายที่ต้องการ โดยควรจะเขียนเป้าหมายนี้ออกมาเป็นลายลักษณ์อักษรที่ชัดเจน เพื่อให้ทีมพัฒนาได้เข้าใจเป้าหมายร่วมกัน

2. ทำความเข้าใจสถาปัตยกรรมของระบบ เพื่อให้ทีมพัฒนาเข้าใจตรงกันในที่นี้หมายถึง Blueprint ที่แสดง E-R Model รวมของระบบความเข้าใจที่ตรงกันทำให้งานดำเนินไปได้เร็วขึ้น

3. เทคโนโลยีที่ใช้ควรอยู่ในวิสัยที่เหมาะสม ทั้งด้านของตัวเงินและความยากง่ายในการเรียนรู้ ทั้งนี้หมายรวมทั้งด้าน Hardware ด้าน Software และเครือข่ายอาจต้องมีการทดสอบและฝึกอบรมก่อนการใช้งานจริง

4. ทีมงานต้องมีวิสัยทัศน์เชิงบวกในการทำงาน เนื่องจากทีมพัฒนามักมาจากส่วนงานด้านเทคโนโลยีสารสนเทศแต่ในเนื้องานจริง ๆ แล้ว ผู้ใช้ขั้นปลายเป็นส่วนงานอื่น ๆ ขององค์กร ดังนั้น จึงจำเป็นอย่างยิ่งที่จะให้ผู้ใช้ขั้นปลายที่เป็นเจ้าของงานเข้ามาร่วมทำงานด้วยตั้งแต่ต้นโครงการ

5. ต้องมั่นใจว่าทีมพัฒนาเข้าใจเป็นอย่างดี ถึงความแตกต่างกันระหว่างฐานข้อมูลปฏิบัติการและฐานข้อมูลสนับสนุนการตัดสินใจ

6. จัดให้มีการฝึกอบรมโดยควรเป็นการฝึกอบรมก่อนเริ่มโครงการ โดยเฉพาะอย่างยิ่งการฝึกอบรมเกี่ยวกับเครื่องมือที่องค์กรจะใช้พัฒนา ทั้งนี้อาจเป็นการฝึกอบรมจากบริษัทผู้ขาย

7. ควรหาบุคลากรที่มีประสบการณ์ในการพัฒนาคลังข้อมูล เพื่อทำหน้าที่เป็นผู้จัดการโครงการหรือถ้าในองค์กรไม่เคยมีประสบการณ์เลยอาจจ้างที่ปรึกษาที่มีความเชี่ยวชาญและมีประสบการณ์ด้านนี้โดยเฉพาะมาช่วยทีมพัฒนา

8. โปรแกรมที่จะใช้นำเสนอข้อมูลในคลังข้อมูล ต้องสามารถเรียนรู้ได้ง่ายและผู้ใช้สามารถใช้งานได้อย่างมีประสิทธิภาพ

หลักเกณฑ์วิธีการออกแบบและพัฒนาคลังข้อมูล

วิธีการนี้ถูกเสนอโดย Kimball ในปี ค.ศ. 1996 เรียกว่า ระเบียบวิธี 9 ขั้นตอน หรือ Nine-Step Methodology (Connolly & Begg, 2002, pp. 3-5) โดยวิธีการนี้ เริ่มจากการออกแบบจากส่วนย่อยที่แสดงถึงแต่ละระบบงานขององค์กร หรือเรียกอีกอย่างหนึ่งว่าดาต้ามาร์ท (datamart) โดยเมื่อออกแบบแต่ละส่วนสำเร็จแล้ว จึงนำมารวมกันเป็นคลังข้อมูลขององค์กรในขั้นสุดท้าย ซึ่งขั้นตอนทั้ง 9 ขั้นตอน มีรายละเอียดดังนี้

1. กำหนด Datamart คือ การเลือกว่าจะสร้าง Datamart ของระบบงานใดบ้าง และระบบใดเป็นระบบงานแรก โดยองค์กรต้องสร้าง E-R Model ที่รวมระบบงานทุกระบบขององค์กรไว้ แสดงการเชื่อมโยงของแต่ละระบบงานอย่างชัดเจน และสิ่งที่

ต้องคำนึงถึงในการเลือกระบบงานที่จะเป็น Datamart นั้น มีปัจจัยที่เกี่ยวข้อง ได้แก่ จะต้องสามารถพัฒนาออกมาได้ทันตามเวลาที่ต้องการ โดยอยู่ในงบประมาณที่กำหนดไว้และต้องตอบปัญหาทางธุรกิจให้แก่องค์กรได้ ดังนั้น Datamart แรก ควรจะเป็นของระบบงานที่นำรายได้เข้ามาสู่องค์กรได้ เช่น ระบบงานขาย เป็นต้น

2. กำหนด Fact Table ของ Datamart คือ การกำหนดเนื้อหาหลักที่ควรจะเป็นของ Datamart โดยการเลือก Entity หลักและกระบวนการที่เกี่ยวกับ Entity นั้น ๆ ออกมาจาก E-R Model ขององค์กรนั่น หมายถึงว่า จะทำให้เราทราบถึง Dimension Table ที่ควรจะมีด้วย

3. กำหนด Attribute ที่จำเป็นในแต่ละ Dimension Table คือ การกำหนด Attribute ที่บอกหรืออธิบายรายละเอียดของ Dimension ได้ ทั้งนี้ Attribute ที่เป็น Primary Key ควรเป็นค่าที่คำนวณได้ กรณีที่มี Datamart มากกว่าหนึ่ง Datamart มี Dimension เหมือนกันนั่นหมายถึง Attribute ใน Dimension นั้นต้องเหมือนกันทุกประการ แต่ก็ไม่อาจจะแก้ไขปัญหาการจัดเก็บข้อมูลซ้ำซ้อน อันจะนำมาสู่ความแตกต่างกันของข้อมูลชุดเดียวกัน ปัญหานี้จึงเป็นการดีที่จะมีการใช้ Dimension Table ร่วมกันในแต่ละ Fact Table ที่จำเป็นต้องมี Dimension ดังกล่าว โดยเรียก Dimension Table ลักษณะแบบนี้ว่า Conformed และเรียกว่า Fact Constellation เราสามารถกำหนดข้อดีของการใช้ Dimension Table ร่วมกันได้ ดังนี้

3.1 แน่ใจได้ว่าในแต่ละรายงานจะออกมาสอดคล้องกัน

3.2 สามารถสร้าง Datamart ในเวลาต่าง ๆ กันได้

3.3 สามารถเข้าถึง Datamart โดยผู้พัฒนากลุ่มอื่น ๆ

3.4 สามารถรวบรวม Datamart หลาย ๆ อันเข้าด้วยกัน

3.5 สามารถออกแบบคลังข้อมูลร่วมกันได้

4. กำหนด Attribute ที่จำเป็นใน Fact Table โดย Attribute หลักใน Fact Table จะมาจาก Primary Key ในแต่ละ Dimension Table นอกจากนี้แล้ว ยังสามารถมี Attribute ที่จำเป็นอื่น ๆ ประกอบอยู่ด้วย เช่น Attribute ที่ได้จากการคำนวณค่าเบื้องต้นที่จำเป็นสำหรับการคงอยู่ของ Attribute อื่นใน Fact Table เรียกอีกอย่างหนึ่งว่า Measure การกำหนด Attribute นี้ไม่ควรจะเลือก Attribute ที่คำนวณค่าไม่ได้ เช่น เป็นตัวหนังสือ

หรือไม่ใช่ตัวเลข เป็นต้น และไม่ควรเลือก Attribute ที่ไม่เกี่ยวข้องกับเนื้อหาของ Fact Table ที่เราสนใจด้วย

5. จัดเก็บค่าการคำนวณเบื้องต้นใน Fact Table คือ การจัดเก็บค่าที่ได้จากการคำนวณให้เป็น Attribute หนึ่งใน Fact Table ถึงแม้ว่าจะสามารถหาค่าได้จาก Attribute อื่น ๆ ก็ตาม ทั้งนี้เพื่อให้การสอบถามมีประสิทธิภาพมากขึ้น สามารถทำงานด้วยความเร็วที่เพิ่มขึ้น เนื่องจากไม่ต้องคำนวณค่าใหม่ทั้งหมด ถึงแม้ว่าจะเกิดความซ้ำซ้อนของข้อมูลในการจัดเก็บบ้างก็ตาม

6. เขียนคำอธิบายของ Dimension Table ทั้งนี้ก็เพื่อให้ผู้ใช้สามารถใช้งาน Datamart ได้อย่างมีประสิทธิภาพ เพราะเกิดความเข้าใจอย่างดีในส่วนต่าง ๆ

7. กำหนดระยะเวลาในการจัดเก็บข้อมูลในฐานข้อมูล โดยอาจจะเป็นการจัดเก็บเพียงช่วงระยะเวลา 1-2 ปี หรือนานกว่านั้น ขึ้นอยู่กับความต้องการขององค์กร เนื่องจากองค์กรแต่ละประเภทมีความต้องการในการจัดเก็บข้อมูลต่างช่วงเวลากัน ทั้งนี้ขึ้นอยู่กับความจำเป็นหรือข้อกำหนดในการดำเนินธุรกิจมีข้อสังเกตอยู่ 2 ประการที่น่าสนใจ และสำคัญสำหรับการออกแบบ Attribute ในเรื่องของการจัดเก็บข้อมูล ดังนี้

7.1 ข้อมูลที่ถูกจัดเก็บไว้นานเกินไปมักเกิดปัญหาการอ่าน หรือแปลข้อมูลนั้น ๆ จากแฟ้มหรือเทปเก่า

7.2 เมื่อมีการนำรูปแบบเก่าของ Dimension Table มาใช้อาจทำให้เกิดการเปลี่ยนแปลงของ Dimension อย่างช้า ๆ ได้

8. การติดตามปัญหาการเปลี่ยนแปลงของ Dimension อย่างช้า ๆ คือ การเปลี่ยนเอา Attribute ของ Dimension Table เก่ามาใช้ส่งผลกระทบต่อข้อมูลปัจจุบันของ Dimension Table โดยสามารถแบ่งประเภทของปัญหาที่เกิดได้เป็น 3 ประเภท ดังนี้

8.1 เกิดการเขียนทับข้อมูลใหม่โดยข้อมูลเก่า

8.2 เกิด Record ใหม่ ๆ ขึ้นใน Dimension

8.3 เกิด Record มีทั้งค่าเก่าและใหม่ปนกันไป

9. กำหนด Query เป็นการออกแบบด้านกายภาพเพื่อให้ผู้ใช้เกิดความสะดวกในการใช้งานและสามารถทำงานได้อย่างมีประสิทธิภาพ

เมื่อดำเนินการทั้ง 9 ขั้นตอนสำหรับแต่ละ Datamart เสร็จแล้ว จึงนำทั้งหมดมารวมกันเป็นภาพของคลังข้อมูลขององค์กร

เทคนิคในการสร้างคลังข้อมูล

ข้อมูลที่จะจัดเก็บภายในคลังข้อมูลมีการเคลื่อนที่ของข้อมูล (information flow) 5 ประเภท ดังนี้ (เศรษฐพงค์ มะลิสุวรรณ, ม.ป.ป.)

1. Inflow คือ การนำข้อมูลจากฐานข้อมูลอื่นเข้าสู่คลังข้อมูลทั้งฐานข้อมูลภายในและภายนอกองค์กร โดยในขั้นนี้อาจมีการเปลี่ยนแปลงโครงสร้างข้อมูล การทำ Denormalize การลบหรือเพิ่มฟิลด์เพื่อให้ข้อมูลทั้งหมดอยู่ในเนื้อหาที่สนใจเดียวกัน ในขั้นตอนนี้อาจใช้เครื่องมือที่เรียกว่า Data Warehouse Tool

2. Upflow เมื่อข้อมูลที่เราต้องการอยู่ในคลังข้อมูลแล้ว บางครั้งอาจต้องมีการเพิ่มคุณค่าให้กับข้อมูลด้วย เพื่อให้ข้อมูลอยู่ในรูปแบบที่เป็นประโยชน์มากที่สุดต่อการนำเครื่องมือมาใช้ ได้แก่ การจัดกลุ่มข้อมูลหาค่าทางสถิติที่ซับซ้อน จัดข้อมูลให้อยู่ในรูปแบบหรือ Template มาตรฐาน

3. Downflow เป็นขั้นตอนของการปรับปรุงเปลี่ยนแปลงข้อมูลที่เก่าและไม่อยู่ในเนื้อหาที่องค์กรสนใจออกไปจากคลังข้อมูลขององค์กร

4. Outflow เป็นขั้นที่ผู้ใช้เรียกใช้ข้อมูลในคลังข้อมูลผ่านเครื่องมือต่าง ๆ โดยการเรียกใช้อาจมีเพียงขอเรียกเป็นครั้งคราว เป็นประจำทุกวัน ทุกเดือน หรือต้องการแบบทันที

5. Metaflow ข้อมูลที่จัดเก็บในคลังข้อมูล จะถูกทำข้อมูลไว้อีกชุดหนึ่งเป็นแหล่งที่มาของข้อมูลนั้น หรือแม้กระทั่งที่อยู่ของข้อมูลนั้นในคลังข้อมูลและข้อมูลอื่นที่เกี่ยวข้อง

กระบวนการโอนถ่ายข้อมูล (Extraction Transformation and Load--ETL)

การโอนถ่ายข้อมูลมีความสำคัญต่อคลังข้อมูล เนื่องจากคลังข้อมูลจะทำการรวบรวมข้อมูลจากแหล่งข้อมูล (data source) ทั้งภายในและภายนอกองค์กร ดังนั้น จึงต้องมีวิธีการในการรวบรวมข้อมูลทั้งหลายเหล่านั้น ซึ่งจะมีการถ่ายโอนข้อมูลจากแหล่งข้อมูล

ต่าง ๆ สู่คลังข้อมูล และการโอนถ่ายข้อมูลจากคลังข้อมูลสู่ Data Marts การโอนถ่ายข้อมูลจะมี 2 ส่วน คือ (กิตติพงษ์ กลมกล่อม, 2548, หน้า 118, 144-145)

1. Data Acquisition เป็นส่วนที่รับและเตรียมข้อมูลก่อนที่จะนำข้อมูลเข้าสู่คลังข้อมูล ไม่ว่าจะเป็นข้อมูลมาจากแหล่งข้อมูลใด หน้าที่หลักของ Data Acquisition คือ

1.1 ตรวจสอบข้อมูลเบื้องต้นพร้อมทั้งแก้ไขข้อมูลที่ผิดพลาดให้มีความถูกต้อง เช่น ข้อมูลที่เป็นตัวเลขต้องมีค่าเป็นตัวเลข เป็นต้น

1.2 ทำการกรองข้อมูลที่ไม่ต้องการ หรือมีความผิดพลาดที่อาจจะทำให้คลังข้อมูลเกิดความผิดพลาดในขณะทำงาน เช่น ไวรัสคอมพิวเตอร์ เป็นต้น

1.3 การรักษาความปลอดภัยในการส่งข้อมูล เช่น การตรวจสอบสิทธิ์ผู้ส่งข้อมูลแจ้งเตือนข้อมูลที่ผิดพลาด

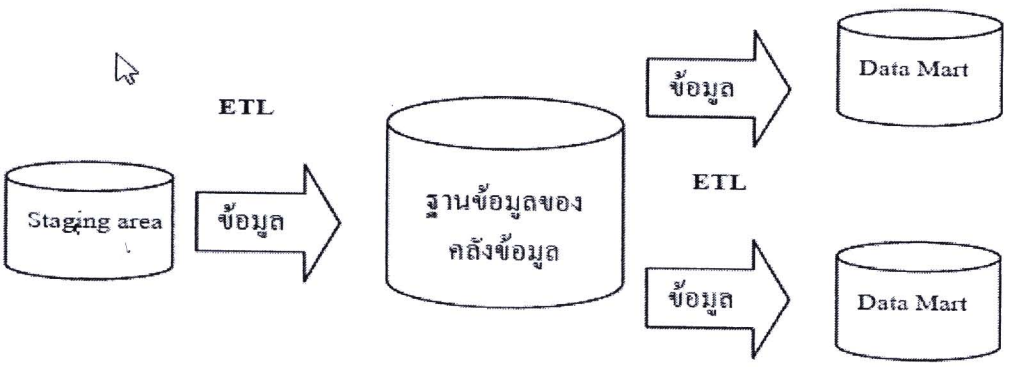
2. Data Staging Area เป็นส่วนแปลงข้อมูลที่ได้รับจาก Data acquisition ให้อยู่ในรูปแบบที่คลังข้อมูลกำหนด หรือตามตัวแบบคลังข้อมูล อีกทั้งยังมีกระบวนการต่าง ๆ ที่อยู่ในส่วนของ Data Staging Area เช่น การเป็นที่พักข้อมูลและสำรองข้อมูลชั่วคราว (temporary backup) ในระหว่างการโอนถ่ายข้อมูลก่อนที่จะนำข้อมูลเข้าสู่คลังข้อมูล กระบวนการทำความสะอาดข้อมูล (data cleansing) และการตรวจสอบข้อมูล ซึ่งการตรวจสอบข้อมูลนั้นจะแตกต่างกับส่วนของ Data Acquisition เนื่องจาก Data Staging Area สามารถติดต่อกับส่วนของคลังข้อมูล ดังนั้น การตรวจสอบข้อมูลในส่วนของ Data Staging Area เป็นการตรวจสอบความถูกต้องและสามารถตรวจสอบความสอดคล้องของข้อมูลที่มีอยู่ในคลังข้อมูลเดิม เช่น การตรวจสอบรหัสสินค้าที่สอดคล้องหรือมีอยู่จริงในคลังข้อมูล เป็นต้น

กระบวนการที่จะนำข้อมูลจากส่วนหนึ่งไปยังอีกส่วนหนึ่ง เรียกว่า Extraction Transformation and Load (ETL) ซึ่งประกอบด้วย 3 กระบวนการ คือ

1. Extract คือ กระบวนการของการดึงข้อมูลออกจากแหล่งข้อมูล

2. Transform คือ กระบวนการแปลงข้อมูลจากโครงสร้างเดิมจากแหล่งข้อมูล (source) ให้อยู่ในรูปแบบโครงสร้างข้อมูลปลายทาง (destination)

3. Load คือ กระบวนการนำข้อมูลที่แปลงรูปแบบแล้ว เข้าสู่แหล่งข้อมูลปลายทาง กระบวนการโอนถ่ายข้อมูล (Extraction Transformation and Load--ETL) ดังแสดงใน ภาพ 5



ภาพ 5 การโอนถ่ายข้อมูล (Extraction Transformation and Load--ETL)

ที่มา. จาก *การออกแบบและพัฒนาระบบคลังข้อมูล*, โดย กิตติพงษ์ กลมกล่อม, 2548, กรุงเทพมหานคร: โรงพิมพ์เคทีพี คอมพ์ แอนด์ คอนซัลท์.

ภาษาที่ใช้ในการจัดการข้อมูลในฐานข้อมูลนั้นอาจจะใช้ภาษา SQL (Structured Query Language) ช่วยในการจัดการข้อมูล โดยภาษา SQL เป็นภาษาทางด้านฐานข้อมูล แบ่งออกเป็น 3 กลุ่ม คือ

1. ภาษาที่ใช้สำหรับนิยามข้อมูล (Data Definition Language--DDL) จะเป็นกลุ่มของคำสั่งที่ใช้ในการกำหนดโครงสร้างต่าง ๆ ของฐานข้อมูล

2. ภาษาที่ใช้ในการจัดการข้อมูล (Data Manipulation Language--DML) จะเป็นกลุ่มคำสั่งที่ใช้ในการเรียกดู เปลี่ยนแปลง เพิ่มหรือลบข้อมูล

3. ภาษาที่ใช้ในการควบคุม (Data Control Language--DCL) เป็นกลุ่มคำสั่งที่ใช้ในการควบคุมข้อมูล เช่น การเรียกใช้ข้อมูลพร้อมกัน การกำหนดความปลอดภัยของข้อมูล เป็นต้น

การดึงข้อมูลจากฐานข้อมูลประจำวัน

การดึงข้อมูลเข้า Data Warehouse คือ ข้อมูลที่ดึงมาจากแหล่งข้อมูลทั้งหลายที่จำเป็นต้องได้รับการแปลงให้อยู่ในรูปแบบที่เหมาะสม ก่อนจะนำมารวมกันเป็นก้อนใน Data Warehouse การแปลงข้อมูลจะเป็นขั้นตอนที่ทำพร้อม ๆ กับการดึงข้อมูลไปด้วยซึ่งจะครอบคลุมไปถึงการจัดรวบรวมข้อมูลที่มีชนิดต่าง ๆ กัน การเปลี่ยนรหัสรวมไปถึงการคัดเลือกเฉพาะข้อมูลที่เราต้องการ นอกจากนั้นการแปลงข้อมูลยังรวมไปถึงการคำนวณที่จำเป็น การสรุปข้อมูลและการ Update ข้อมูลจากหลาย ๆ แหล่งให้ตรงกัน ในการนี้ก็จะมีเครื่องมือที่ได้รับการออกแบบมาโดยเฉพาะสำหรับการแปลงข้อมูล ดังนั้น เครื่องมือจะมีหน้าที่ดึงข้อมูลและแปลงข้อมูลจากหลาย ๆ แหล่ง แล้วทำการ Map ข้อมูลต้นฉบับไปสู่ปลายทางและสร้าง Data Definition Language (DDL) รวมไปถึงสร้างโปรแกรม ที่จะทำหน้าที่จัดการกับข้อมูลและแปลงข้อมูลก่อนจะโหลดไปเก็บไว้ในตำแหน่งที่เหมาะสมในฐานข้อมูลต่อไป เครื่องมือที่มีความสำคัญ คือ Data Scrubbing และ Normalization Tools มีหน้าที่ทำความสะอาดข้อมูลให้มี Consistency สูง เช่น ค้นหาและทำการ Merge ที่ซ้ำกันเข้าด้วยกัน จัดมาตรฐานให้กับคำย่อที่ใช้ให้ตรงกันทั้งหมดและปรับความยาวของ Field ของข้อมูลชนิดเดียวกันให้ตรงกันฐานข้อมูลที่ทำการ Normalization ไว้ได้ดี จะทำให้ระบบที่นำไปใช้จริงมีประสิทธิภาพสูงแต่ก็ทำให้เราต้อง Join ข้อมูลมากเช่นกัน ตอนเราทำการ Query นั้นเครื่องมือที่ใช้ในการแปลงข้อมูลจะมีหน้าที่ De-Normalization กลับ ในขั้นตอนการสร้างหรือรวบรวมข้อมูลสำหรับ Data Warehouse เครื่องมือรวบรวมและแปลงข้อมูลบางตัวสามารถตั้งให้ทำการ Update ข้อมูลตามตารางเวลาได้ การสร้าง Data Warehouse บางระบบก็อาจจะไม่จำเป็นต้องมีการแปลงข้อมูลเลยก็ได้ ถ้าเป็นกรณีที่ Data Warehouse นั้นมีโครงสร้างไม่ซับซ้อนในขณะที่ข้อมูลจากฐานข้อมูลต้นฉบับเป็นชนิดเดียวกันหมด การค้นหาและดึงข้อมูลทำได้ง่ายรวมไปถึงมีเครื่องมือ Front End ที่ช่วยให้เข้าถึงข้อมูลได้ (เอื้อน ปิ่นเงิน, 2541)

การส่งถ่ายข้อมูล Middleware

กิตติพงษ์ กลมกล่อม (2552, หน้า 125) อธิบายว่า การส่งถ่ายข้อมูล Middleware ซึ่งเป็นตัวกลางระหว่างระบบฐานข้อมูลประจำวันและระบบ Data Warehouse ทำหน้าที่

เป็นตัวกลางในการโอนย้ายข้อมูล Middleware จำเป็นต้องรองรับหลาย Protocol โดยมากเครื่องมือช่วยติดต่อเหล่านั้นจะอยู่ในรูปของ Gateway โดยเครื่องมือที่ช่วยติดต่อนี้จะเชื่อมแหล่งข้อมูลจากหลาย ๆ แหล่งด้วยกันและเข้าสู่ Data Warehouse สร้างเป็นระบบ 3 ชั้น ขึ้นเพื่อเชื่อมข้อมูลเข้าสู่ Server ทั้งหมดก่อนที่จะส่งเข้าไปที่เครื่อง Client ไปในคราวเดียวกัน Middleware จะอยู่ตรงกลางระหว่าง Data Warehouse จุดเข้าถึงฐานข้อมูลและระบบวิเคราะห์ข้อมูล สิ่งที่ต้องคำนึงถึงในการย้ายข้อมูลจากแหล่งข้อมูล มีหลายอย่างด้วยกัน เช่น อัตราการ Update วิธีการคัดลอกข้อมูลการ Maintenance และการปรับข้อมูลให้เข้ากับความต้องการทางธุรกิจ ข้อมูลเมื่อได้รับการอ่านขึ้นมาจากต้นฉบับแล้วก็จะถูกทำความสะอาดและปรับปรุงให้อยู่ในสภาพที่ดีขึ้น การทำความสะอาด คือ การรับโครงสร้างข้อมูลไม่ว่าจะเป็นการเติมค่าใน Field ที่หายไปตรวจสอบความถูกต้องของข้อมูล (consistency และ integrity check)

ส่วนการปรับสภาพข้อมูลให้ดีขึ้นนั้น คือ การเพิ่ม Attribute ที่บอกค่าของช่วงเวลาลงไป ทำการสรุปข้อมูลหรือคำนวณค่าที่จะเป็นผลลัพธ์ต่าง ๆ เมื่อข้อมูลได้รับการทำความสะอาดและปรับสภาพแล้วจึงจะถูก Map เข้ายังตำแหน่งใน Data Warehouse ต่อไป การนำข้อมูลดังกล่าวไปใช้ในฐานข้อมูลของ Data Warehouse อาจจะทำได้โดยผ่านทาง SQL ก็ได้

Metadata

Metadata หรือ Information Directory

เอื้อน ปิ่นเงิน (2541) อธิบายว่า การที่จะได้ประโยชน์จากระบบ Data Warehouse อย่างเต็มที่ ผู้ใช้จำเป็นต้องมีเครื่องอำนวยความสะดวกที่ช่วยพวกเขาหาข้อมูลเกี่ยวกับตัวข้อมูล หรือสิ่งอื่น ๆ ที่เกี่ยวข้องกับข้อมูลได้ เช่น Query ตัววิเคราะห์หรือรายงานที่มีและวิธีที่นำสิ่งเหล่านี้มาช่วยในการตัดสินใจวางแผนกลยุทธ์สิ่งอำนวยความสะดวก คือ "Information Directory" นั่นเอง จะทำหน้าที่เหมือนเป็นคลังสำหรับจัดการเก็บ Metadata หรือข้อมูลที่เกี่ยวกับเนื้อหาที่ Data Warehouse มีอยู่โดย Metadata จะช่วยให้ผู้ใช้รู้ว่าตัว Data Warehouse นั้นเก็บอะไรไว้ เข้าใจความหมายของข้อมูลแต่ละตัวและสามารถ

เรียกใช้ Query หรือตัววิเคราะห์เพื่อเข้าถึงข้อมูลภายใน Data Warehouseได้ถูกต้อง นอกจากผู้ใช้ปลายทางแล้วนักพัฒนาหรือผู้ดูแลระบบก็มีความจำเป็นต้องใช้ Metadata นี้ด้วยเช่นกัน โดยทั่วไปแล้วระบบ Information Directory ก็จะเตรียมเครื่องมือให้จำนวนหนึ่งเพื่อช่วยรวบรวมข้อมูลดูแลรักษาและดูข้อมูล Metadata ซึ่งเป็นตัวบอกถึงเนื้อหาภายใน Data Warehouse อีกที Metadata ที่ว่านี้อาจจะเป็น Metadata ทางเทคนิคหรือทางธุรกิจก็ได้ และอาจจะสร้างโดย Graphical Information Directory Interface ซึ่งเป็นเครื่องมือจัดการ Metadata ที่เป็นกราฟิกอินเตอร์เฟสหรืออาจจะสร้างโดยการ Inport เอา Metadata มาจากผลิตภัณฑ์ หรือ Data Warehouse ตัวอื่นก็ได้ ข้อมูล Metadata ทางเทคนิคก็จะเป็นข้อมูลเกี่ยวกับ Data Warehouse ระบบนั้นที่ใช้ โดยผู้ออกแบบ Data Warehouse และผู้ดูแลระบบในขั้นตอน การพัฒนาและการเชื่อมโยงประเด็นทางธุรกิจเข้าสู่ตัวข้อมูลใน Data Warehouse ด้วย

นอกจากนั้นก็ยังให้นิยามและบอกรายละเอียดถึง Query และรายงานแต่ละตัวให้นิยามแค่ศัพท์ทางธุรกิจและศัพท์ทางเทคนิคที่เกี่ยวข้อง นอกจากนั้นก็ยังให้รายละเอียดเกี่ยวกับผู้ดูแลรักษาข้อมูลแต่ละตัว นอกจากนั้น Metadata ยังให้ข้อมูลในขั้นตอนการออกแบบและให้นิยามฐานข้อมูลของ Data Warehouse แหล่งข้อมูลของ Data Warehouse นอกจากนั้น ยังกำหนดกฎสำหรับใช้กำจัดขยะและปรับปรุงข้อมูลในฐานข้อมูลด้วย

ข้อมูลที่เก็บอยู่ใน Metadata

Metadata จะเป็นข้อมูลที่บอกถึงข้อเท็จจริงในแง่มุมต่าง ๆ ของคลังข้อมูล ซึ่งได้แก่ คำนิยามต่าง ๆ รวมถึงข้อมูลที่ใช้ในการดำเนินการกับคลังข้อมูลซึ่งทั้งสองสามารถแบ่งออกได้อีกเป็นข้อเท็จจริงโดยทั่วไป (general facts) และข้อเท็จจริงที่มีอยู่เฉพาะในองค์กร (organization-customized facts) (กิตติพงษ์ กลมกล่อม, 2552, หน้า 134-135)

1. คำนิยามของข้อมูล (data definition) จะเก็บแหล่งที่มาของข้อมูลความสัมพันธ์ในเชิงการประมวลผลของข้อมูลและการใช้งานข้อมูล โดยข้อมูลไม่ได้หมายถึงเฉพาะข้อมูลที่อยู่ในรูปของ Entity หรือ Fact Table เท่านั้น แต่หมายถึงข้อมูลในทุก ๆ ระดับ ไม่ว่าจะเป็น Entity Attribute หรือแม้แต่ Value ของแต่ละ Attribute เรียกว่า หน่วยของ

ข้อมูล (data unit) ภายในคลังข้อมูล หน่วยข้อมูลนั้นจะต้องมีคำอธิบายเพราะเป็นระบบขนาดใหญ่ จะต้องมีเวลาที่ระบบจะต้องถูกใช้งานยาวนาน ซึ่งอาจจะมีการเปลี่ยนแปลงผู้ใช้ หรือผู้ดูแลระบบได้ เหตุผลดังกล่าวทำให้คำอธิบายในแต่ละหน่วยข้อมูลมีความจำเป็นอย่างมาก ทั้งสำหรับผู้ดูแลระบบที่อาจจะต้องมีการเปลี่ยนแปลงหรือสำหรับผู้ใช้งาน

2. แหล่งที่มาและการเข้าของข้อมูล (data entry) ข้อมูลที่จำเป็นสำหรับกระบวนการนำเข้าข้อมูลทั้งหมดจะถูกเก็บไว้ใน Metadata ข้อมูลที่เข้ามาสู่ระบบคลังข้อมูลนั้นเรียกว่า Data Acquisition File ซึ่งประกอบด้วยหลาย ๆ Entity โดย Metadata จะเก็บรายละเอียดว่า Entity แต่ละตัวจะอยู่ใน Data Acquisition File แต่ละตัวมีรอบระยะเวลาในการเข้ามาของข้อมูลถี่มากน้อยเพียงใด เช่น รายเดือนหรือรายวัน นอกจากนี้ Metadata จะเก็บข้อมูลเกี่ยวกับกฎการตรวจสอบข้อมูลและกระบวนการที่ใช้ดำเนินการ เมื่อข้อมูลเกิดความผิดพลาดขึ้นสำหรับ Data Item แต่ละตัวไว้และในการนำข้อมูลเข้าสู่ Data Warehouse Database นั้นข้อมูลของ ETL ที่ใช้สำหรับการ Transform จะถูกเก็บไว้ใน Metadata ด้วย

3. ความสัมพันธ์ในเชิงการประมวลผลข้อมูล (data derivation) Data Warehouse เป็นที่รวมของ Data และ Information โดยการประมวลผลเพื่อใช้งานอาจมีการซับซ้อน ซึ่งข้อมูลตัวใดตัวหนึ่งอาจจะถูกนำไปประมวลผลหลายทอด กว่าจะได้ผลลัพธ์ที่ต้องการ โดยทั่วไปข้อมูลจะต้องถูกประมวลผลโดยการประมวลในรูปแบบหนึ่ง เช่น ด้วยการใช้ Query ด้วยการใช้ระบบการสร้าง OLAP หรือในกรณีที่ซับซ้อนมาก ๆ อาจถูกประมวลผลโดยโปรแกรมซึ่งส่วนนี้จะถูกเก็บไว้ในคลังข้อมูลด้วย ดังนั้น จึงเป็นหน้าที่ของ Metadata ที่จะต้องบันทึกว่า Derived Data ตัวหนึ่ง ๆ เกิดจาก Data หรือ Derived Data ตัวใดบ้าง และด้วยกระบวนการหรือ Process ใด

4. การใช้งานข้อมูล (data usage) จุดมุ่งหมายของการใช้ Data Warehouse คือ ความสามารถในการสร้างรายงานที่มีประสิทธิภาพ เพื่อการวิเคราะห์ข้อมูลซึ่งสิ่งที่จะนำมาใช้สร้างรายงาน คือ ข้อมูลใน Data Mart ที่นำมาจาก Data Warehouse Database มี Metadata เป็นตัวเก็บข้อมูล เกี่ยวกับการสร้างผลลัพธ์ด้วยลักษณะการเก็บข้อมูลคล้าย ๆ กับ Data Derivation แต่วัตถุประสงค์ที่แตกต่างกันเพราะ Data Derivation

บอกถึงความสัมพันธ์ระหว่างข้อมูลใน Data Warehouse Database แต่ Metadata จะทำหน้าที่บอกความสัมพันธ์ระหว่างข้อมูลใน Data Warehouse Database และ Data Mart

5. ข้อมูลอื่น ๆ ที่จำเป็น Metadata ยังสามารถเก็บข้อมูลอื่น ๆ ที่จำเป็นได้ เช่น เก็บผู้ที่มีหน้าที่ดูแลรับผิดชอบในข้อมูลต่าง ๆ หรือเก็บข้อมูลผู้ใช้คนใดใช้ข้อมูลตัวใดบ้าง เป็นต้น

แนวทางการออกแบบ Metadata Repository

เป็นการออกแบบ Data Model ทั้งส่วนที่เป็น Logical Data Model และ Physical Data Model ของ Metadata ซึ่งสามารถกระทำได้โดยใช้วิธีการออกแบบที่แบ่งเนื้อที่ใน Data Model สำหรับ Metadata ออกเป็น Subject Area ต่าง ๆ ที่จำเป็นและออกแบบ Logical Data Model รวมทั้ง Physical Data Model ตามแนวทางการออกแบบ Data Model สำหรับ Data Warehouse โดยแบ่ง Subject Area ของ Metadata ออกเป็น 3 Subject Area ด้วยกัน คือ (กิตติพงษ์ กลมกล่อม, 2552, หน้า 139)

1. Data Subject Area อธิบายข้อมูลต่าง ๆ ใน Data Warehouse

2. Process Subject Area Process ต่าง ๆ ที่ต้องมีใน Data Warehouse

3. Definition Subject Area ให้คำจำกัดความกับข้อมูลและกระบวนการต่าง ๆ ที่มีอยู่ใน Data Warehouse

การเปลี่ยนแปลงแก้ไข Metadata

Metadata ทำหน้าที่ในการเก็บรวบรวมข้อมูลต่าง ๆ ที่สำคัญเกี่ยวกับส่วนต่าง ๆ ของ Data Warehouse ไว้เพื่อประโยชน์ในการควบคุมการทำงานของ Data Warehouse แต่ในการสร้าง Metadata Repository การเปลี่ยนแปลงที่เกิดขึ้นกับส่วนต่าง ๆ ของ Data Warehouse นั้นจะเป็นการเปลี่ยนแปลงที่เกิดขึ้นใน Metadata ก่อนแล้วจึงเปลี่ยนแปลงส่วนประกอบต่าง ๆ โดย Metadata จะต้องเก็บข้อมูลที่ถูกต้องของระบบเสมอโดยการเปลี่ยนแปลงแก้ไขระบบ ทำได้ 2 วิธี ดังนี้ (กิตติพงษ์ กลมกล่อม, 2552, หน้า 154)

1. การเปลี่ยนแปลงแบบ Metadata Driven Approach คือ ไม่ว่าเกิดการเปลี่ยน-แปลงใด ๆ กับระบบ Metadata Repository จะทำหน้าที่เป็นส่วนด่านหน้าในการรับ

การร้องขอจากภายนอกเสมอและข้อมูลของระบบที่อยู่ใน Metadata ถูกเปลี่ยนแปลงเป็นที่แรกก่อนที่จะมีการเปลี่ยนแปลงในส่วนอื่นของระบบเกิดขึ้น การเปลี่ยนแปลงแบบ Metadata Driven Approach เหมาะสมกับ Data Warehouse ที่มีปริมาณข้อมูลมาก มีความสัมพันธ์ระหว่างข้อมูลมีมากและมีกระบวนการต่าง ๆ ที่ซับซ้อนเป็นระบบที่มีความสำคัญสูง เนื่องจากมีบุคคลได้กลุ่มได้รับผลกระทบจากการเปลี่ยนแปลงระบบ ซึ่งการเปลี่ยนแปลงระบบแต่ละครั้งจะมีโอกาสสูงที่จะเกิดผลกระทบอย่างมากกับการดำเนินการของระบบและองค์กร

2. การเปลี่ยนแปลงแบบ Component Driven Approach คือ การเปลี่ยนแปลงของระบบ Data Warehouse ที่จะเกิดขึ้นในแต่ละส่วนก่อนที่การเปลี่ยนแปลง Metadata จะเกิดขึ้น แนวทาง Component Driven Approach นี้ การเปลี่ยนแปลงสามารถเกิดได้จากแต่ละส่วนของระบบ หลังจากนั้น Metadata Repository จะรับรู้การเปลี่ยนแปลงที่เกิดขึ้นและทำหน้าที่เปลี่ยนแปลงข้อมูลใน Metadataให้สอดคล้องกับการเปลี่ยนแปลงที่เกิดขึ้น การเปลี่ยนแปลงแก้ไขระบบแบบ Component Driven Approach เหมาะสมกับ Data Warehouse ที่มีปริมาณข้อมูลไม่มากนักความซับซ้อนและกระบวนการต่าง ๆ มีน้อยและมีผู้ที่ได้รับผลกระทบจากการเปลี่ยนแปลงแก้ไขระบบไม่มากนัก

OLTP และ OLAP

Online Transaction Processing (OLTP)

สมพร จิวรสกุล (2545, หน้า 696) อธิบายว่า ตามปกติแล้วระบบงานที่ถูกสร้างเพื่อให้รองรับกับ Transaction ที่เกิดขึ้นในแต่ละวันทั้งการ Insert, Update, Delete หรือกล่าวได้ว่า อยู่ในลักษณะการทำงานที่เป็นแบบ Online Transaction Processing (OLTP) เช่น ระบบการขายและการจัดส่ง ระบบจองตั๋วล่วงหน้า ระบบการโอนบัญชีธนาคาร เป็นต้น ในการออกแบบการจัดเก็บระบบฐานข้อมูลของ OLTP มักเป็นการสร้างความสัมพันธ์กันในลักษณะของตารางที่ลดความซ้ำซ้อนของข้อมูล ด้วยการใช้คีย์หลักและคีย์นอกเป็นตัวเชื่อมโยงระหว่างกัน เพื่อเป็นไปตามหลักของ Normalization เท่าที่สามารถจะทำได้ โดยทั่วไปแล้ว OLTP มักถูกสร้างเพื่อให้รองรับกับวัตถุประสงค์ ดังต่อไปนี้

1. การทำงานกับข้อมูลที่เกิดขึ้นถูกสร้างเป็น Transaction

2. ข้อมูลมีความถูกต้องแน่นอน โดยกำจัดความซ้ำซ้อนของข้อมูลออกไป

3. เพื่อให้มั่นใจได้ว่าข้อมูลมีความสมบูรณ์ในตัวเอง

4. สร้างเอกสารหรือรายงานตามระยะเวลาสั้น ๆ เช่น ใบส่งของหรือรายงานประจำวัน เป็นต้น

5. ช่วยปรับปรุงและเพิ่มเติมความสามารถในการทำงานให้มีประสิทธิภาพ

ปัจจัยที่จะต้องคำนึงถึงในการจัดทำระบบ OLTP มีดังนี้

1. ขนาดและตำแหน่งของ Rollback Segment

2. ดัชนีการจัดกลุ่มและการคำนวณตำแหน่งที่อยู่ (hashing)

3. การออกแบบข้อมูลธุรกรรมให้เหมาะกับงานประยุกต์

4. หน่วยเก็บและเนื้อที่ว่างสำหรับการเก็บข้อมูลใหม่

5. ความเข้าใจลักษณะงานประยุกต์และการเขียนคำสั่งสำหรับค้นคืนข้อมูล

6. การปรับปรุงสมรรถนะของระบบอย่างต่อเนื่อง

ระบบ Online Transaction Processing (OLTP) ที่พัฒนาขึ้นโดยใช้เทคนิคด้านฐานข้อมูลตามปกติ มักจะไม่สามารถรองรับกับปริมาณข้อมูลที่เพิ่มขึ้นอย่างมากมายเป็นประจำทุกวันได้ การนำระบบเช่นนี้มาใช้จึงมีความเสี่ยงที่จะเกิดความผิดพลาดเสียหายขึ้น วิธีการแก้ไข คือ การแยกฐานข้อมูลออกมาเป็นส่วนให้เหมาะกับการใช้งาน

Online Analytical Processing (OLAP)

Train and Jacobson (2001) อธิบายว่า การสร้างระบบคลังข้อมูล เป็นการรวบรวมและจัดเก็บข้อมูลที่มาจากหลายแหล่ง ให้สามารถตอบสนองต่อการหาคำตอบของ Query จากผลลัพธ์ของการดำเนินงานที่ผ่านมาตลอดจนการคาดเดาข้อมูลที่อาจเกิดขึ้นในอนาคต หรือกล่าวได้ว่าอยู่ในลักษณะการทำงานที่เป็นแบบ Online Analytical Processing (OLAP) เช่น Decision Support System (DSS) และ Executive Information System (EIS) เป็นต้น ในการออกแบบการจัดเก็บระบบฐานข้อมูลของ OLAP จะตรงข้ามกับ OLTP โดยสิ้นเชิง กล่าวคือ เป็นการกระจายข้อมูลออกด้วยการกระทำ De-Normalization จากตารางที่เคย Normalize ถึงแม้ว่าจะทำให้ข้อมูลมีความซ้ำซ้อนเกิดขึ้นแต่จะไม่ทำให้เกิดปัญหาตามมา

เพราะข้อมูลอยู่ในลักษณะของการอ่านเพียงอย่างเดียว อีกทั้งยังได้มาซึ่งประสิทธิภาพของการ Query ที่รวดเร็วยิ่งขึ้น เนื่องจากการลด Join ของตารางออกไปนั่นเอง โดยทั่วไปแล้ว OLAP มักถูกสร้างเพื่อให้รองรับกับวัตถุประสงค์ ดังต่อไปนี้

1. ตอบสนองกับการแสดงและนำเสนอข้อมูลได้เป็นอย่างดี
2. สร้างความยืดหยุ่นและง่ายต่อการทำรายงานในการ Queries ต่าง ๆ
3. การทำวิเคราะห์ข้อมูลได้ในทุกมุมมองขององค์ประกอบที่เกี่ยวข้องทั้งหมด
4. ช่วยคาดคะเนหรือทำนายผลที่จะตามมาในอนาคต

ปัจจัยที่ต้องคำนึงถึงในการจัดทำระบบ OLAP มีดังนี้

1. ความสามารถการวิเคราะห์ข้อมูลแบบหลายมิติ (multidimensional analysis)
2. สถาปัตยกรรมแบบ Client/Server
3. ความสามารถในการจัดเก็บข้อมูลและความเร็วในการเข้าถึงข้อมูลขนาดใหญ่
4. เครื่องมือในการพัฒนาระบบ (application development tool)
5. การดึงข้อมูล (transformation) จากแหล่งต่าง ๆ เช่น ระบบคลังข้อมูล Flat File, Spreadsheet เป็นต้น
6. ความสามารถในการจัดลำดับชั้นของข้อมูล (hierarchy)

ปัจจัยที่จะทำให้ได้ตามกล่าวนี้มีอยู่ 3 ข้อ คือ

1. จะต้องมีระบบจัดคำสั่งค้นคืนข้อมูลให้ทำงานได้รวดเร็วที่สุด
2. การจัดดัชนีจัดกลุ่มข้อมูลและการคำนวณตำแหน่งที่อยู่ข้อมูล
3. การประมวลผลคำสั่งค้นคืนในแบบขนานโดยเฉพาะเมื่อใช้หน่วยเก็บแบบ RAID

แม้ระบบ OLTP และ OLAP นี้จะเกี่ยวข้องกับข้อมูลธุรกรรมเหมือนกัน แต่ก็มีความแตกต่างกันมาก ในกระบวนการทำงานที่เกี่ยวกับข้อมูล หากพบว่าการอ่านข้อมูลจากฐานข้อมูลมาประมวลผลมีช่วงเวลาโต้ตอบ (response time) ช้ามากและต้องการปรับการเก็บโดยการจัดทำดัชนีเพิ่มเติม ให้การค้นคืนข้อมูลได้สะดวกขึ้น ก็จะส่งผลให้การบันทึกข้อมูลกลับต้องช้าลง เพราะต้องเสียเวลาดำเนินการกับดัชนีมากขึ้นกว่าระบบเดิม ด้วยเหตุนี้จึงเป็นเรื่องยากที่จะปรับระบบทั้งสองให้มีสมรรถนะดีมากขึ้นพร้อมกัน ปัจจุบันนี้แนวทางแก้ไขปัญหาข้างต้น คือ การแยกระบบ OLTP และ OLAP ออกจากกัน

ให้เป็นคนละระบบ โดยให้ระบบ OLTP สามารถจัดเก็บข้อมูลจำนวนมากได้อย่างรวดเร็วมีประสิทธิภาพและระบบ OLAP ก็สามารถค้นคืนและวิเคราะห์ข้อมูลตามความต้องการของผู้ใช้ได้อย่างรวดเร็วระบบ OLTP นั้น ปกติยังคงปล่อยให้เป็นแบบเดิมหากใช้คอมพิวเตอร์ขนาดใหญ่ เช่น เครื่องเมนเฟรมและใช้ระบบจัดการฐานข้อมูลขนาดใหญ่อยู่แล้ว ก็เพียงแต่ปรับให้สามารถบันทึกจัดเก็บข้อมูลให้เร็วขึ้น จากนั้นก็จัดทำระบบขึ้นใหม่ ให้แยกข้อมูลพื้นฐานออกจากฐานข้อมูลในระบบเดิม แล้วนำข้อมูลมาจัดทำดัชนีใหม่ เพื่อให้ผู้บริหารวิเคราะห์

การออกแบบโครงสร้างของ Dimensional Database

องค์ประกอบของ Dimensional Database

ภายใน Dimensional Database ประกอบด้วย ตารางสำคัญ 2 ตาราง คือ (กิตติพงษ์ กลมกล่อม, 2552, หน้า 85)

1. Fact Table เป็นตารางหลักของ Dimensional Database ซึ่งมีลักษณะคล้ายกันกับตารางประเภท Transaction ของ OLTP ซึ่งเป็นตารางที่มีการเกิด Transaction Insert, Delete, Update โดยส่วนใหญ่แล้วจะมีขนาดมากกว่า 80% ของข้อมูลทั้งหมดโดยภายในประกอบด้วยคอลัมน์ที่สำคัญ 2 ประเภท คือ

1.1 Fact Table Key เป็นคอลัมน์ที่ใช้เชื่อมโยงไปยัง Dimension Table ต่าง ๆ ดังนั้น จำนวนคอลัมน์ของ Fact Table Key จะเพิ่มขึ้นตามจำนวนของ Dimension Table และคอลัมน์ทั้งหมดนี้ ยังสามารถนำมาใช้สร้างให้เป็น Primary Key ของตารางด้วย

1.2 Measure เป็นคอลัมน์ที่มีประเภทของข้อมูลเป็นตัวเลข ทำหน้าที่เก็บจำนวนหรือปริมาณที่เกิดขึ้นแต่ละ Transaction นอกจากนี้ ยังสามารถเก็บผลลัพธ์ ที่ได้จากการคำนวณได้

2. Dimension Tables เป็นตารางแกนต่าง ๆ ของ Dimensional Database ซึ่งมีลักษณะคล้ายกันกับตารางประเภท Master ของ OLTP ซึ่งเป็นตารางที่เก็บข้อมูลคำอธิบาย Entity ต่าง ๆ โดยภายในประกอบด้วยคอลัมน์ที่ใช้เป็น Key เพื่อเชื่อมโยงไปยัง Fact Table Key และคอลัมน์ที่ให้ความหมายเพิ่มเติมแก่ Entity เช่น Entity นี้ประกอบด้วย

Attribute อะไรบ้าง สำหรับ Dimensional Table สามารถนำไปสร้างเป็น Dimension ของ OLAP Cube ตามลักษณะต่าง ๆ ดังต่อไปนี้

2.1 Standard Dimension มาจาก Dimension Table ซึ่งแต่ละคอลัมน์จะอธิบายข้อมูลของ Entity นั้น ๆ เพียงอย่างเดียว

2.2 Time Dimension โดยส่วนใหญ่มาจาก Dimension Table ที่สร้างขึ้นเพื่อเก็บข้อมูลเกี่ยวกับวันที่และเวลาซึ่งมีความสัมพันธ์กันกับ Fact Table โดยเฉพาะและจะมีอย่างน้อย 1 คอลัมน์ที่มีประเภทของข้อมูลเป็นแบบ Date time นอกจากนี้ Time Dimension ยังสามารถมาจาก Fact Table ที่มีคอลัมน์ซึ่งเก็บข้อมูลประเภท Date Time จะมีความสัมพันธ์ภายในกันเอง

3. OLAP Cube ภายหลังจากที่ได้มีการออกแบบโครงสร้างของตารางข้อมูลตามลักษณะของ Dimensional พร้อมทั้งโหลดข้อมูลต่าง ๆ ลงสู่ตารางเรียบร้อย แล้วจึงทำการสร้าง Cube เพื่อคำนวณหา Aggregate หรือผลลัพธ์ต่าง ๆ จากข้อมูล Measure ในส่วนของ Fact Table ตามแต่ละ Entity หรือรายการของ Dimension Table ที่เกี่ยวข้องในทุกมุมมองที่สามารถเป็นไปได้ทั้งหมด พร้อมทั้งเก็บค่าเหล่านี้ไว้ เพื่อสร้าง Query ข้อมูลต่อไปสำหรับ Query ที่เข้ามายังระบบ จะได้รับการตอบกลับอย่างรวดเร็วเนื่องจากผลลัพธ์ที่ต้องการได้รับการหาคำตอบไว้ก่อนแล้ว Cube แต่ละแกนเป็นตัวแทนของ Dimension Table ส่วนข้อมูลที่อยู่ในแต่ละ Cell คือ ค่าที่คำนวณของแต่ละ Dimension ภายใน OLAP Cube สามารถประกอบกันเพื่อแสดงข้อมูล

4. OLAP Storage ผลลัพธ์ที่ได้จากการคำนวณของ OLAP Cube จะถูกเก็บลงยัง OLAP Storage เพื่อให้สามารถนำกลับมาใช้ได้ตลอดเวลา

OLAP Cube

กิตติพงษ์ กลมกล่อม (2552, หน้า 90) อธิบายว่า ภายหลังจากที่ได้มีการออกแบบโครงสร้างของตารางข้อมูลตามลักษณะของ Dimensional พร้อมทั้งโหลดข้อมูลต่าง ๆ ลงสู่ตารางเรียบร้อยแล้ว จึงทำการสร้าง Cube เพื่อคำนวณหา Aggregate หรือผลลัพธ์ต่าง ๆ จากข้อมูล Measure ในส่วนของ Fact Table ตามแต่ละ Entity หรือรายการของ Dimension Table ที่เกี่ยวข้องในทุกมุมมองที่เป็นไปได้ทั้งหมด พร้อมทั้งเก็บค่าเหล่านี้ไว้

เพื่อสร้าง Query ข้อมูลต่อไป สำหรับ Query ที่เข้ามายังระบบ จะได้รับการตอบกลับอย่างรวดเร็ว เนื่องจากผลลัพธ์ที่ต้องการได้รับการหาคำตอบไว้ก่อนแล้ว

Cube แต่ละแกนเป็นตัวแทนของ Dimension Table ส่วนข้อมูลที่อยู่ในแต่ละ Cell คือ ค่าที่คำนวณของแต่ละ Dimension ภายใน OLAP Cube สามารถประกอบกันเพื่อแสดงข้อมูลได้มากถึง 64 Dimension

OLAP Storage

ผลลัพธ์ที่ได้จากการคำนวณของ OLAP Cube จะถูกเก็บลงยัง OLAP Storage เพื่อให้สามารถนำกลับมาใช้ได้ตลอดเวลา สำหรับ SQL Server 2005 Analysis Services สนับสนุน OLAP Storage 3 ประเภท ดังนี้ (กิตติพงษ์ กลมกล่อม, 2552, หน้า 93)

1. Multi-Dimensional OLAP (MOLAP) เป็น OLAP Storage ที่มีประสิทธิภาพสูง เนื่องจากข้อมูลทั้งที่เป็น Details และ Aggregates ถูกเก็บให้อยู่ในลักษณะของ Multi-Dimensional ดังนั้น จึงทำให้สามารถตอบสนองต่อ Query ได้รวดเร็วที่สุด นอกจากนี้ยังมีประโยชน์ในแง่ระบบการจัดเก็บข้อมูล ที่มีอัลกอริทึมของการบีบอัดที่มีประสิทธิภาพ แต่จะต้องเสียเนื้อที่ในการเก็บเพิ่มขึ้นเนื่องจากเป็นการนำข้อมูลจาก Dimensional Database มาเก็บไว้บน Analysis Services อีกทีเสมือนเป็นการซ้ำซ้อนของข้อมูล

2. Relational OLAP (ROLAP) เป็นการกำหนดโครงสร้างของข้อมูลไว้บน Analysis Services เท่านั้นแต่ข้อมูลที่เป็น Details ยังคงอยู่ที่ Relational Database ส่วนข้อมูลที่เป็น Aggregates ถูกสร้างให้เก็บไว้เป็นตารางในชื่อ Materialized Views ของ Dimensional Database พร้อมทั้งสร้าง Index เพื่อสร้างให้สามารถเข้าถึงข้อมูลได้อย่างรวดเร็วแต่การตอบสนองของ Query จึงทำได้ช้ากว่า MOLAP เนื่องจากต้องนำข้อมูลทั้งหมดจาก Relational Database มาประกอบกับโครงสร้างของข้อมูลก่อนจึงจะสามารถตอบคำถามได้

3. Hybrid OLAP (HOLAP) เป็นการผสมผสานกันระหว่าง MOLAP, ROLAP คือ ข้อมูลของ Details จะยังคงอยู่ที่ Relational Database ส่วนข้อมูลที่เป็น Aggregates จะถูกสร้างให้อยู่ในลักษณะของ Multi-Dimensional ดังนั้น จึงทำให้สามารถตอบสนองต่อ Query ได้ในระดับปานกลางในขณะที่เสียเนื้อที่ในการเก็บน้อยกว่า MOLAP การ

Processing Time จะพอ ๆ กันกับ MOLAP เนื่องจากจำนวนข้อมูลจะถูกโหลดเข้าสู่หน่วยความจำในปริมาณที่เท่า ๆ กัน การอ่านหรือการเขียนข้อมูลจึงเป็นไปในลักษณะเดียวกัน

หลักการทำงานของคลังข้อมูล

การนำเอาคลังข้อมูลไปใช้งาน

เอื้อนุ ปิ่นเงิน (2541) อธิบายว่า การนำข้อมูลจากฐานข้อมูลจากระบบประจำวันไปไว้ใน Data Warehouse เป็นกระบวนการที่ประกอบด้วยหลายขั้นตอนด้วยกัน อย่างน้อยต้องมีการ Mapping การคัดเลือกและดึงข้อมูลและแปลงข้อมูลให้อยู่ในรูปแบบที่ต้องการ ซึ่งในขั้นตอนนี้อย่างน้อยก็ต้องสร้างโปรแกรมหรือระบบที่สามารถเข้าใจความสัมพันธ์ระหว่างฐานข้อมูลประจำวันกับฐานข้อมูลของ Data Warehouse หลังจากนั้นก็เลือกดึงเอาเฉพาะข้อมูลที่ต้องการ และจัดรูปแบบใหม่ให้อยู่ในรูปแบบที่เหมาะสำหรับระบบเป้าหมาย เช่น ระบบการตัดสินใจ เป็นต้น

ส่วนประกอบหนึ่งที่มีความสำคัญ คือ Database Gateway มันจะทำหน้าที่เป็นประตูเชื่อมระหว่างฐานข้อมูลประจำวัน มีรูปแบบหลากหลายหรือแม้กระทั่งมีระบบที่หลากหลายเข้าด้วยกัน นอกจากนั้นแล้ว Gateway ยังมีหน้าที่ในการเชื่อมระหว่างเครือข่าย 2 ชนิดที่ต่างกัน โปรโตคอลสื่อสารที่ต่างกันและเชื่อม 2 ระบบ ที่มีวิธีการเก็บหรือจัดรูปแบบข้อมูลที่ต่างกัน เช่น ระหว่างเครื่องเมนเฟรมกับระบบฐานข้อมูลที่ทำงานอยู่บนระบบปฏิบัติการต่าง ๆ ข้อมูลที่ถูกวิเคราะห์แล้วว่าเป็นที่ต้องการและจำเป็นสำหรับระบบช่วยการตัดสินใจ ก็จะถูกดึงจากระบบฐานข้อมูลประจำวันหรือจากแหล่งอื่น ๆ เช่น จากบริษัทเฉพาะทางที่ประกอบอาชีพขายข้อมูล เช่น Metro Mail, A.C. Nielsen หรือ Information Source Inc. (IRI) โดยทั่วไปแล้ว การพยายามนำเอาข้อมูลมาจากหลาย ๆ แหล่ง นับเป็นเรื่องธรรมดาสำหรับโครงสร้างของ Data Warehouse แหล่งข้อมูลระดับ Field อาจจะมาจากหลายแหล่ง หลายฐานข้อมูล หลายระบบซึ่งแน่นอนว่าจะต้องประกอบด้วยข้อมูลหลาย ๆ ชนิด ในหลาย ๆ แบบ

ประโยชน์ของคลังข้อมูล

โดยทั่วไปแล้วข้อมูล Operational Database จะเก็บข้อมูลในรูปแบบ Transaction Systems เมื่อมีความต้องการข้อมูลในอันที่จะนำมาใช้ช่วยในการตัดสินใจก็จะประสบปัญหาต่าง ๆ เช่น (ทบวงมหาวิทยาลัย, โครงการเครือข่ายสารสนเทศเพื่อพัฒนาการ-ศึกษา, 2544)

1. บุคลากรทางด้าน Information Systems จำเป็นต้องเรียกข้อมูลจากฐานข้อมูลขนาดใหญ่ ซึ่งมีข้อมูลมากเกินความต้องการส่งผลให้ประสิทธิภาพของ Transaction Operational Database ทำงานได้ช้าลง

2. ข้อมูลจะเป็นรูปแบบข้อมูลตารางเท่านั้น

3. ข้อมูลจะถูกนำเสนอในรูปแบบที่ตายตัว ไม่สามารถเปลี่ยนแปลงตามความต้องการของผู้ใช้

4. ไม่ตอบสนองความต้องการของการตัดสินใจเพราะข้อมูลสำหรับการตัดสินใจมีความสลับซับซ้อนสูง มีการรวมตัวกันของข้อมูลจากตารางต่าง ๆ หลาย ๆ ตารางข้อมูล

5. ไม่ตอบสนองการสอบถามข้อมูล (data queries) สำหรับผู้ใช้

6. มีข้อมูลย้อนหลังน้อย (historical data)

7. ข้อมูลถูกจัดเก็บกระจัดกระจายตามที่ต่าง ๆ ยากต่อการเรียกใช้ ทำให้ขาดความสัมพันธ์ทางธุรกิจ อันอาจจะต้องเสียเวลาในการทำให้สอดคล้องหรือเกิดความ-ซ้ำซ้อนของข้อมูลได้

จากอุปสรรคที่กล่าวมา คลังข้อมูลจึงได้ถูกออกแบบมาเพื่อตอบสนองงานในรูปแบบการตัดสินใจ โดยการแยกฐานข้อมูลออกจาก Operational Database และเก็บข้อมูลในรูปแบบข้อมูลสรุป (summary data) ข้อมูลสรุปนี้ จะเลือกแต่เฉพาะข้อมูลที่มีความสำคัญต่อการตัดสินใจ หรือเพื่อใช้ในการบริหารไปจนถึงการกำหนดแผนงานในอนาคต ในระบบคลังข้อมูลข้อมูลที่ซับซ้อนจะถูกเก็บรวบรวมหรือเปลี่ยนแปลงให้ง่ายต่อการจัดเก็บและสามารถกลับมาใช้ได้อย่างรวดเร็วและถูกต้อง โดยข้อมูลต่าง ๆ เหล่านี้ จะถูกนำมาใช้สำหรับการวิเคราะห์และช่วยในเรื่องการตัดสินใจ โดยอาศัยเครื่องมือที่อยู่ในเครื่องคอมพิวเตอร์ที่เป็นซอฟท์แวร์ มาใช้ในการจัดการทำรายงานและเพิ่มประสิทธิภาพสำหรับการตัดสินใจให้รวดเร็วยิ่งขึ้น โดยผู้บริหาร นักวางแผนและ

นักวิเคราะห์ข้อมูลสามารถเรียกหาข้อมูลหรือสอบถาม (query) เพื่อให้ได้รับคำตอบในรูปแบบรายงานตารางหรือรายงานกราฟ เพื่อมาทำการวิเคราะห์ข้อมูลด้วยตนเอง เช่น

1. การเปรียบเทียบยอดขายระหว่างช่วงเวลาในอดีตกับปัจจุบันจนถึงการทำพยากรณ์ยอดขายในอนาคต (forecasting)

2. การหายอดขายสูงสุดหรือต่ำสุด

3. การเปรียบเทียบยอดขายต้นทุนกำไร ในรูปแบบตารางรายงานหรือรายงานกราฟ

เครื่องมือนี้ถือได้ว่าเป็นสิ่งสำคัญ ในการที่จะนำองค์กรไปสู่ความสำเร็จในกระบวนการตัดสินใจในปัจจุบัน เครื่องมือที่ตอบสนองงานเพื่อช่วยผู้บริหารสำหรับการตัดสินใจมีอยู่มากมายในตลาด ทั้งนี้ ก็เป็นทางเลือกของผู้ใช้ในการที่จะเลือกเครื่องมือที่มีประสิทธิภาพสูงสุด เพื่อตอบสนองงานของผู้บริหารในกระบวนการตัดสินใจต่อไป จะเห็นได้ว่าการจัดทำคลังข้อมูลจะมีความสำคัญมากขึ้นในอนาคต เพราะปัจจุบันนี้ผู้ใช้และผู้บริหารของหน่วยงาน เริ่มมีความเข้าใจความสำคัญของข้อมูลมากขึ้น และเริ่มตระหนักว่าหากนำข้อมูลมาวิเคราะห์ให้เข้าใจสถานภาพหรือเหตุการณ์ที่เกิดขึ้นแล้วจะทำให้หน่วยงานหรือบริษัทสามารถตอบสนองต่อเหตุการณ์ที่เกิดขึ้นได้ดียิ่งขึ้น และจะทำให้หน่วยงานหรือบริษัททำงานบรรลุวัตถุประสงค์และเป้าหมายได้ดียิ่งขึ้น ตามไปด้วย คลังข้อมูลจะทำหน้าที่ในการรวบรวมคำนวณ จัดเก็บตลอดจนสามารถบริหารและจัดการข้อมูลเหล่านี้ เพื่อให้ผู้บริหารรับข้อมูลที่ถูกต้องแม่นยำต่อการตัดสินใจในแนวทางของธุรกิจ โดยทั่วไปแล้วคลังข้อมูลมักจะมีลักษณะดังต่อไปนี้

1. รวบรวมข้อมูลจากหลาย ๆ แหล่ง ปกติมักกำหนดให้อยู่ในลักษณะของการอ่านได้เพียงอย่างเดียวเท่านั้น

2. สนับสนุนข้อมูลที่ช่วยในการตัดสินใจ

3. เก็บผลของการคำนวณไว้ล่วงหน้า เพื่อช่วยให้รายงานข้อมูลได้เร็วขึ้น

4. ช่วยในการวิเคราะห์แนวทางของธุรกิจได้ตามระยะเวลาที่ผ่านมา

5. มีการโหลดข้อมูลจากแหล่งต่าง ๆ อยู่เป็นประจำเพื่อให้ข้อมูลที่ถูกอัพเดตหรือเพิ่มเติมเข้ามาอยู่ในการวิเคราะห์ด้วยเสมอ

การออกแบบโครงสร้างในการจัดเก็บข้อมูลของคลังข้อมูล

การออกแบบโครงสร้างในการจัดเก็บข้อมูลของคลังข้อมูล สามารถแบ่งออกเป็น (ภัทรชัย ลลิตโรจน์วงศ์, 2547, หน้า 51-52)

1. กำหนดความต้องการ ซึ่งประกอบด้วย ความต้องการทางธุรกิจของผู้ใช้และความต้องการทางเทคนิค กำหนดขอบเขตงานและกำหนดแหล่งที่มาของข้อมูลที่ต้องใช้ โดยศึกษาจากงานที่ใช้อยู่เดิม ความต้องการและรูปแบบของรายงานที่ใช้สำหรับวิเคราะห์

2. ออกแบบและสร้าง Database โครงสร้างของ Database สำหรับงานคลังข้อมูลจะแตกต่างจากการออกแบบงานที่เป็น Relational Database สำหรับคลังข้อมูลนั้น จะเน้นการออกแบบให้ De-Normalization มากที่สุด คือ ไม่จำเป็นต้องเป็นไปตามกฎของ Normalize เพื่อให้ Query ได้ผลเร็วที่สุด จึงยอมให้เกิดการซ้ำซ้อนของข้อมูลในแต่ละ Table ได้ถ้าจำเป็น ซึ่งการซ้ำซ้อนของข้อมูลนี้จะผิดกฎของการ Normalization ซึ่งในการออกแบบ Database ของคลังข้อมูลประกอบด้วย Table เป็น 2 อย่าง คือ

2.1 Fact Table เป็น Table หลักที่เก็บข้อมูล และสามารถตอบคำถามที่ต้องการได้เพียงพอ การออกแบบ Fact Table จะเลือกเฉพาะคอลัมน์ที่ต้องการใช้งานเท่านั้น และพยายามลดขนาดของคอลัมน์ที่มีความยาวมากเกินไป โดยไม่จำเป็นข้อมูลจะไม่มีการแก้ไขหรือเปลี่ยนแปลง ยกเว้นแต่การเพิ่มข้อมูลใหม่เข้าไปใน Table

2.2 Dimension Table เป็น Table ที่เก็บความหมายของรหัสที่ใช้ใน Fact Table มีประโยชน์เพื่อช่วยให้ Query แสดงคำอธิบายได้ชัดเจนขึ้นกว่าการแสดงแบบรหัสอย่างเดียวนอกจากนี้ Dimension Table ยังจัดข้อมูลเป็นหลายระดับการดูข้อมูลของ Dimension เวลาจะเลือกดูได้ตั้งแต่ระดับปี ไตรมาสและเดือน ซึ่งทำได้โดยใช้ Tool ของ OLAP เพื่อทำการ Drill Down หรือ Roll Up ไปที่ Dimension นั้นได้

3. เลือกชนิดของข้อมูลที่เป็นตัวเลขสำหรับวิเคราะห์ ในทางคลังข้อมูล จะเรียกว่า Measure ได้แก่ การเลือกคอลัมน์ที่มีชนิดข้อมูลเป็นตัวเลข เช่น ยอดขายและยอดซื้อ เป็นต้น

4. การเตรียมข้อมูล คือ การเตรียมแผนในการดึงข้อมูลจากระบบ OLTP มาสร้างเป็นข้อมูลของระบบคลังข้อมูล

5. ออกแบบการเพิ่มหรือปรับปรุงข้อมูลในคลังข้อมูล การเพิ่มข้อมูลในคลังข้อมูลเป็นสิ่งปกติ และสำคัญเนื่องจากจะต้องมีการนำข้อมูลจากระบบ OLTP มาเพิ่มที่คลังข้อมูลเพื่อให้มีข้อมูลทันสมัยนอกจากนี้ยังอาจจะมีการปรับเปลี่ยนโครงสร้างที่เก็บข้อมูลเพราะเปลี่ยนแนวความคิด หรือเพิ่มวิธีในการวิเคราะห์ข้อมูล เป็นต้น การเพิ่มและปรับปรุงข้อมูลทำได้ 3 วิธี คือ

5.1 Incremental Update คือ การเพิ่มข้อมูลต่อท้ายข้อมูลที่มีอยู่เดิมโดยไม่ได้เปลี่ยนแปลงโครงสร้างของข้อมูล และข้อมูลส่วนนี้ยังไม่เคยนำเข้าในคลังข้อมูล

5.2 Refresh Data คือ การทำให้ข้อมูลทั้งหมด มีความถูกต้องทันสมัยโดยทำการประมวล์ผลข้อมูลใหม่ทั้งหมด มักเกิดจากการเปลี่ยนแปลงโครงสร้างของ Database ในคลังข้อมูล เช่น การเพิ่ม Dimension การเพิ่ม Measure เป็นต้น

5.3 Rebuild the Dimension Structure คือ การปรับปรุงโครงสร้างที่ใช้สำหรับการวิเคราะห์ เช่น การเพิ่ม Dimension การเพิ่ม Measure เป็นต้น

6. การบำรุงรักษาระบบ Database คือ การเตรียมแผนและจัด Backup ข้อมูลจากระบบคลังข้อมูลอย่างสม่ำเสมอ รวมทั้งการทดสอบ และวางแผนการกู้ระบบ เมื่อมีความเสียหายเกิดขึ้นด้วย

การออกแบบโครงสร้างคลังข้อมูล ตามแนวคิดของ สิริวัฒน์ ธนุรเวท (2543) มีดังนี้

ขั้นที่ 1 กำหนดความต้องการของระบบ ประกอบด้วย ความต้องการทางธุรกิจของผู้ใช้ ความต้องการทางเทคนิค กำหนดแหล่งที่มาของข้อมูล ความต้องการรูปแบบรายงานที่ต้องการวิเคราะห์

ขั้นที่ 2 ออกแบบและสร้าง Database การออกแบบจะต่างจากการออกแบบทั่วไปเพราะใช้ Demoralize มากที่สุด เพื่อให้คิวรี (query) ได้ผลเร็วที่สุด

ขั้นที่ 3 เลือกชนิดของข้อมูลสำหรับวิเคราะห์ คือ การเลือกข้อมูลในการวิเคราะห์ส่วนใหญ่เป็นข้อมูลตัวเลข

ขั้นที่ 4 เตรียมข้อมูล คือ การดึงข้อมูลจากระบบ (Online Transaction Processing--OLTP) ซึ่งเป็นระบบที่มีการจัดการกับข้อมูลรายวันและมีการเปลี่ยนแปลงข้อมูลอยู่ตลอดเวลา มาสร้างเป็นข้อมูล Data Warehouse

ขั้นที่ 5 ออกแบบวิธีการเพิ่มและปรับปรุงข้อมูลในคลังข้อมูล เนื่องจากจะต้องมีการนำข้อมูลจาก OLAP มาเพิ่มเพื่อความทันสมัยของคลังข้อมูล

ขั้นที่ 6 การบำรุงรักษาระบบ Database คือ เตรียมแผนและจัด Backup ระบบคลังข้อมูลอย่างสม่ำเสมอ รวมทั้งมีการทดสอบการกู้ระบบเมื่อมีความเสียหายที่เกิดขึ้น

การออกแบบโครงสร้างในการจัดเก็บข้อมูลของคลังข้อมูล คือ การจัดรูปแบบของตารางเก็บข้อมูล โดยใช้วิธี De-Normalization ผลลัพธ์ที่เกิดขึ้นนี้จะเป็นการสร้างฐานข้อมูลให้อยู่ในลักษณะของ Dimensional โดยองค์ประกอบสำคัญในตารางมี 2 ประเภท คือ (Inmon, 1996, p. 98)

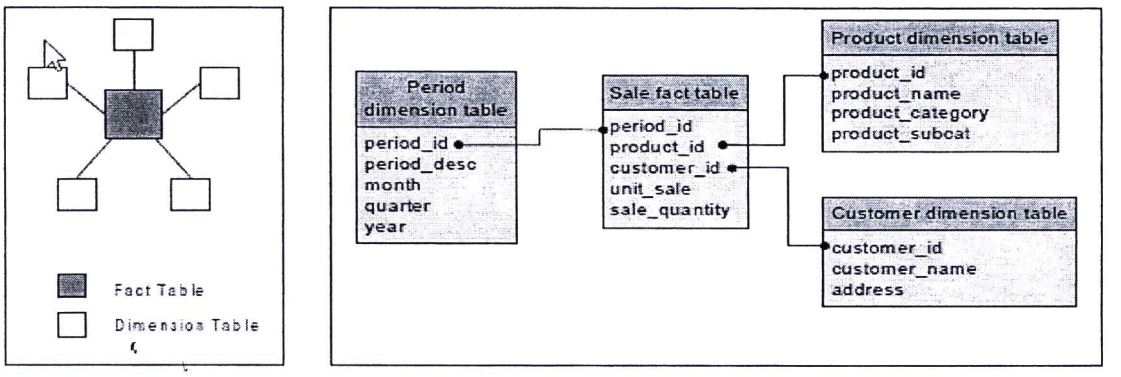
ตารางหลัก (measure/fact table) เป็นตารางหลักที่เก็บข้อมูลที่ต้องการจะนำมาวิเคราะห์ เป็นข้อมูลที่สามารถตอบคำถามที่ต้องการได้เพียงพอ เช่น ถ้าเก็บข้อมูลเกี่ยวกับการขาย ก็จะเก็บข้อมูลยอดรวมของการขายและมีรายละเอียดเกี่ยวกับ รหัสของลูกค้า รหัสสินค้า เวลา ยอดขาย เป็นต้น ปกติข้อมูลในตารางนี้มักจะไม่มีการแก้ไขหรือเปลี่ยนแปลง ยกเว้นแต่การเพิ่มข้อมูลใหม่เข้าไป

ตารางมิติ (dimension table) เป็นตารางที่เก็บความหมายของรหัสที่ใช้ในตารางหลัก มีประโยชน์เพื่อช่วยในการสืบค้นหาคำอธิบายที่ชัดเจนขึ้นของรหัสที่ใช้ในตารางหลัก เช่น ระบบการขาย จะมีตารางมิติที่เป็นตารางมิติสินค้าเก็บข้อมูลของรหัสสินค้า ชื่อ และรายละเอียดของสินค้า เป็นต้น นอกจากนี้ยังอาจจะจัดข้อมูลเป็นหลายระดับ คือ ระดับใหญ่ ระดับรอง เช่น ถ้าเลือกมิติของเวลา สามารถจัดให้มีระดับมิติที่ใหญ่สุด คือ ปี ระดับต่อมา คือ ไตรมาส และระดับถัดไป คือ เดือน เป็นต้น

การออกแบบโครงสร้างการเชื่อมโยงระหว่างตารางเก็บข้อมูล ตามลักษณะของ Dimensional แบ่งออกเป็น 2 ประเภท คือ

1. โครงสร้างแบบดวงดาว (star schema) โครงสร้างนิยมใช้ในการออกแบบคลังข้อมูล จะมีตารางหลักเป็นศูนย์รวมของข้อมูลเพียงตารางเดียว และมีตารางมิติที่มีรายละเอียดของรหัสที่ใช้ในตารางหลัก ตารางมิติจะมีจำนวนเท่าใดก็ได้ และจะมีคีย์ที่สัมพันธ์ไปยังตารางหลักเท่านั้น โครงสร้างชนิดนี้จะช่วยเพิ่มความเร็วในการสืบค้นข้อมูล เนื่องจากความสัมพันธ์ระหว่างตารางที่ไม่ซับซ้อน ลักษณะของโครงสร้าง

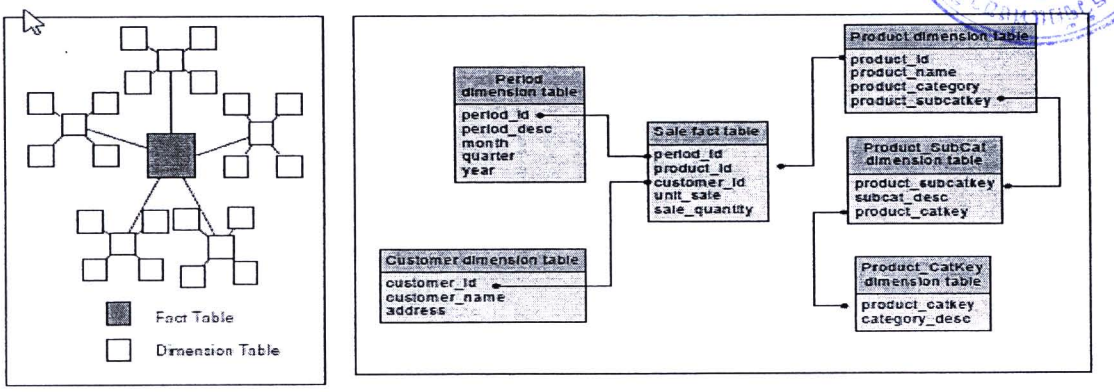
แบบดวงดาวที่สำคัญ คือ ข้อมูลเป็นแบบ Denormalized ทั้งนี้เพื่อให้สามารถค้นหาข้อมูลได้อย่างรวดเร็ว (สุนีย์ พงษ์พินิจภิญโญ, ม.ป.ป.) ดังแสดงในภาพ 6



ภาพ 6 โครงสร้างแบบดวงดาว (star schema)

ที่มา. จาก *คลังข้อมูล*, โดย สุนีย์ พงษ์พินิจภิญโญ, ม.ป.ป., ค้นเมื่อ 4 เมษายน 2549, จาก http://www.Thaicyberu.go.th/officiatcu/main_09_DatabaseSystem/public_html/lesson15/ms21.html

2. โครงสร้างแบบเกล็ดหิมะ (snowflake schema) มีโครงสร้างที่แตกต่างจากโครงสร้างแบบดวงดาว (star schema) คือ ตารางมิติ มีหลายระดับและมีคีย์ที่โยงไปยังตารางมิติอื่นอีก ดังนั้น โครงสร้างแบบนี้จะซับซ้อนมากขึ้น รวมทั้งมีผลให้การสืบค้นยากขึ้นด้วย ลักษณะของข้อมูลจะมีความเป็น Normalized (Youness, 2005, p. 70) ดังแสดงในภาพ 7



ภาพ 7 โครงสร้างแบบเกล็ดหิมะ (snowflake schema)

ที่มา. จาก *Professional Data Warehousing with SQL Server 7.0 and OLAP Service*, by S. Youness, 2005, Birmingham, England: Wrox Press.

การทำงานในการสร้างระบบคลังข้อมูล

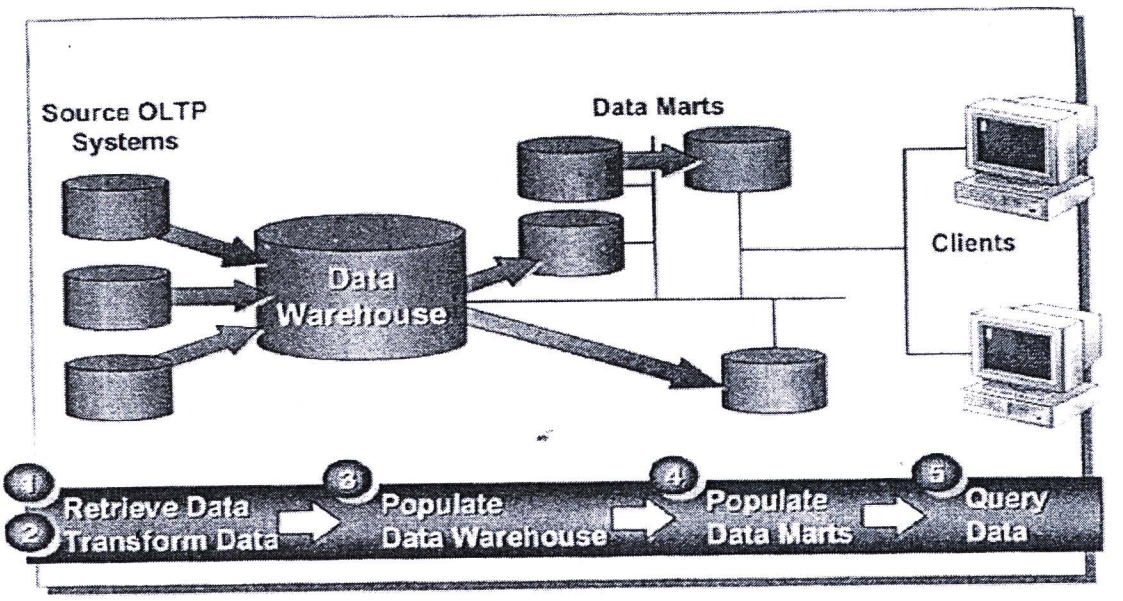
คลังข้อมูลเป็นแหล่งรวบรวมข้อมูลที่ได้จาก Transaction ที่เกิดขึ้นของ Online Transaction Processing (OLTP) ดังนี้ (สมพร จิวรสกุล, 2545, หน้า 697)

1. Design the Data Warehouse เป็นการออกแบบโครงสร้างตารางเก็บข้อมูลที่มาจาก OLTP เพื่อปรับให้อยู่ในลักษณะของ OLAP เป็นไปตามแบบของ Star Schema หรือ Snowflake Schema

2. Populate the Data Warehouse เป็นการถ่ายโอนข้อมูลตามที่ได้ออกแบบไว้ โดยใช้กระบวนการ (Extract Transform Load--ETL) เป็นเครื่องมือหลัก พร้อมทั้งมีการตรวจสอบข้อมูลให้อยู่ในรูปแบบที่ถูกต้อง เช่น การใช้ตัวย่อที่แตกต่างกันของข้อมูลในแต่ละระบบ แต่มีความหมายเดียวกันในทางปฏิบัติ นอกจากนี้ยังทำให้เรื่องของการรวมคอลัมน์ การคำนวณผลลัพธ์ ตลอดจนการแปลงอักษรให้เป็นค่าตัวเลข

3. Create OLAP Cubes เป็นการแสดง Cube ให้สนองตอบต่อการวิเคราะห์ข้อมูลที่ต้องการ รวมถึงสามารถนำไปใช้ทำนายแนวโน้มที่อาจเกิดขึ้นต่อไปในอนาคต

4. Query Data เป็นส่วนของการนำข้อมูลใน Cube ไปใช้งานแสดงผลในแบบต่าง ๆ ตามที่ออกแบบไว้ (การทำงานของระบบคลังข้อมูล) ดังแสดงในภาพ 8



ภาพ 8 การทำงานของระบบคลังข้อมูล

ที่มา. จาก *คู่มือการติดตั้งและใช้งาน Microsoft SQL Server 2000* (หน้า 697), โดย สมพร จิวรสกุล, 2545, นนทบุรี: โรงพิมพ์อินโฟเพรส.

จากภาพ 8 หลักการทำงานของระบบคลังข้อมูล แสดงถึงหลักการทำงานของระบบคลังข้อมูลซึ่งเป็นการจัดเก็บข้อมูลจากหลาย ๆ แหล่ง เพื่อให้ตอบสนองต่อการนำเสนอรายงานในรูปแบบต่าง ๆ และเพื่อใช้เป็นเครื่องมือที่ช่วยสนับสนุนการวิเคราะห์และการตัดสินใจ

การสร้างและการใช้งานระบบคลังข้อมูล

เพื่อให้ได้มาซึ่งระบบคลังข้อมูลที่สมบูรณ์ และสามารถนำไปใช้งานได้จริงในระดับ Client โดยมีองค์ประกอบที่สำคัญตลอดกระบวนการ ซึ่งมีรายละเอียดดังต่อไปนี้ (Gardner, 1998, pp. 52-53)

1. OLTP Source คือ แหล่งข้อมูล OLTP ต้นทาง ซึ่งอาจจะมาจากหลายระบบ เช่น Mainframe, AS/400, ERP หรือ RDBMS ต่าง ๆ เป็นต้น

2. Data Transformation Service (DTS) คือ คอมโพเนนต์ของ SQL Server 2005 ที่ทำหน้าที่ในการโอนข้อมูลจาก OLTP Source มายัง Data Warehouse Storage

3. Data Warehouse Storage คือ ศูนย์เก็บข้อมูลปลายทางที่ได้มีการรวบรวมข้อมูลมาจาก OLTP Source ต่าง ๆ ทั้ง SQL Server 2005 หรือ RDBMS ตัวอื่น เช่น SQL Server 6.5/7.0, Oracle, DB2, Informix, Sybase หรือแม้แต่ Microsoft Access เป็นต้น

4. Object Linking and Embedding Database (OLE DB), Open Database Connectivity (ODBC) คือ ตัวกลางซึ่งทำหน้าที่ในการติดต่อกับแหล่งข้อมูลประเภทต่าง ๆ

5. Analysis Service ทำหน้าที่เกี่ยวกับการจัดการและบริหาร OLAP Cube

6. OLE DB for OLAP เช่นเดียวกับ OLE DB ทั่วไป แต่มีไว้เพื่อเชื่อมต่อกับ Analysis Service

7. Active X Data Objects Multi Dimensional (ADO-MD) เป็นอีกช่องทางหนึ่งที่สามารถติดต่อกับ Analysis Service ได้ แต่เน้นไปที่การพัฒนา Application ขึ้นมาใช้เองจาก Tool ต่าง ๆ ที่สนับสนุน เช่น Visual Basic, Visual C++ เป็นต้น

8. Clients คือ Application ทางฝั่ง Client ซึ่งนำมาเชื่อมต่อกับ Analysis Service เพื่อให้สามารถนำข้อมูลมาวิเคราะห์ได้ เช่น Microsoft Excel, Application จาก Third-Party หรือ Application ที่พัฒนาขึ้นมาใช้เอง

การวิเคราะห์และการออกรายงาน

สามารถใช้โปรแกรมพวก Spreadsheet อย่าง Lotus 1-2-3 และ Microsoft Access ระบบฐานข้อมูลบนเครื่องเดสก์ทอปและโปรแกรมที่ทำหน้าที่เป็น Front End สำหรับการเข้าถึงข้อมูล เช่น Business Objects และ Forest & Tree เป็นต้น เครื่องมือพวกนี้ทำให้สามารถเข้าใช้งานข้อมูลใน Data Warehouse ได้โดยไม่ต้องรู้โครงสร้างภายในหรือไม่รู้วิธีการเรียกใช้คำสั่ง SQL เลย นอกจากนี้สามารถเจาะลึกเข้าไปในข้อมูล หรือสำรองออกมาเก็บไว้ใน Spreadsheet เพื่อจะเอามาวิเคราะห์ต่อไป สำหรับการเข้าถึงข้อมูลสามารถแบ่งได้ ดังนี้ (Bontempo & Zagelow, 1998, pp. 38-40)

1. เครื่องมือเข้าถึงข้อมูลและ Query พวกนี้จะเป็น GUI เพื่อช่วยเราเข้าใช้ Data Warehouse ทำให้ผู้ใช้สามารถติดต่อกับ Table ได้โดยตรง

2. ตัวสร้างรายงาน เป็นตัวให้บริการรูปแบบเพิ่มเติมของรายงาน และยังทำให้ผู้ใช้สามารถติดต่อกับ Table ได้โดยตรงเช่นกัน

3. ระบบฐานข้อมูลแบบหลายมิติ (multi-dimension database system) ให้บริการ Advance Metric และสามารถทำ Slice-and-Dice ได้ด้วย

4. ระบบช่วยการตัดสินใจระดับสูง (advanced decision support) ให้บริการวิเคราะห์แบบหลายมิติขั้นสูง โดยกระทำที่ Relational (DBMS) โดยตรงเครื่องมือประเภทนี้มักจะมี Drill Down รวมทั้งสามารถจัดการกับ Metadata ที่ใช้ร่วมกันได้

5. ระบบสารสนเทศผู้บริหาร ใช้บริการ Interface แบบมีรูปแบบเพื่อสร้างรายงานให้ข้อมูลที่สรุปเป็นพิเศษเกี่ยวกับองค์กรโดยรวม

ระบบสนับสนุนการตัดสินใจ

ระบบสนับสนุนการตัดสินใจ (decision support system) เป็นระบบที่ช่วยในการวิเคราะห์ข้อมูลที่ถูกเก็บอยู่ในคลังข้อมูล เพื่อให้สามารถทำนายแนวโน้มของข้อมูลที่ช่วยให้ผู้บริหารสามารถรู้ล่วงหน้าได้ เพื่อทำให้การบริหารงานมีประสิทธิภาพมากยิ่งขึ้น ระบบสนับสนุนการตัดสินใจมีลักษณะและความสามารถ ดังต่อไปนี้ (กิตติ ภักดีวัฒนะกุล, 2550, หน้า 21)

1. สามารถสนับสนุนการตัดสินใจทั้งในสถานการณ์ของปัญหาแบบกึ่งโครงสร้างและปัญหาแบบไม่มีโครงสร้าง

2. สามารถรองรับการใช้งานของผู้บริหารได้ทุกระดับ ตั้งแต่ผู้บริหารระดับล่างไปจนถึงผู้บริหารระดับสูง

3. สามารถส่งเสริมการตัดสินใจแบบกลุ่มและแบบเดี่ยวได้ เนื่องจากปัญหาที่เกิดขึ้นมีความแตกต่างกัน ส่งผลให้ต้องใช้ผู้ตัดสินใจจำนวนต่างกันด้วย กล่าวคือ บางปัญหาอาจอาศัยการตัดสินใจของบุคคลเพียงคนเดียวได้ แต่บางปัญหาอาจต้องอาศัยการตัดสินใจร่วมกันเป็นทีม ดังนั้น ระบบจึงต้องสามารถสนับสนุนการตัดสินใจทั้งแบบกลุ่มและแบบเดี่ยวได้

4. ระบบต้องสามารถสนับสนุนการตัดสินใจได้ทั้งปัญหาแบบเกี่ยวพัน และ/หรือ ปัญหาแบบต่อเนื่อง

5. ระบบต้องสามารถส่งเสริมกระบวนการตัดสินใจในขั้นตอน Intelligence Phase, Design Phase, Implementation Phase และ Choice Phase ของกระบวนการตัดสินใจ (decision making process)

6. ต้องสนับสนุนกระบวนการและรูปแบบการตัดสินใจที่มีความหลากหลายได้อย่างมีประสิทธิภาพ

7. ต้องมีความยืดหยุ่นสูง คือ ต้องสามารถดัดแปลงระบบเพื่อนำไปใช้กับปัญหาต่าง ๆ ได้อย่างกว้างขวาง

8. ผู้ใช้งานระบบสนับสนุนการตัดสินใจ ล้วนแต่เป็นผู้บริหารที่ไม่มีความชำนาญในการใช้คอมพิวเตอร์ทั้งสิ้น ดังนั้น การสร้างระบบจึงต้องใช้งานได้ง่ายและเข้าใจขั้นตอนการทำงานของระบบได้ง่าย เพื่อเหมาะกับการใช้งานของผู้บริหารงานในทุกระดับ

9. การพัฒนาระบบสนับสนุนการตัดสินใจ เน้นหนักไปทางด้านการทำงานที่สำเร็จตรงตามเป้าหมายมากกว่าค่าใช้จ่ายในการพัฒนาระบบ

10. เป้าหมายของระบบสนับสนุนการตัดสินใจ คือ ช่วยสนับสนุนการตัดสินใจของผู้บริหารเท่านั้น ไม่ได้ทำหน้าที่แทนผู้ตัดสินใจ ดังนั้น ต้องเข้าใจถึงขั้นตอนและกระบวนการตัดสินใจ (decision making process) ให้มากที่สุด

11. ผู้ใช้อาจทำการสร้างและปรับปรุงระบบสนับสนุนการตัดสินใจขนาดเล็กที่ทำงานอย่างง่าย ๆ ได้ด้วยตนเอง แต่สำหรับระบบสนับสนุนการตัดสินใจขนาดใหญ่ที่มีความซับซ้อน ควรขอความช่วยเหลือจากผู้เชี่ยวชาญเท่านั้น

12. ระบบสนับสนุนการตัดสินใจ เป็นระบบที่ใช้วิเคราะห์สถานการณ์ในการตัดสินใจด้วยแบบจำลองแบบต่าง ๆ ระบบจึงต้องสามารถสร้างแบบจำลอง เพื่อทดสอบป้อนค่าตัวแปรและเปลี่ยนค่าไปเรื่อย ๆ เพื่อสร้างทางเลือกใหม่ ๆ

13. สามารถเข้าถึงแหล่งเก็บข้อมูลได้หลากหลาย และต้องสามารถแสดงผลข้อมูลในรูปแบบที่หลากหลายได้เช่นกัน

ประโยชน์ของระบบสนับสนุนการตัดสินใจ

การพัฒนาระบบสนับสนุนการตัดสินใจ ช่วยลดภาระหน้าที่ของผู้บริหารในการตัดสินใจด้านต่าง ๆ และช่วยให้การตัดสินใจบริหารองค์กรเป็นไปอย่างราบรื่น โดยประโยชน์ที่ได้รับจากการนำระบบสนับสนุนการตัดสินใจมาใช้ในองค์กร ดังนี้ (กิตติ ภักดีวัฒนะกุล, 2550, หน้า 22-23)

1. พัฒนาประสิทธิภาพการทำงานส่วนบุคคล โดยเฉพาะงานที่เกี่ยวข้องกับการตัดสินใจ ซึ่งเป็นงานหลักของผู้บริหาร เนื่องจากระบบช่วยจัดเตรียมสารสนเทศที่มีประสิทธิภาพ เพื่อใช้ประกอบการตัดสินใจ

2. พัฒนาประสิทธิภาพการแก้ไขปัญหา โดยช่วยให้ผู้ตัดสินใจสามารถแก้ปัญหาได้รวดเร็ว และถูกต้องมากยิ่งขึ้น และยังสามารถช่วยตัดสินปัญหากึ่งโครงสร้างและปัญหาที่ไม่มีโครงสร้างได้อย่างมีประสิทธิภาพ โดยระบบอาจมีการจัดเตรียมสารสนเทศเกี่ยวกับการตัดสินใจของปัญหาในลักษณะเดียวกับในอดีต และผลที่ได้รับจากการตัดสินใจนั้น ๆ เพื่อพิจารณาประกอบการตัดสินใจของผู้ใช้ระบบ ซึ่งช่วยให้การตัดสินใจมีความถูกต้อง รวดเร็ว และน่าเชื่อถือมากยิ่งขึ้น

3. ช่วยอำนวยความสะดวกในการติดต่อสื่อสารสำหรับระบบสนับสนุนการตัดสินใจที่มีการทำงานในลักษณะกลุ่มที่เรียกว่า “Groupware” ทำให้ผู้บริหารสามารถตัดสินใจปัญหาที่ต้องอาศัยการตัดสินใจร่วมกันของกลุ่มผู้บริหารได้ โดยทำการปรึกษาประชุม และเรียกใช้สารสนเทศเพื่อประกอบการตัดสินใจผ่านเทคโนโลยี เครือข่ายคอมพิวเตอร์ช่วยให้ประหยัดเวลา งบประมาณ และช่วยให้การประชุมติดต่อระหว่างผู้บริหารต่าง ๆ เป็นไปโดยสะดวก ช่วยให้การตัดสินใจมีประสิทธิภาพมากขึ้น

4. ส่งเสริมการเรียนรู้หรือการฝึกหัด เนื่องจากระบบสนับสนุนการตัดสินใจมีกระบวนการทำงานคล้ายกระบวนการตัดสินใจของมนุษย์ ดังนั้น เมื่อมีการใช้ระบบสนับสนุนการตัดสินใจซ้ำ ๆ จึงช่วยพัฒนาการเรียนรู้และช่วยฝึกหัดการใช้งานระบบให้กับผู้ใช้ โดยผู้ใช้สามารถศึกษากระบวนการให้เหตุผลของระบบสนับการตัดสินใจผ่านการสอบถามถึงลักษณะปัญหา ขั้นตอนการวิเคราะห์ปัญหา กระบวนการให้ข้อเสนอแนะและกระบวนการให้เหตุผล โดยระบบสนับสนุนการตัดสินใจที่ช่วยส่งเสริม

พัฒนาการเรียนรู้และการฝึกหัดของผู้ใช้ คือ ระบบผู้เชี่ยวชาญ

5. เพิ่มประสิทธิภาพการควบคุมองค์กร เนื่องจากระบบสนับสนุนการตัดสินใจมีส่วนช่วยเพิ่มประสิทธิภาพการตัดสินใจให้กับผู้บริหาร ทำให้สามารถบริหารควบคุมองค์กรได้ดียิ่งขึ้น การบริหารและควบคุมองค์กรเป็นงานที่ต้องอาศัยการตัดสินใจหลาย ๆ ด้าน ดังนั้น การตัดสินใจที่มีประสิทธิภาพจะช่วยให้องค์กรสามารถดำเนินการได้อย่างราบรื่น เป็นไปตามวัตถุประสงค์ขององค์กร

ระบบสารสนเทศการบริหารโรงพยาบาล

ระบบสารสนเทศการบริหารโรงพยาบาล มีส่วนประกอบหลัก 3 ประการ ดังนี้ (พงษ์พิสุทธิ์ จงอุดมสุข, สมชาติ โตรักษา และสุกิจ พึ่งเกศสุนทร, 2545, หน้า 108, 210)

1. การวางระบบงานเพื่อปฏิบัติงาน

2. การควบคุมดูแลการปฏิบัติงานตามระบบที่วางไว้

3. การพัฒนาระบบงานโดยการทำให้ทรัพยากรที่มีอยู่ให้เกิดประสิทธิผลและประสิทธิภาพสูงสุด

ระบบสารสนเทศการบริหารโรงพยาบาล เป็นเครื่องมือสำคัญในการบริหารการบริการและพัฒนาโรงพยาบาล ระบบสารสนเทศในโรงพยาบาล ประกอบด้วยระบบสารสนเทศเพื่อการบริการ ระบบสารสนเทศเพื่อการสนับสนุนการบริการและงานวิชาการ และระบบสารสนเทศเพื่อการบริหาร ซึ่งประกอบด้วย ข้อมูลสารสนเทศและเทคโนโลยีสารสนเทศ โดยผู้บริหารต้องทำความเข้าใจเพื่อให้สามารถวางแผนการให้บริการอย่างถูกต้องรวดเร็ว ทันท่วงที เพื่อการวางแผนงาน การกำหนดนโยบายและกลยุทธ์ทิศทางในการบริหารจัดการองค์กร และเพื่อนำมาใช้ประกอบการตัดสินใจอย่างมีประสิทธิภาพ

งานวิจัยที่เกี่ยวข้อง

งานวิจัยในประเทศ

คงเดช บุญยกิจสมบัติ (2553) ศึกษาเรื่อง *การพัฒนาระบบคลังข้อมูลและการทำเหมืองข้อมูลสำหรับผู้ติดเชื้อ HIV ในโรงพยาบาลสังกัดสำนักอนามัย กรุงเทพมหานคร* มีวัตถุประสงค์เพื่อพัฒนาระบบคลังข้อมูลสำหรับข้อมูลผู้ติดเชื้อเอชไอวี เพื่อรวบรวมข้อมูลที่แตกต่างกันมาไว้บนฐานข้อมูลเดียว ระบบสามารถคัดกรองข้อมูลที่สำคัญและวิเคราะห์ข้อมูลตามความต้องการได้อย่างรวดเร็ว ผู้ใช้งานโปรแกรมสามารถเลือกกำหนดเงื่อนไขต่าง ๆ และเข้าถึงข้อมูลได้เองในรูปแบบ Data Cube และทำการพัฒนาระบบเหมืองข้อมูลโดยใช้อัลกอริธึมต้นไม้ตัดสินใจ ให้สามารถกำหนดรูปแบบของข้อมูลได้ ระบบนี้พัฒนาขึ้นโดยใช้ด้วยภาษา MS Visual Basic 6.0 และใช้ฐานข้อมูล MS SQL Server 2000 ในการทดสอบได้นำโปรแกรมติดตั้งและใช้งานโดยทดลองใช้งานกับผู้เชี่ยวชาญ 2 กลุ่ม คือ ผู้เชี่ยวชาญทางด้านเทคโนโลยีคอมพิวเตอร์ และผู้เชี่ยวชาญทางด้านโรคเอดส์ โดยการประเมินผลทำการประเมิน 3 ด้าน คือ (1) ด้านความสามารถการทำงานตามหน้าที่ (2) ด้านสภาพแวดล้อมของโปรแกรม และ (3) ด้านเสถียรภาพของโปรแกรม ผลจากการประเมินโดยรวมได้ค่าเฉลี่ยเลขคณิต 4.38 ส่วนเบี่ยงเบนมาตรฐาน 0.49 สรุปได้ว่า ด้านความสามารถทำงานตามหน้าที่และเสถียรภาพของระบบได้ระดับดีและด้านสภาพแวดล้อมของโปรแกรมได้ระดับปานกลาง

ศันสนีย์ ชาตตระกูล (2552) ศึกษาเรื่อง *การพัฒนาคลังข้อมูลเพื่อระบบสนับสนุนการตัดสินใจกรณีศึกษา: ระบบแถวคอย แผนกอายุรกรรม โรงพยาบาลนครธนบุรี* การวิจัยนี้เป็นการนำแนวคิด เรื่องคลังข้อมูลมาประยุกต์เข้ากับระบบสนับสนุนการตัดสินใจเพื่อใช้สำหรับโรงพยาบาล ปัจจุบันโรงพยาบาลได้มีการจัดทำระบบจัดการฐานข้อมูลไว้เก็บข้อมูลจำนวนมาก แต่ไม่ได้นำข้อมูลเหล่านั้นมาใช้ประโยชน์อย่างเต็มที่ เนื่องจากไม่ได้นำข้อมูลที่มีอยู่มาทำการวิเคราะห์และสรุปผล การวิจัยนี้จัดทำรายงานสรุปทั่วไปเกี่ยวกับผู้ป่วย รายรับโรงพยาบาลและรายงานสรุปเพื่อประกอบการตัดสินใจในเรื่องระบบแถวคอย เพื่อแสดงให้ผู้บริหารทราบว่าในปัจจุบันจำนวนแพทย์ผู้ให้การรักษามีจำนวนพอเหมาะกับจำนวนผู้ป่วยที่มารับบริการหรือไม่ โดยแบ่งช่วงการทำรายงานสรุป

เป็น 2 ช่วง คือ ในช่วงวันจันทร์ถึงวันศุกร์ และวันเสาร์ถึงวันอาทิตย์ ทั้งนี้การวิเคราะห์ระบบแถวคอยใช้แบบจำลองคณิตศาสตร์ โดยลักษณะแถวคอยที่ทำการศึกษาตรงกับรูปแบบการวิเคราะห์ระบบแถวคอยของ Markov Chains แบบ M/M/1 และ M/M/C ทำการเก็บรวบรวมข้อมูลโดยการสัมภาษณ์ผู้บริหาร เจ้าหน้าที่โรงพยาบาลและให้ผู้ป่วยกรอกแบบสอบถาม ข้อมูลที่นำมาสรุปได้จากฐานข้อมูลที่มีอยู่แล้วและทำการเพิ่มข้อมูลบางอย่างที่ไม่ได้มีเก็บไว้ลงไป ส่วนการพัฒนาระบบคลังข้อมูลใช้โปรแกรมไมโครซอฟท์-แอคแซส เพื่อจัดทำคลังข้อมูลโดยเชื่อมข้อมูลมาจากฐานข้อมูลของโรงพยาบาลที่มีอยู่แล้ว ใช้โปรแกรมวิชวลเบสิคแสดงผลในรูปแบบกราฟฟิคเพื่อง่ายต่อการใช้งาน ผลการวิจัยทำให้ทราบถึงผลการดำเนินงาน ที่ผ่านมาเกี่ยวกับผู้ป่วย รายรับของโรงพยาบาลและผลการวิเคราะห์ระบบแถวคอย แสดงให้เห็นว่ามีบางช่วงเวลาที่ยังมีแถวคอยอยู่มาก หรือบางช่วงแพทย์ผู้ให้บริการว่าง ผลการวิเคราะห์นี้เป็นแนวทางให้ผู้บริหารใช้ประกอบการตัดสินใจในการเลือกวางแผนกำหนดจำนวนแพทย์ในโรงพยาบาลต่อไป ผู้บริหารควรนำปัจจัยอื่น เช่น ค่าจ้างแพทย์ มาประกอบการตัดสินใจด้วย การพัฒนาคลังข้อมูลเพื่อประกอบการตัดสินใจ ได้แสดงผลอย่างชัดเจนว่าสามารถทำให้การดึงข้อมูลสรุปเป็นไปอย่างรวดเร็วและนำข้อมูลที่มีอยู่มาใช้ให้เกิดประโยชน์

สันติ อาริยทรัพย์, เทอดพงษ์ ผลจันทร์ และปรีชา น้อยอำคา (2552) ศึกษาเรื่อง *การวิเคราะห์ข้อมูลเพื่อการตัดสินใจด้วย Business Intelligence* มีการนำเอาข้อมูลสารสนเทศที่มีอยู่มาก่อให้เกิดประโยชน์สูงสุด เพื่อช่วยให้การตัดสินใจที่ถูกต้องและแม่นยำ โดยใช้เทคโนโลยีเป็นส่วนประกอบที่ทำให้ประสบผลสำเร็จ เป้าหมายของ Business Intelligence คือ การนำข้อมูลมากมายมาก่อให้เกิดประโยชน์ กระบวนการในการจัดทำ Business Intelligence เริ่มต้นที่มีการกำหนดแหล่งข้อมูล (data sources) ที่จะนำเข้ามาสู่คลังข้อมูล โดยแหล่งข้อมูลสามารถแบ่งแหล่งข้อมูลออกได้เป็น 2 ประเภท คือ แหล่งข้อมูลภายใน (internal data sources) และแหล่งข้อมูลภายนอก (external data sources) แหล่งข้อมูลภายใน ได้แก่ ข้อมูลการดำเนินงาน (operation transaction) ข้อมูลอดีต (legacy data) เป็นต้น แหล่งข้อมูลภายนอก (external data sources) ได้แก่ ข้อมูลสถิติจากสถาบันต่าง ๆ ข้อมูลของโครงการสารสนเทศอื่น ๆ บทวิเคราะห์และบทความวิชาการต่าง ๆ ซึ่งในการกำหนดแหล่งข้อมูล จำเป็นจะต้องคำนึงถึงผลลัพธ์ที่ต้องการ

เพื่อที่ว่าข้อมูลที่นำเข้ามาใช้งานจะสามารถสอดคล้องกับผลลัพธ์ที่ต้องการ เมื่อมีการกำหนดแหล่งข้อมูลที่แน่ชัด ขั้นตอนถัดไป คือ การออกแบบคลังข้อมูล (data warehouse design) เพราะว่า Business Intelligence จำเป็นต้องอาศัยแหล่งข้อมูลจากคลังข้อมูล (data warehouse) เป็นหลัก ซึ่งการออกแบบคลังข้อมูล มีอยู่ด้วยกัน 3 แบบ เช่น คลังข้อมูลแบบ Star Schema หรือ Multidimensional Schema คลังข้อมูลแบบ Relational Schema และ Snowflake Schema ดังนั้น Business Intelligence ส่วนใหญ่จะนิยมใช้คลังข้อมูลแบบ Star Schema เป็นฐานข้อมูล

กสุภพ ประศาสน์ศรีสุภาพ และเกียรติชุมพล สุทธิศิริกุล (2552) ศึกษาเรื่อง *ระบบฐานข้อมูล Data Warehouse* ซึ่งเป็นระบบฐานข้อมูลรูปแบบหนึ่งที่แตกต่างจากฐานข้อมูลทั่วไป (relational database) ระบบฐานข้อมูล Data Warehouse มีการจัดเก็บข้อมูลในลักษณะ Multidimensional ใช้ Microsoft SQL Server 2000 Analysis Service มาทดลองสร้างระบบ Data Warehouse โดยได้ศึกษาการจัดเก็บข้อมูลในลักษณะต่าง ๆ กัน คือ MOLAP, ROLAP และ HOLAP และได้ทดสอบเปรียบเทียบประสิทธิภาพ การจัดเก็บข้อมูลทั้ง 3 แบบ นอกจากนี้พัฒนาโปรแกรมประยุกต์ Client ขึ้นมาติดต่อ Analysis Service ซึ่งพัฒนาขึ้นทั้งที่ทำงานบน LAN และบน Web เพื่อให้ผู้ใช้เข้าถึงฐานข้อมูล Data Warehouse ได้สะดวกและยืดหยุ่นมากยิ่งขึ้น

กานต์ ผู้ภักดี และหิรัณย์ อโนวรรณพันธ์ (2552) ศึกษาเรื่อง *คลังข้อมูลเชิงวัตถุ-สัมพันธ์* ซึ่งในปัจจุบันระบบ Data Warehousing ที่ใช้ฐานข้อมูลแบบวัตถุเชิงสัมพันธ์ (object-relational database) ยังมีไม่มากนัก เพราะ Data Warehouse ส่วนใหญ่จะใช้ฐานข้อมูลแบบเชิงสัมพันธ์ (relational database) ซึ่งมีการใช้งานที่ง่าย ฐานข้อมูลแบบวัตถุเชิงสัมพันธ์นั้น จะสนับสนุนการพัฒนาระบบ Data Warehouse ที่ไม่สามารถทำได้เมื่อใช้ฐานข้อมูลแบบเชิงสัมพันธ์และฐานข้อมูลแบบวัตถุ (object-oriented database) ซึ่งฐานข้อมูลแบบวัตถุเชิงสัมพันธ์ ได้มีการรวมเอาข้อดีของฐานข้อมูลแบบเชิงสัมพันธ์และฐานข้อมูลแบบวัตถุเข้าไว้ด้วยกัน เทคโนโลยีแบบ (object-relation) นั้นสามารถที่จะแก้ปัญหาในข้อมูลประเภทโครงสร้าง (slowly changing attributes) ของ Entity ได้ โดยใช้โครงสร้างของข้อมูลซับซ้อนที่เกี่ยวกับเวลา ซึ่งไม่สามารถแก้ได้ ถ้าใช้ฐานข้อมูลแบบเชิงสัมพันธ์ เพราะข้อมูลประเภทโครงสร้างเป็นประเด็นหลักในการพัฒนา Data

Warehousing วิทยานิพนธ์ฉบับนี้ได้พัฒนารูปแบบของระบบ Data Warehousing ที่ใช้ฐานข้อมูลแบบวัตถุเชิงสัมพันธ์เพื่อแก้ปัญหาข้อมูลประเภทโครงสร้างและปัญหาความซ้ำซ้อนของข้อมูล ซึ่งไม่สามารถแก้ไขได้โดยฐานข้อมูลแบบเชิงสัมพันธ์ หรือฐานข้อมูลแบบวัตถุ

จุฑามาศ นันท์ธนะวานิช (2552) ศึกษาเรื่อง *การออกแบบคลังข้อมูลสำหรับการไฟฟ้าส่วนภูมิภาค* ทำการสร้างฐานข้อมูล Metadata เพื่อควบคุมการทำงานของคลังข้อมูลให้อยู่ในกรอบ โดยออกแบบให้รองรับการทำงานที่ต่อเนื่องกันเป็นชุด เพื่อให้โปรแกรมสามารถทำงานได้เองโดยใช้คนคุมให้น้อยที่สุด การวิจัยครั้งนี้ใช้ข้อมูลที่เกี่ยวกับรายได้และค่าใช้จ่ายของการไฟฟ้าส่วนภูมิภาคจากระบบงานหลัก 4 ระบบ คือ ระบบบัญชีและการเงิน ระบบบริการผู้ใช้ไฟ ระบบพัสดุคงคลัง และระบบบริหารบุคคล ซึ่งจะใช้ข้อมูลจากระบบบุคลากรโดยที่ข้อมูลจากระบบงานเหล่านี้ มีการจัดเก็บในรูปของแฟ้มข้อมูล ข้อมูลที่คัดเลือกเป็นข้อมูลนำเข้าคลังข้อมูลนี้ได้จากการสัมภาษณ์ การสำรวจจากรายงานที่เสนอผู้บริหารของการไฟฟ้าส่วนภูมิภาค และโครงสร้างแฟ้มข้อมูลจากระบบงานหลักทั้ง 4 ระบบ การพัฒนาคลังข้อมูลมี 4 ส่วน คือ การพัฒนาฐานข้อมูล Metadata การนำข้อมูลเข้าคลัง การพัฒนาฐานข้อมูลสำหรับคลังข้อมูล และการเรียกใช้ข้อมูลจากคลังข้อมูล ใช้หลักการการเชื่อมโยงรูปแบบดาวในการออกแบบการเชื่อมโยงข้อมูลในคลังข้อมูล เพื่อรองรับการวิเคราะห์จำแนกตามมิติ การเรียกใช้ข้อมูลเป็นตัวอย่างโปรแกรมสำหรับทดสอบคลังข้อมูล โดยพัฒนาให้สามารถตอบคำถามเชิงสรุปเกี่ยวกับข้อมูลย้อนหลังและคำถามเชิงเปรียบเทียบเพื่อการวิเคราะห์ได้เป็นอย่างน้อย

สุมนา ตัณฑเศรษฐี และปัณณวิชญ์ วงศ์วิวัฒนานนท์ (2551) ได้ศึกษาเกี่ยวกับระบบสารสนเทศเพื่อการบริหารสำหรับการให้บริการทางกายภาพบำบัด โรงพยาบาลราชวิถี โดยการจัดเก็บข้อมูลทางสถิติผู้ป่วยรายวันของงานกายภาพบำบัด กลุ่มงานเวชศาสตร์ฟื้นฟู โรงพยาบาลราชวิถี นักกายภาพบำบัดจะเก็บข้อมูลโดยการบันทึกด้วยลายมือในสมุดบันทึกประจำตัว ซึ่งหากต้องการข้อมูลทางสถิติจะต้องใช้กำลังคน และเวลานานในการรวบรวม สำหรับเจ้าหน้าที่ธุรการนั้นจะลงบันทึกด้วยคอมพิวเตอร์ในระบบฐานข้อมูลทั่วไป (operational database) ทำให้ข้อมูลเกิดความซ้ำซ้อนมาก และใช้เวลานานในการเรียกดูข้อมูล เนื่องจากบางข้อมูลมีขนาดใหญ่

จากเหตุผลดังกล่าวจะเป็นอุปสรรคต่อการดำเนินการต่าง ๆ ได้ เช่น การตัดสินใจของผู้บริหาร การนำข้อมูลมาวิเคราะห์หรือทำวิจัยเพื่อพัฒนากระบวนการทำงานต่าง ๆ ด้วยเหตุนี้ การพัฒนาระบบสารสนเทศ เพื่อการบริหารสำหรับการให้บริการทางกายภาพ-บำบัดในครั้งนี้ จึงเอาแนวคิดเรื่องคลังข้อมูล (data warehouse) ทางด้านสุขภาพมาใช้ เพื่อให้เกิดความยืดหยุ่นของข้อมูลมากกว่าระบบฐานข้อมูลทั่วไป การนำมาประมวลผลโดยใช้ระบบ OLAP (Online Analytic Processing) สะดวกและสามารถสืบค้นได้ทันทีผ่านระบบออนไลน์ ข้อมูลที่ได้จะเป็นลักษณะ Multi-Dimention Database ทำให้เลือกช่วงข้อมูลและสร้างความสัมพันธ์ระหว่างข้อมูลต่าง ๆ ต่อมาเป็นการติดตั้งแม่ข่าย-คลังข้อมูล ที่ตอบสนองการสืบค้นข้อมูลแบบออนไลน์ได้ในทันทีทันใด โดยใช้ Microsoft OLAP Server ร่วมกับ Microsoft SQL Server 2000 Person Edition ซึ่งสามารถติดตั้งใช้งานได้ในเครื่องคอมพิวเตอร์ระบบ Window XP ทั่วไป ส่วนการสร้างโมเดล การสืบค้นนั้นจะใช้โมเดลที่ชื่อว่า Cube และสุดท้ายเป็นการออกแบบและสร้างโปรแกรมที่เรียกใช้งานและสืบค้นข้อมูลได้ง่าย โดยเลือกใช้ Microsoft Front Page ในการสร้างเว็บแอพพลิเคชั่นโดยใช้เพียงโปรแกรมอ่านเว็บ (web browser) จึงเป็นข้อได้เปรียบตรงที่บุคคลทั่วไปที่เข้าใจระบบคอมพิวเตอร์พื้นฐานและไม่ซับซ้อนก็สามารถใช้งานได้ง่าย ผลการศึกษานี้ทำให้ได้ระบบสารสนเทศวิธีใหม่ เพื่อพัฒนาการทำงานด้านข้อมูล คือ ระบบคลังข้อมูล (data warehouse) และระบบ OLAP สำหรับการให้บริการทางกายภาพ-บำบัด ซึ่งยังไม่เคยมีหน่วยงานกายภาพบำบัดแห่งใดจัดทำมาก่อน แก้ปัญหาความซ้ำซ้อนของการบันทึกข้อมูลและแสดงผลรายงาน การให้บริการทางกายภาพบำบัดที่ถูกต้องได้อย่างสะดวก รวดเร็ว การใช้งานทำได้ง่าย ข้อมูลสารสนเทศที่ได้มีมาตรฐาน ที่ผู้บริหารสามารถนำไปใช้ประกอบในการตัดสินใจบริหารองค์กรได้

อมร กั่งเม่ง (2550) ศึกษาเรื่อง *ระบบคลังข้อมูลการบริการผู้ป่วยใน* โดยทำการออกแบบระบบคลังข้อมูลการบริการผู้ป่วยใน โดยใช้แนวคิดของ Multidimention ประกอบด้วย ตารางหลัก (fact table) และตารางมิติ (dimention table) ตารางหลักมีคีย์ของตารางมิติและข้อมูลที่วัดปริมาณได้ ตารางมิติมีข้อมูลลักษณะที่สนใจในการวิเคราะห์ การออกแบบ Database Model ใช้แบบ Star Schema และ Snowflake Schema โดยใช้ฐานข้อมูล Microsoft SQL Server 2000 และซอฟต์แวร์ Business Intelligent Tools

ชื่อ Cognos Power Play เชื่อมโยงกับ OLAP Cube จัดการกับข้อมูลให้อยู่ในรูปแบบ Multidimentional Analysis ซึ่งเหมาะสำหรับการวิเคราะห์ข้อมูลในมุมมองต่าง ๆ และแสดงผลเป็นรูปแบบกราฟได้หลายแบบ ช่วยให้ตอบคำถามของผู้ใช้งานได้ตามความต้องการโดยไม่ต้องทำรายงานที่ซ้ำซ้อน และพบว่า ระบบคลังข้อมูลการบริการผู้ป่วยมีการนำเสนอรายงานเพื่อการวิเคราะห์ที่มีความยืดหยุ่นต่อผู้ใช้งาน ซึ่งสามารถเลือกดูข้อมูลตามมิติที่ต้องการได้ง่าย ซึ่งสามารถ Drill-Down ดูรายละเอียดข้อมูลและ Roll-Up เพื่อดูข้อมูลยอดสรุปได้อย่างสะดวก รวดเร็ว มีการนำเสนอข้อมูลในรูปแบบที่เข้าใจได้ง่าย ทำให้มีข้อมูลการวิเคราะห์ เพื่อประกอบการตัดสินใจเพื่อใช้ในการบริหารงานโรงพยาบาลให้มีประสิทธิภาพมากยิ่งขึ้น

พรทิพย์ พงษ์สวัสดิ์ (2550) ศึกษาเรื่อง *ระบบคลังข้อมูลสำหรับสนับสนุนการตัดสินใจและติดตามการใช้งบประมาณ* มีวัตถุประสงค์เพื่อพัฒนาระบบคลังข้อมูลเพื่อสนับสนุนการตัดสินใจและติดตามการใช้งบประมาณ เพื่อรวบรวมข้อมูลที่ต่างกันมาไว้ที่เดียวกัน ระบบสามารถคัดกรองข้อมูลที่สำคัญและวิเคราะห์ข้อมูลตามความต้องการได้อย่างรวดเร็ว ผู้ใช้โปรแกรมสามารถเลือกและกำหนดเงื่อนไขต่าง ๆ และเข้าถึงข้อมูลได้เอง ในรูปแบบ Data Cube และเรียกดูรายงานต่าง ๆ ผ่าน Web Browser

ระบบนี้พัฒนาขึ้นโดยใช้โปรแกรม Microsoft SQL Server 2005 เป็นตัวที่ใช้จัดการฐานข้อมูลใช้ Microsoft SQL Server 2005 Reporting Service ในการแสดงผลรายงานและใช้ระบบปฏิบัติการ Microsoft Window 2003 Server ในการทดสอบได้นำโปรแกรมติดตั้งและใช้งาน ทดลองใช้งานกับผู้เชี่ยวชาญและผู้ใช้งานข้อมูลที่ได้จากระบบเป็นประโยชน์ต่อผู้บริหารและผู้ใช้งานที่สามารถเลือกสรรข้อมูล ที่ได้เพื่อนำไปใช้สนับสนุนการตัดสินใจได้อย่างมีประสิทธิภาพ ผลการประเมินคุณภาพและความพึงพอใจหลังการทดลองระบบ พบว่าอยู่ในระดับดี (ค่าเฉลี่ยเท่ากับ 4.05 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.62) สรุปได้ว่าสามารถนำระบบคลังข้อมูลสำหรับสนับสนุนการตัดสินใจและติดตามการใช้งบประมาณ ไปใช้ตามวัตถุประสงค์ได้อย่างเหมาะสม

วีรอร อุดมพันธ์ (2549) ศึกษาเรื่อง *ระบบคลังข้อมูลเพื่อการจัดทำระบบรายงานสำหรับการบริหารงบประมาณ* พัฒนาขึ้นเพื่อนำเสนอข้อมูลด้านงบประมาณที่มีอยู่ในองค์กรเพื่อให้เกิดประโยชน์สูงสุด ซึ่งข้อมูลที่ใช้ในการวิเคราะห์และแสดงผล นำมาจาก

ฐานข้อมูลงบประมาณ พัสดุ การเงินและบัญชีกองทุน โดยเกณฑ์พึงรับ-พึงจ่าย ลักษณะ 3 มิติ มหาวิทยาลัยรามคำแหง นำมาจัดทำเป็นรายงานสรุปในด้านต่าง ๆ ที่แสดงผลในรูปแบบหลากหลายมุมมอง ซึ่งรายงานที่ได้จะเป็นรายงานเพื่อใช้ประกอบการตัดสินใจในการบริหารจัดการงบประมาณของมหาวิทยาลัยรามคำแหง

ระบบนำเสนอรูปแบบรายงานสำหรับการบริหารจัดการงบประมาณเป็นหลัก ซึ่งพัฒนาขึ้นโดยใช้ภาษา ASP และ Microsoft SQL Server 2000 เป็นระบบจัดการฐานข้อมูล และใช้ Data Transformation Service เป็นเครื่องมือในการนำข้อมูลเข้าสู่ระบบ หลังจากนั้นได้ใช้ Microsoft Analysis Service ในการวิเคราะห์ข้อมูลและจัดทำลูกบาศก์ข้อมูลหลายมิติ เพื่อนำมาแสดงผลเป็นรายงานโดยใช้ Pivot Table ของ Microsoft Excel แล้วนำรายงานที่ได้ไปแสดงที่หน้าเว็บเพจ การพัฒนาระบบนี้เป็นทางเลือกใหม่ ได้นำเทคโนโลยีที่ทันสมัยเข้าช่วยในการบริหารจัดการงบประมาณและสามารถจัดทำรายงานเพื่อตอบสนองต่อความต้องการได้รวดเร็วยิ่งขึ้น

ศิรินุช เทียนรุ่งโรจน์ (2548) ได้ศึกษาเกี่ยวกับ ระบบคลังข้อมูลเพื่อการบริหารมหาวิทยาลัย โดยพัฒนาระบบคลังข้อมูลเพื่อการบริหารมหาวิทยาลัย เพื่อนำไปสู่การพัฒนาระบบสารสนเทศเพื่อสนับสนุนการตัดสินใจ ในการจัดการระบบสารสนเทศโดยทั่วไปอาจแบ่งออกเป็น 2 กลุ่มหลัก คือ ระบบสารสนเทศ เพื่อการสนับสนุนการปฏิบัติงานขององค์กร (Transaction Processing--TP) เช่น ระบบการลงทะเบียนเรียน ระบบการจ่ายเงินเดือนและระบบสารสนสารสนเทศเพื่อสนับสนุนการบริหารงานองค์กร (Management Information System--MIS) เช่น ระบบภาระงานสอน เป็นต้น

ระบบคลังข้อมูล (data warehouse) เมื่อกลับมาวิเคราะห์ปัญหาต่าง ๆ ที่ระบบ SUPREME ในปัจจุบันประสบอยู่ พบว่า ปัญหาส่วนหนึ่งมาจากการพัฒนาระบบที่เริ่มต้นด้วยการสร้างระบบเพื่อสนับสนุนการปฏิบัติงานขององค์กร (TP) ดังนั้น ระบบฐานข้อมูล SUPREME จึงถูกออกแบบมาเพื่อจัดเก็บข้อมูลเป็นปัจจุบัน เพื่อใช้ในการปฏิบัติงานรายวัน ซึ่งข้อมูลเหล่านั้นจะมีการเคลื่อนไหวตลอดเวลา เมื่อถูกร้องขอสารสนเทศ เพื่อนำไปใช้ในการสนับสนุนการบริหารงานของผู้บริหาร ที่มักมีรูปแบบไม่แน่นอน และต้องการข้อมูลที่เกิดขึ้นในช่วงเวลาต่าง ๆ ที่ผ่านมา จึงมักเกิดปัญหา เพราะระบบฐานข้อมูล SUPREME ไม่ได้ถูกออกแบบไว้ล่วงหน้า เช่น ต้องการจำนวนนิสิต

ที่ลงทะเบียนเรียนในวิชาต่าง ๆ ก่อนการเพิ่มถอนรายวิชาและหลังเพิ่มถอนรายวิชาในแต่ละภาคการศึกษา เพื่อนำมาใช้ในการวางแผนการจัดห้องเรียน เป็นต้น นอกจากนี้ พบว่า ปัญหาจากความต้องการข้อมูลภาครัฐก็ยังส่งผลต่อความต้องการระบบฐานข้อมูลในมิติ (dimension) ที่หลากหลายด้วย

ปัญหาเหล่านี้ จึงนำไปสู่แนวคิดเรื่องระบบคลังข้อมูล ที่สามารถนำไปใช้เพื่อสร้างระบบสารสนเทศเพื่อสนับสนุนการบริการองค์กรต่อไป ระบบคลังข้อมูลนับเป็นฐานข้อมูลขนาดใหญ่ขององค์กร ที่เก็บรวบรวมข้อมูลมาจากระบบฐานข้อมูลรายวัน (operational database) แล้วนำมาผ่านกระบวนการวิเคราะห์และออกแบบระบบเพื่อสร้างเป็นเสมือนฐานข้อมูลที่มีมิติที่หลากหลาย (data warehouse หรือ dimensional database) เพื่ออำนวยความสะดวกในการสืบค้นข้อมูลให้แก่ผู้ใช้ได้ ในกระบวนการสร้างระบบคลังข้อมูล (data warehousing) นั้นจำเป็นต้องนำข้อมูลจากระบบฐานข้อมูล SUPREME และข้อมูลจากภายนอก (external data) มาผ่านกระบวนการ ที่เรียกว่า ETL (Extract Transform and Loading) & Cleaning เพื่อให้ได้คลังข้อมูลที่สมบูรณ์พร้อมใช้งานต่อไป แต่ในการเรียกใช้ข้อมูลนั้นมักพบว่า ต้องนำเสนอข้อมูลตามกลุ่มข้อมูลหรือตามงานซึ่งจำแนกออกเป็นมิติ (dimension) ต่าง ๆ ได้ เช่น เป็นรายภาคการศึกษา รายปีการศึกษา ตามแผนงานและโครงการ เป็นต้น ดังนั้น เพื่อตอบสนองความต้องการข้อมูลที่หลากหลายเช่นนี้ จึงต้องมีการสร้างระบบฐานข้อมูลตามงาน (subject data mart) ขึ้นมาอีกชั้นหนึ่ง เพื่อให้ผู้ใช้สามารถเรียกใช้สารสนเทศจากระบบคลังข้อมูลได้สะดวกและคล่องตัว เช่น ฐานข้อมูลนิสิต (Student Data Mart--SDM) ฐานข้อมูลงานการเงิน (Finance Data Mart--FDM) ฐานข้อมูลบุคลากร (Human Resource Data Mart--HRDM) เป็นต้น

ซึ่งสรุปเป็นภาพรวมของการพัฒนาระบบคลังข้อมูล ทำให้ได้รับผลพลอยได้ คือ ในกระบวนการสร้างระบบคลังข้อมูลขององค์กร คือ ผู้ใช้ได้รับข้อมูลอีกประเภทหนึ่งที่เรียกว่า Metadata Repository ซึ่งเป็นข้อมูลเกี่ยวกับระบบข้อมูลทั้งหมดขององค์กรที่จัดเก็บอยู่ในระบบฐานข้อมูลของระบบ SUPREME และในระบบคลังข้อมูลของระบบ SUPREME-Plus ซึ่งเป็นระบบสารสนเทศเพื่อสนับสนุนการบริหารงานของผู้บริหารและเพื่อสนับสนุนการประสานงานข้อมูลกับหน่วยงานภาครัฐ

คันธรส ชำนาญกิจ (2544) ศึกษาเรื่อง *การพัฒนาคลังข้อมูลสำหรับงานสถิติ: กรณีศึกษาข้อมูลการสำรวจภาวการณ์ทำงานของประชากร* โดยการพัฒนาคลังข้อมูลประกอบด้วย 4 ขั้นตอน คือ (1) การศึกษาและวิเคราะห์เพื่อทำความเข้าใจระบบเดิม (2) ดำเนินการออกแบบ ซึ่งประกอบด้วย การกำหนดตัววัดค่าเชิงปริมาณที่ต้องการทราบ (measure) จำแนกและวิเคราะห์ลำดับขั้นของสิ่งหรือมิติที่สัมพันธ์กับตัววัดค่า (dimension) (3) ออกแบบตารางหลักและสร้างแบบจำลองแสดงความสัมพันธ์ระหว่างตารางหลักและตารางที่จัดเก็บมิติ และ (4) เตรียมข้อมูลเข้าไปยังคลังข้อมูล โดยใช้ข้อมูลจากการสำรวจภาวะการทำงานของประชากรเป็นข้อมูลตัวอย่าง ผลการวิจัยพบว่า ผู้ใช้สามารถเข้าถึงข้อมูล่ในคลังข้อมูลได้อย่างรวดเร็ว ยืดหยุ่นมากขึ้นและยังสามารถใช้ข้อมูลเชิงวิเคราะห์เพื่อช่วยในการตัดสินใจ

เพ็ญณี หวังเมธีกุล (2542) ได้ศึกษาเกี่ยวกับการพิจารณาเลือกใช้ชุดเครื่องมือในการพัฒนาระบบสารสนเทศเชิงวิเคราะห์ โดยทำการวิเคราะห์และออกแบบระบบคลังข้อมูล เพื่อพัฒนาระบบสารสนเทศสำหรับผู้บริหารระดับสูงของสถานศึกษา โดยพิจารณาความสัมพันธ์ระหว่างปัจจัยพื้นฐานที่จำเป็นต่อการบริหาร ได้แก่ งบประมาณ นักศึกษาและบุคลากร คลังข้อมูลที่ออกแบบใช้โครงสร้างสตาร์ (star schema) โดยใช้ระบบจัดการฐานข้อมูล Oracle Developer 2000 และการแสดงผลข้อมูลผลลัพธ์โดยใช้โปรแกรม Pilot Decision Support Suit

งานวิจัยในต่างประเทศ

Schubart and Einbinder (2000) ศึกษาเกี่ยวกับ Evalution of a Data Warehouse in an Acadamic Health Sciences Center เป็นการประเมินคลังข้อมูลของศูนย์วิชาการ-วิทยาศาสตร์สุขภาพ แห่งมหาวิทยาลัยเวอร์จิเนีย เพื่อประเมินว่าคลังข้อมูลนั้นทำงานได้อย่างที่ได้มีการออกแบบจากผู้รับบริการหรือไม่ และผลลัพธ์เป็นไปตามที่ต้องการหรือไม่ โดยอาศัยทฤษฎีการกระจายนวัตกรรมของ Rogers คณะผู้วิจัยได้ทำการประเมิน CDR โดยการสำรวจด้วยแบบสอบถามและการสัมภาษณ์ผู้นำระดับบริหารสำคัญ ๆ ผลการวิจัยชี้ถึงปัจจัยที่มีอิทธิพลต่อการตัดสินใจ ขั้นต้นที่จะใช้แหล่งข้อมูลทดสอบผลกระทบของช่องทางการสื่อสารและเน้นถึงประเด็นสำคัญ ๆ ที่เป็นตัวบ่งชี้การใช้

คลังข้อมูลด้านสุขภาพและความสำเร็จของคลังข้อมูลสุขภาพ ผลการศึกษาพบว่า ผู้รับบริการ CDR 69% เคยส่งคำถามโดยระบบ CDR ซึ่งแสดงให้เห็นว่า ผู้รับบริการบางส่วนแม้จะมี ID สำหรับใช้ CDR แต่ก็ไม่เคยใช้ระบบเลย จำนวนผู้รับบริการ 26% ใช้ระบบ CDR อย่างไรก็ดีผลการศึกษาบ่งชี้ว่า ความสัมพันธ์ระหว่างเครือข่ายทางสังคมและการตัดสินใจใช้ระบบ CDR ไม่มีนัยสำคัญทางสถิติ

นอกจากนี้ผลการศึกษายังบ่งชี้ว่า ผู้รับบริการส่วนใหญ่ต้องการฝึกอบรมก่อนการใช้และต้องการเวลาทำความคุ้นเคย นอกจากนี้ปัจจัยอื่น ๆ ที่มีผลต่อการใช้ระบบ CDR คือ คุณภาพของข้อมูล การเข้าใช้ข้อมูลลับบางอย่าง เช่น ข้อมูลลับผู้ป่วย ยังมีความยุ่งยาก์ อย่างไรก็ดีผู้รับบริการ 74% เชื่อว่าระบบ CDR จะเป็นระบบสำคัญในอนาคต โดยรวมแล้ว ผู้รับบริการสนใจที่จะเรียนรู้ระบบ CDR และสนับสนุนให้มีการพัฒนาระบบ CDR นี้ต่อไป

Brandas (2007) ได้ศึกษาวิจัยเรื่อง *การนำเสนอระบบสนับสนุนการตัดสินในการบริหารจัดการงบประมาณ* โดยนำเทคโนโลยีสำหรับปัญญาทางธุรกิจ เช่น OLAP คลังข้อมูลและเหมืองข้อมูล (data mining) เป็นเครื่องมือที่ใช้ในการวางแผน และบริหารจัดการภายในองค์กร นำมาใช้ในการรวบรวมและวิเคราะห์ข้อมูลที่เกี่ยวกับด้านงบประมาณ โดยงานวิจัยนี้ได้นำเสนอโมเดลที่ช่วยในการตัดสินใจ คือ Decision Support System Model (B-Admin) ผลการวิจัยพบว่า ระบบสนับสนุนการตัดสินใจในการบริหารจัดการงบประมาณ สามารถช่วยให้ผู้บริหารสามารถนำข้อมูลที่ได้ไปทำการวิเคราะห์กลยุทธ์และนำมาซึ่งความสำเร็จขององค์กร

Katarina (2007) ศึกษาเรื่อง *การนำเสนอระบบปัญญาทางธุรกิจและบริหาร-จัดการกระบวนการทางธุรกิจทางด้านธุรกรรมทางการเงิน* โดยนำเทคโนโลยีทางด้านคลังข้อมูลและ OLAP มาใช้วิเคราะห์พฤติกรรมของลูกค้า ความต้องการของลูกค้า เพื่อความได้เปรียบทางธุรกิจเหนือคู่แข่งรายอื่น จากผลการดำเนินการ พบว่า ระบบดังกล่าวนำมาช่วยในการตัดสินใจ วิเคราะห์พฤติกรรมลูกค้า การจัดการความเสี่ยง แนวโน้มของตลาด การตอบสนองต่อการเปลี่ยนแปลงของตลาดอย่างรวดเร็วและช่วยเพิ่มคุณภาพในการดำเนินงานให้สูงขึ้น

งานวิจัยข้างต้นสามารถสรุปได้ดังนี้ ระบบคลังข้อมูลที่พัฒนาขึ้นสามารถสร้างระบบคลังข้อมูลได้ตามวัตถุประสงค์ และสามารถเลือกดูข้อมูลตามมุมมองต่าง ๆ ที่ผู้วิจัยได้กำหนดไว้ได้เป็นอย่างดี สามารถเรียกดูข้อมูลได้ถูกต้องและได้ด้วยตนเอง การเลือกดูข้อมูลตามมุมมองต่าง ๆ ทำให้เกิดการเพิ่มคุณค่าของสารสนเทศที่มีอยู่ให้กลายเป็นความรู้ที่สามารถนำมาใช้ให้เกิดประโยชน์ในการดำเนินงานของผู้บริหารมากยิ่งขึ้นด้วย

จากการศึกษางานวิจัยที่เกี่ยวข้อง ระบบคลังข้อมูลมีเครื่องมือที่ให้ผู้ใช้งานสามารถเลือกดูข้อมูลที่ง่ายต่อการใช้งานและมีความหลากหลาย ผู้ใช้สามารถเข้าถึงข้อมูลได้เอง การแสดงข้อมูลในรูปแบบหลายมิติ สามารถตอบสนองการเลือกมุมมองของข้อมูลได้ งานวิจัยที่ได้ทำขึ้นนี้จึงได้นำเอาความสามารถในส่วนนี้มาประยุกต์ใช้เพื่อให้ระบบมีความสมบูรณ์และเกิดประโยชน์ในการใช้งานมากยิ่งขึ้น