การออกแบบและพัฒนาคลังข้อมูล

Data Warehouse Design and Development

สายสุนีย์ เจริญสุข

บทคัดย่อ

ความนิยมของคลังข้อมูลการวิเคราะห์ข้อมูลที่มีการเติบโตอย่างมาก ระบบคลังข้อมูลช่วยให้ผู้บริหารขององค์กรบูรณาการได้รับข้อมูลจากที่แตกต่างกันของแหล่งที่มาและการค้นหาฐานข้อมูลขนาดใหญ่ได้อย่างมีประสิทธิภาพ การสร้างคลังข้อมูลต้องใช้การออกแบบและเทคนิคการใช้งานอย่างสมบูรณ์แตกต่างจากระบบฐานข้อมูล และการพัฒนาระบบคลังข้อมูลเริ่มจากการออกแบบฐานข้อมูล ซึ่งวิธีการหนึ่งเรียกว่า ระเบียบวิธี 9 ขั้นของ Kimball จะเน้นที่การออกแบบจากระบบงานย่อยหรือ ดาต้ามาร์ทของแต่ละระบบงานในองค์กรก่อนจึงนำส่วนย่อยๆ นั้นมารวมเป็นระบบคลังข้อมูลขององค์กรต่อไป ซึ่งจะต้องมีการวิเคราะห์ข้อมูลในคลังข้อมูลหลายมิติ เช่น การเจาะลึก การรวมขึ้น การเฉือนบาง การหมุน การเจาะข้าม และการเจาะทะลุ รวมทั้งมีการออกแบบฐานข้อมูลระบบคลังข้อมูลที่มีโครงสร้าง 2 แบบคือ โครงสร้างแบบดาวและโครงสร้างแบบเกล็ดหิมะ ทั้งนี้กระบวนการหนึ่งที่สำคัญมากในการพัฒนาระบบคลังข้อมูล คือ การนำข้อมูลจากแหล่งข้อมูลเข้าสู่ดาต้ามาร์ทของแต่ละระบบ หรือเรียกว่าการแปลงข้อมูล โดยจะต้องกำหนดการส่งข้อมูล รวบรวมหรือสร้างข้อมูลภายนอก วางแผนและสร้างรูทีนการแปลงข้อมูล จึงตรวจสอบความถูกต้องของข้อมูลที่ได้ก่อนนำเข้าสู่คลังข้อมูล เพื่อให้เป็นข้อมูลที่เหมาะสมที่จะนำไปวิเคราะห์ต่อไป

คำสำคัญ: การออกแบบ, คลังข้อมูล, วิเคราะห์ข้อมูล

Abstract

The popularity of data warehouse for analysis of data has grown tremendously. Data warehousing systems enable enterprise managers to acquire and integrate information from heterogeneous sources and to query very large databases efficiently. Building a data warehouse requires adopting design and implementation techniques completely different from those underlying information systems. And the development of data warehouse systems from the design database. This method is called the ninth stage of the Kimball methodology. It emphasizes on the design of each data mart, small operation system, in the organization and then consolidates them into a data warehouse. This will require a multi-dimensional analysis of the data in the data warehouse as well as Drill Down, Roll up , Slice, Pivot or Rotate , Drill Across and Drill Through. Including database design, data warehousing are two types of structure Star Schema and Snowflake Schema. However, the most important processes of data warehouse development are data extraction, transformation, and loading. It manipulates information from its source into each data mart system. After that we need to define its data map, plan and build the data transformation routine. Finally, the process investigates and verifies all derived data before they are transformed into the data warehouse in order to create the suitable data for the further analysis.

Keywords: design, data warehouse, data analysis

บทนำ

ปัจจุบันองค์กรที่จะประสบความสำเร็จประกอบด้วยปัจจัยหลายอย่าง และปัจจัยหนึ่งคือข้อมูลที่มีอยู่และใช้ประจำวัน จะมีข้อมูลเพิ่มขึ้นทุกๆวันทำให้ผู้บริหารองค์กรต้องเพิ่มความพยายามอย่างมากในการวางแผนและควบคุมการดำเนินงานขององค์กรให้เกิดผลดีที่สุดต่อองค์กรและบรรลุเป้าหมายที่ตั้งไว้ ดังนั้นการบริหารจัดการที่ดีผู้บริหารต้องมีระบบจัดการข้อมูลที่ดีและการใช้ข้อมูลเป็นเครื่องมือสำคัญในการตัดสินใจขององค์กร เพื่อให้สามารถแข่งขันกับคู่แข่งขันได้ ฉะนั้นอาจกล่าวได้ว่าการมีข้อมูลมากทำให้มีโอกาสและมีชัยเหนือคู่แข่งในระดับหนึ่ง แต่ในทางกลับกัน การมีข้อมูลจำนวนมากแต่ขาดการจัดการที่เป็นระบบ จะทำให้เกิดความยุ่งยากในการเข้าถึงเพื่อนำมาใช้ประโยชน์และยังต้องเสียค่าใช้จ่ายในการเก็บรักษาข้อมูลนั้นไว้โดยไม่จำเป็น

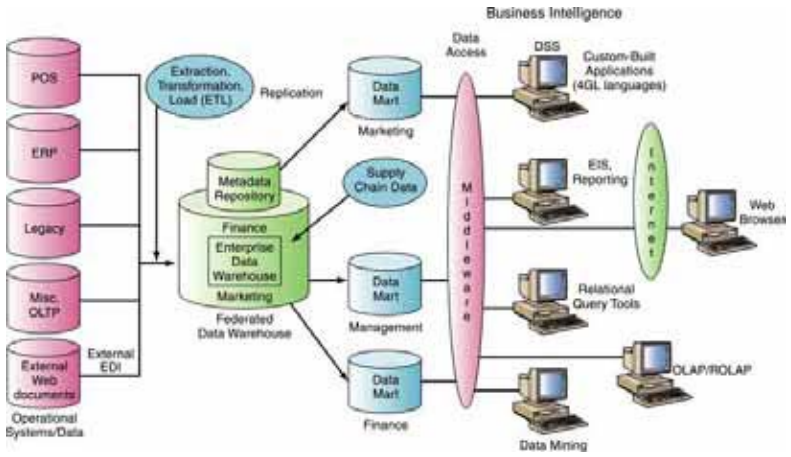
ดังนั้นระบบคลังข้อมูลจะช่วยจัดเก็บข้อมูลจำนวนมากหรือมีขนาดใหญ่เพื่อสามารถนำมาใช้ให้เกิดประโยชน์สูงสุด การที่ระบบคลังข้อมูลจะทำงานได้ดีนั้นต้องเกิดจากการรวบรวมข้อมูลที่ครบรอบด้าน ทุกมิติและควรเก็บข้อมูลในอดีตที่ยาวนานเพียงพอ จึงจะทำให้การพยากรณ์แม่นยำ นั่นคือองค์กรควรมีหน่วยงานทำงานเฉพาะการรวบรวม คัดแยกข้อมูล และสร้างข้อมูลเพื่อนำมาใส่ในระบบ และเป็นผู้ประสานงานจากทุกๆหน่วย เพื่ออำนวยความสะดวกให้กับหน่วยงานต่างๆนำเอาข้อมูลสารสนเทศมาใส่ในระบบ

แนวคิดเกี่ยวกับคลังข้อมูล

1. ความหมายของคลังข้อมูล

คลังข้อมูล (Data Warehouse) หมายถึง ระบบการจัดเก็บ รวบรวมข้อมูลที่ถูกดึงมาจากฐานข้อมูล (database) ที่ได้เก็บรวบรวมรายการต่างๆที่เกิดขึ้นในแต่ละวัน ที่มีอยู่ในระบบปฏิบัติการต่างๆขององค์กร โดยข้อมูลเหล่านั้นจะอยู่กระจัดกระจาย ให้มารวมไว้เป็นศูนย์กลางข้อมูลขององค์กร และสามารถเก็บข้อมูลย้อนหลังได้หลายๆปี เพื่อใช้เป็นข้อมูลสนับสนุนการตัดสินใจ หรือใช้ในการวิเคราะห์ข้อมูล ที่ถูกต้อง และมีประสิทธิภาพ โดยการวิเคราะห์ต้องทำได้แบบหลายมิติ (Multidimensional Analysis) ตลอดจนการวิเคราะห์ทางธุรกิจ เช่น การพยากรณ์(Forecasting) เหมืองข้อมูล (Data Mining)

ดังนั้นจะเห็นว่าข้อมูลที่อยู่ในคลังข้อมูลจะเกี่ยวข้องกับเวลาในอดีตเสมอ แต่อย่างไรก็ตามข้อมูลที่จะนำมาสร้างเป็นคลังข้อมูลต้องถูกนำมาจากฐานข้อมูลที่ใช้อยู่ในช่วงเวลาปัจจุบันแต่อาจต้องมีการปรับเปลี่ยนรูปแบบข้อมูลเพื่อให้มีความเหมาะสมและสะดวกในการนำไปใช้งานด้วย สถาปัตยกรรมของคลังข้อมูลสามารถแสดงได้ดังภาพ 1



ภาพ 1 สถาปัตยกรรมคลังข้อมูล

ที่มา. จาก สถาปัตยกรรมคลังข้อมูล, โดย Studyblue, ค้นจาก https://www.studyblue.com/#file/view/10640500

จากภาพ 1 แสดงถึงสถาปัตยกรรมคลังข้อมูลที่เริ่มต้นด้วยการนำข้อมูลมาจากหลากหลายแหล่งข้อมูล เช่น POS ERP Legacy LegacyMisc.OLTP และ External Webdocuments ผ่านกระบวนการ 3 กระบวนการที่สำคัญดังนี้

1.Extraction คือ การดึงข้อมูลที่ต้องการใช้มาจากแหล่งข้อมูล เช่น ฐานข้อมูลที่ใช้งานอยู่ทั้งภายในและภายนอกองค์กร

2.Transformation คือ เกิดมาจากการที่ข้อมูลจากแหล่งต่างๆอาจจะมีรูปแบบการเก็บข้อมูลต่างกัน แต่ข้อมูลเหล่านั้นหมายถึงสิ่งเดียวกัน การจัดเก็บข้อมูลที่มีความหมายเดียวกันจากแหล่งต่างๆ เข้าสู่คลังข้อมูลจึงจำเป็นต้องแปลงข้อมูลเหล่านั้นให้อยู่ในรูปแบบมาตรฐานเดียวกัน

3.Loading คือ การนำข้อมูลที่ผ่านกระบวนการข้างต้นแล้ว ข้อมูลต่างๆจะอยู่ในรูปแบบที่พร้อมจะนำเข้าสู่คลังข้อมูล เพื่อเตรียมพร้อมใช้งานสำหรับผู้ใช้

2. คุณลักษณะเฉพาะของคลังข้อมูล

ลักษณะของระบบคลังข้อมูล ที่แตกต่างจากฐานข้อมูลปฏิบัติการประกอบด้วย 4 ประการดังนี้ (Inmon, 1994)

1. การจัดกลุ่มตามเนื้อหาของข้อมูล (Subject-Oriented Data) ข้อมูลที่เกิดขึ้นในระดับปฏิบัติการจะมีจำนวนมาก และส่วนใหญ่ไม่ได้นำมาใช้ในการวิเคราะห์หรือช่วยในการตัดสินใจ ดังนั้น คลังข้อมูลสามารถนำมาใช้ในการวิเคราะห์เรื่องใดเรื่องหนึ่ง ข้อมูลในคลังข้อมูลจึงเลือกเก็บข้อมูลในระดับปฏิบัติการเฉพาะส่วนที่นำมาใช้ในเชิงวิเคราะห์หรือเชิงตัดสินใจ มากกว่าการเก็บข้อมูลเพื่อตอบคำถามแบบรายละเอียดปลีกย่อย

2. การรวมข้อมูลเป็นหนึ่งเดียว(Integrated Data) ข้อมูลที่เข้ามาสู่คลังข้อมูลนั้นมีที่มาจากหลายๆแหล่ง ข้อมูลที่มีความหมายเดียวกันแต่อาจจะมีชนิดข้อมูลที่แตกต่างกันไป เช่น จากระบบปฏิบัติงานประจำวันขององค์กร จากระบบสารสนเทศต่างๆ ในองค์กร หรือจากแหล่งข้อมูลภายนอก ดังนั้นเมื่อแหล่งข้อมูลมีมากมายหลากหลาย จึงเป็นไปได้ที่จะเกิดความซ้ำซ้อนของข้อมูลขึ้น หน้าที่ของคลังข้อมูล คือ การกำจัดความซ้ำซ้อนของข้อมูล ทำให้ชนิดของข้อมูลที่แตกต่างกันรวมเป็นชนิดเดียวกัน ทั้งนี้เพื่อไม่ให้เกิดความสับสนในการวิเคราะห์ข้อมูล

3. ความสัมพันธ์กับเวลา (Time Referenced Data) ข้อมูลในคลังข้อมูลแตกต่างจากข้อมูลในระบบปฏิบัติการตรงที่ ในข้อมูลปฏิบัติการจะเก็บเฉพาะข้อมูลในปัจจุบัน ซึ่งมีผลต่อการดำเนินธุรกิจขององค์กร ณ ขณะนั้น แต่ระบบคลังข้อมูลจะมุ่งเน้นไปที่การเก็บข้อมูลเพื่อการวิเคราะห์ข้อมูลตามช่วงเวลา ดังนั้นระบบคลังข้อมูลต้องจัดเก็บข้อมูลในอดีตและปัจจุบันขององค์กร โดยเมื่อนำมาใช้งานระบบจะสามารถแสดงข้อมูลย้อนหลังได้เช่น 3 เดือน 6 เดือน 12 เดือน

4. ความเสถียรของข้อมูล (Non-Volatile Data) เพื่อรักษาคุณสมบัติที่มีความสัมพันธ์กับเวลา ข้อมูลในคลังข้อมูลจึงเป็นข้อมูลที่คงอยู่ตลอดไปไม่ว่าข้อมูลจะเก่าเพียงใด ยังจะอยู่ในคลังข้อมูลตลอด ไม่เปลี่ยนแปลง เรียกคุณสมบัตินี้ว่า ความเสถียรของข้อมูล ทั้งนี้เพื่อให้การวิเคราะห์ข้อมูลแบบช่วงเวลาให้ผลลัพธ์ที่มีประสิทธิภาพ ถึงแม้ข้อมูลจะไม่ถูกลบจากคลังข้อมูล แต่เพื่อการจัดการพื้นที่ในคลังข้อมูลอย่างมีประสิทธิภาพ สามารถถูกเปลี่ยนแปลงรูปแบบไป เพื่อให้ข้อมูลมีขนาดเล็กลงได้ โดยการเปลี่ยนแปลงรูปแบบดังกล่าวไม่กระทบต่อการวิเคราะห์ข้อมูล

3. รูปแบบการวิเคราะห์ข้อมูลในคลังข้อมูล

เนื่องจากข้อมูลทางธุรกิจมีลักษณะเป็นหลายมิติ (Multidimensional) เช่น ข้อมูลยอดขาย มีลักษณะเป็นยอดขายต่อสินค้า มองในด้านยอดขายต่อเวลา มองในด้านสถานที่จำหน่าย เรียกข้อมูลมีหลายมุมมองว่า มิติ (Dimension)ของข้อมูล เช่น สินค้าจะมีมิติขนาดของ

สินค้าและเป็นประเภทสินค้า หรือมิติของเวลาเป็นวันที่ เป็นสัปดาห์ เป็นเดือน และเป็นปี หรือมิติของสถานที่จำหน่ายสินค้า เป็นตำบล เป็นอำเภอ และเป็นจังหวัด เป็นต้น ดังนั้นการวิเคราะห์ข้อมูลหลายมิติ จึงประกอบด้วย

1.การเจาะลึก (Drill Down) หมายถึงการวิเคราะห์ข้อมูลโดยดูข้อมูลเป็นลำดับชั้น ดูข้อมูลจากระดับบนแล้วค่อยๆ ซอยย่อยเจาะลึกลงไปในระดับล่าง เช่น ดูยอดขายประจำปี แล้วเจาะลึกลงไปดูรายเดือน รายสัปดาห์

2.การรวมขึ้น (Roll up) จะตรงกันข้ามกับ การเจาะลึก (Drill Down) คือเริ่มดูข้อมูลจากส่วนรายละเอียดแล้วค่อยดูสรุปรวม หรือการเปลี่ยนระดับชั้นขึ้นหนึ่งระดับเหมือนกับการลดจำนวนมิติ

3.การเฉือนบาง (Slice) และการตัดบางส่วน (Dice) คือการมองข้อมูลจากหลายมุม เช่น การดูยอดขายสินค้าทั้งหมดต่อเดือนแล้วเปลี่ยนมาดูยอดขายต่อจังหวัด

4.การหมุน(Pivot or Rotate) คือขั้นตอนการเปลี่ยนมุมมองของลูกบาศก์วิธีการคือการสลับแกนการแสดงลูกบาศก์ เช่น มองในมุมของเวลา(ปีพ.ศ.) เปลี่ยนมุมมองเป็นมองในมุมของชนิดสินค้า

ตาราง 1

มองในมุมของเวลา (ปี พ.ศ.)

ปี พ.ศ.	ปากกา	สมุด
2551	150000	250000
2552	250000	310000

ตาราง 2

มองในมุมมองชนิดสินค้า

รายการสินค้า	2551	2552
ปากกา	150000	250000
สมุด	250000	310000

5.การเจาะข้าม (Drill Across) คือ การเจาะข้อมูลลงในตารางข้อมูลจริงมากกว่าหนึ่งตาราง

6.การเจาะทะลุ(Drill Through) คือ การเจาะข้อมูลลงไปในระดับที่ต่ำกว่าลูกบาศก์ขั้นต่ำที่สุด โดยส่งไปยังฐานข้อมูลที่เป็นที่มาของข้อมูล หรือการเจาะข้อมูลลงไปในระดับฐานข้อมูล

4. วิธีการออกแบบฐานข้อมูลสำหรับคลังข้อมูล

ปัจจัยหนึ่งที่จะทำให้การสร้างระบบคลังข้อมูลได้สำเร็จจึงต้องรวบรวมข้อมูลของทั้งองค์กรที่ต้องการทั้งหมดมาก่อน เพื่อข้อมูลจะได้สอดคล้องกัน จากนั้นจึงแบ่งข้อมูลเหล่านั้นออกมาเป็นส่วนย่อยๆ เพื่อนำมาสร้างตลาดข้อมูล (Data Mart) เฉพาะเรื่องใดเรื่องหนึ่ง เมื่อทำตลาดข้อมูล (Data Mart) ครบทุกเรื่องแล้วมารวมกันเป็น

ระบบคลังข้อมูล จะได้ระบบคลังข้อมูลที่ใช้งานได้อย่างมีประสิทธิภาพ

วิธีการนี้ถูกเสนอโดย คิมบอล(Kimball) ในปี 1996 เรียกว่าระเบียบวิธี 9 ขั้น หรือ Nine-Step Methodology โดยวิธีการนี้เริ่มจากการออกแบบจากส่วนย่อยที่แสดงถึงแต่ละระบบงานขององค์กร หรือเรียกอีกอย่างหนึ่งว่า ดาต้ามาร์ท (Data Mart) โดยเมื่อออกแบบแต่ละส่วนสำเร็จแล้ว จึงนำมารวมกันเป็นคลังข้อมูลขององค์กร ซึ่งขั้นตอนทั้ง 9 ขั้นตอน(สามารถดูตัวอย่างจากรูปที่ 3 ได้) มีรายละเอียดดังนี้

ขั้นตอนที่ 1 กำหนดดาต้ามาร์ท คือ การเลือกว่าจะสร้างดาต้ามาร์ทของระบบงานใดบ้าง และระบบงานใดเป็นระบบงานแรกโดยองค์กรจะต้องสร้าง E-R Model ที่รวมระบบงานทุกระบบขององค์กรไว้ แสดงการเชื่อมโยงของแต่ละระบบงานอย่างชัดเจน เช่น การเลือกพัฒนาระบบดาต้ามาร์ทด้านการขาย

ขั้นตอนที่ 2 กำหนด fact table ของดาต้ามาร์ท คือกำนหดเนื้อหาหลักที่ควรจะเป็นของดาต้ามาร์ท โดยเลือกเอนทริตีหลักและกระบวนการที่เกี่ยวข้องเอนทริตีนั้นๆ ออกมาจาก E-R Model ขององค์กร เช่น กำหนด face table คือ Sales Face Table และกำหนด Dimension Table คือ time, brance, item และ location

ขั้นตอนที่ 3 กำหนดแอตทริบิวต์ที่จำเป็นในแต่ละ dimension table คือการกำหนดแอตทริบิวต์ที่บอกหรืออธิบายรายละเอียดของ dimension ได้ ทั้งนี้แอตทริบิวต์ที่เป็น primary key ควรเป็นค่าที่คำนวณได้ กรณีที่มีดาต้ามาร์ทมากกว่าหนึ่งดาต้ามาร์ทมี dimension เหมือนกัน นั่นหมายถึง แอตทริบิวต์ใน dimension นั้นจะต้องเหมือนกันทุกประการเช่น ตาราง time จะประกอบด้วยแอตทริบิวต์คือ time_key, day, day_of_the_week, month, quarterและ year เป็นต้น

ขั้นตอนที่ 4 กำหนดแอตทริบิวต์ที่จำเป็นใน fact table โดยแอตทริบิวต์หลักใน fact table จะมาจาก primary key ในแต่ละ dimension table นอกจากนี้แล้ว ยังสามารถมแอตทริบิวต์ที่จำเป็นอื่นๆ ประกอบอยู่ด้วย เช่น แอตทริบิวต์ที่ได้จากการคำนวณค่าเบื้องต้นที่จำเป็นสำหรับการคงอยู่ของแอตทริบิวต์อื่นใน fact table เรียกอีกอย่างหนึ่งว่า measure เช่น ตาราง Sales Fact Table มีแอตทริบิวต์ units_sold, money_soldและ avg_salesเป็น measues

ขั้นตอนที่ 5 จัดเก็บค่าคำนวณเบื้องต้นใน fact table คือการจัดเก็บที่ได้จากการคำนวณให้เป็นแอตทริบิวต์ใน fact table

ขั้นตอนที่ 6 เขียนคำอธิบาย dimension table ทั้งนี้เพื่อให้ผู้ใช้สามารถใช้งานดาต้ามาร์ทได้อย่างมีประสิทธิภาพเพราะเกิดความเข้าใจอย่างดีในส่วนต่างๆ

ขั้นตอนที่ 7 กำหนดระยะเวลาในการจัดเก็บข้อมูลในฐานข้อมูล โดยอาจจะเป็นการจัดเก็บเพียงระยะเวลา 1-2 ปี หรือนานกว่านั้นขึ้นอยู่กับความต้องการขององค์การ

ขั้นตอนที่ 8 การติดตามปัญหาการเปลี่ยนแปลงของ dimension อย่างช้าๆ คือ การเปลี่ยนเอาแอตทริบิวต์ของ dimension table เก่ามาใช้แล้วส่งผลกระทบต่อข้อมูลปัจจุบันของ dimension table โดยสามารถแบ่งประเภทของปัญหาที่เกิดได้ 3 ประเภท ดังนี้

1) เกิดการเขียนทับข้อมูลใหม่โดยข้อมูลเก่า

2) เกิดเรคอร์ดใหม่ๆ ขึ้นใน dimension

3) เกิดเรคอร์ดที่มีทั้งค่าเก่าและใหม่ปนกันไป

ขั้นตอนที่ 9 กำหนดคิวรี่เป็นการออกแบบด้านกายภาพเพื่อให้ผู้ใช้เกิดความสะดวกในการใช้งานและสามารถทำงานได้อย่างมีประสิทธิภาพ

5. การออกแบบฐานข้อมูลระบบคลังข้อมูล

การออกแบบฐานข้อมูลระบบคลังข้อมูล จะใช้ data modeling ที่เรียกว่า Dimensional Modeling เป็นเครื่องมือสำหรับช่วยในการออกแบบโครงสร้างพื้นฐานของระบบฐานข้อมูล การออกแบบ Dimension model ของงาน โอแลป (OLAP) จะทำในลักษณะของลูกบาศก์

(Cube) ซึ่งเปรียบเหมือนกับรูปลูกบาศก์ที่มีมุมมองหลากหลาย แต่ละมุมมองจะทำให้เกิดการสืบค้นข้อมูลได้หลากหลายแบบ ลูกบาศก์ (Cube) ประกอบด้วยส่วนประกอบพื้นฐานที่สำคัญ 2 ส่วนคือ ตารางแกนมิติ (Dimension) และค่าที่ต้องการวัด (Measure) โดยลูกบาศก์ (Cube) มีโครงสร้าง 2 แบบ คือ โครงสร้างแบบดาว (Star Schema) และโครงสร้างแบบเกล็ดหิมะ(Snowflake Schema) มีรายละเอียดดังนี้

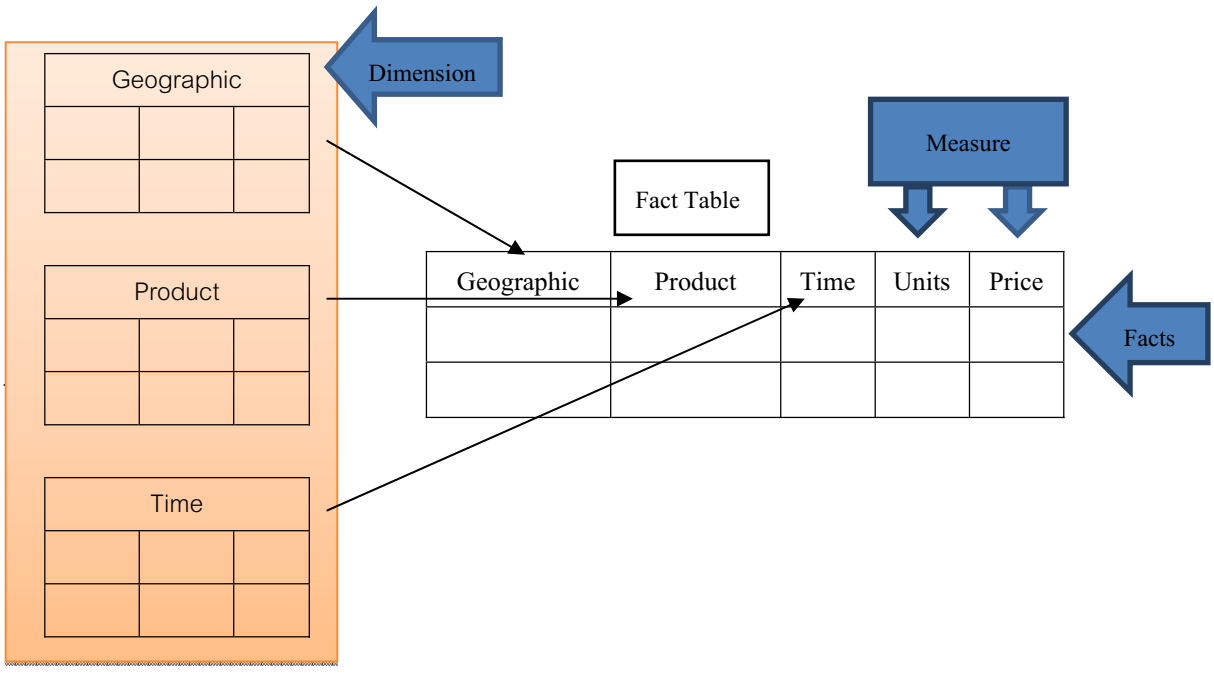
1.ส่วนประกอบของแบบจำลอง Data Warehouse

ก่อนที่จะออกแบบ data warehouse หรือ data mart ต้องศึกษาส่วนประกอบของแบบจำลองพื้นฐานซึ่งประกอบด้วย

ตาราง 3

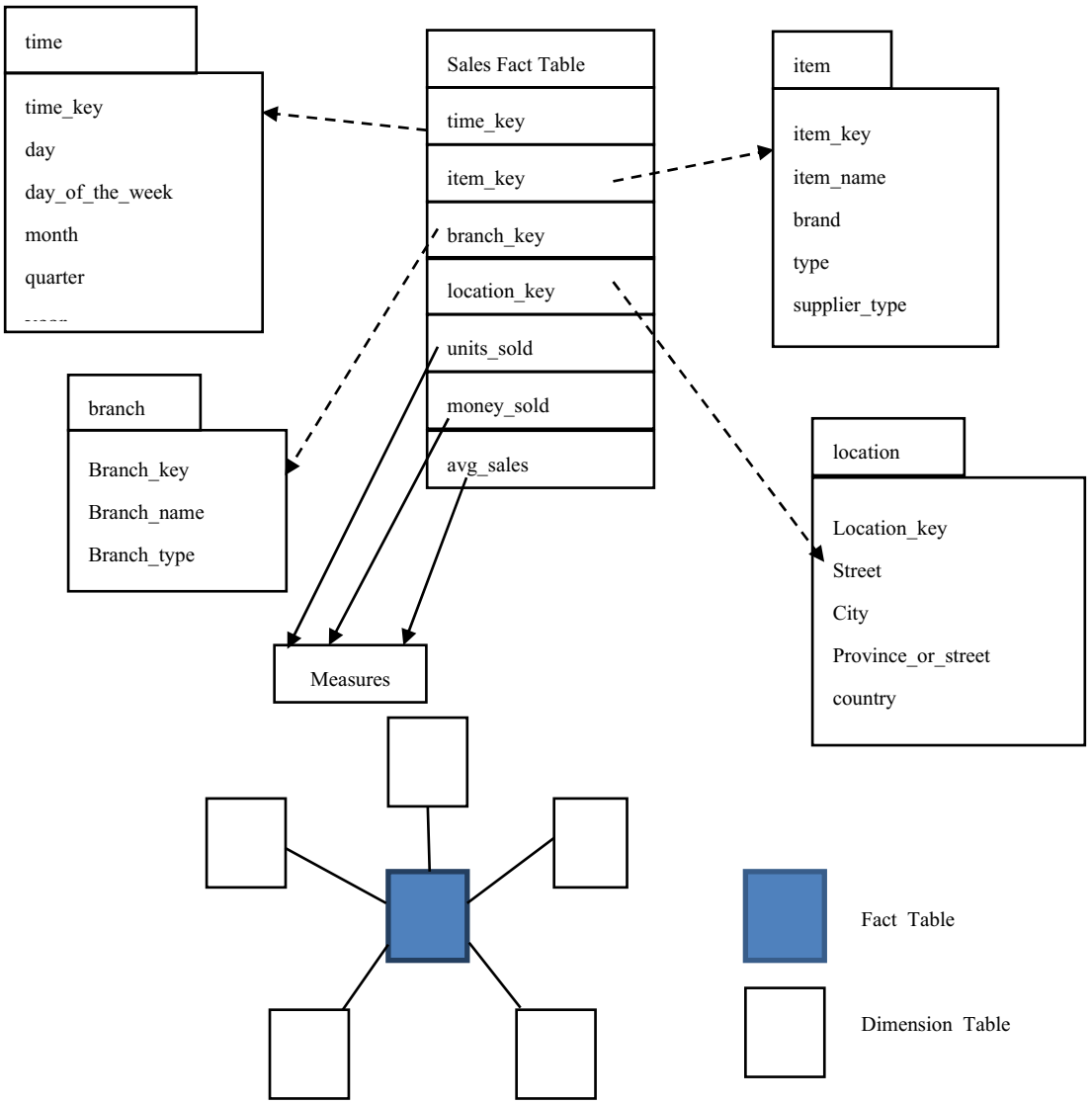
ส่วนประกอบของ Data Warehouse

ส่วนประกอบ	รายละเอียด
Fact table	เป็นตารางศูนย์กลางที่ใช้ใน data warehouse และ data mart ที่จะเก็บจำนวน measures และรายละเอียดสำคัญในเชิงธุรกิจ
Fact	คือ row ในตาราง โดย fact จะเก็บค่าตัวเลขที่ใช้วัดเหตุการณ์ที่เกิดขึ้นกับ data
Measure	คือ ปริมาณ , จำนวน column ใน fact table โดย measure จะแสดงให้เห็นถึงค่าที่ถูกวิเคราะห์แล้ว
Dimension	คือ เอกลักษณ์ทางธุรกิจเป็นลักษณะทางกายภาพ
Dimension table	เป็นตารางใน data warehouse หรือ data mart ที่อธิบายข้อมูลใน fact table



ภาพ 2 ส่วนประกอบของ Data Warehouse Star Schema model

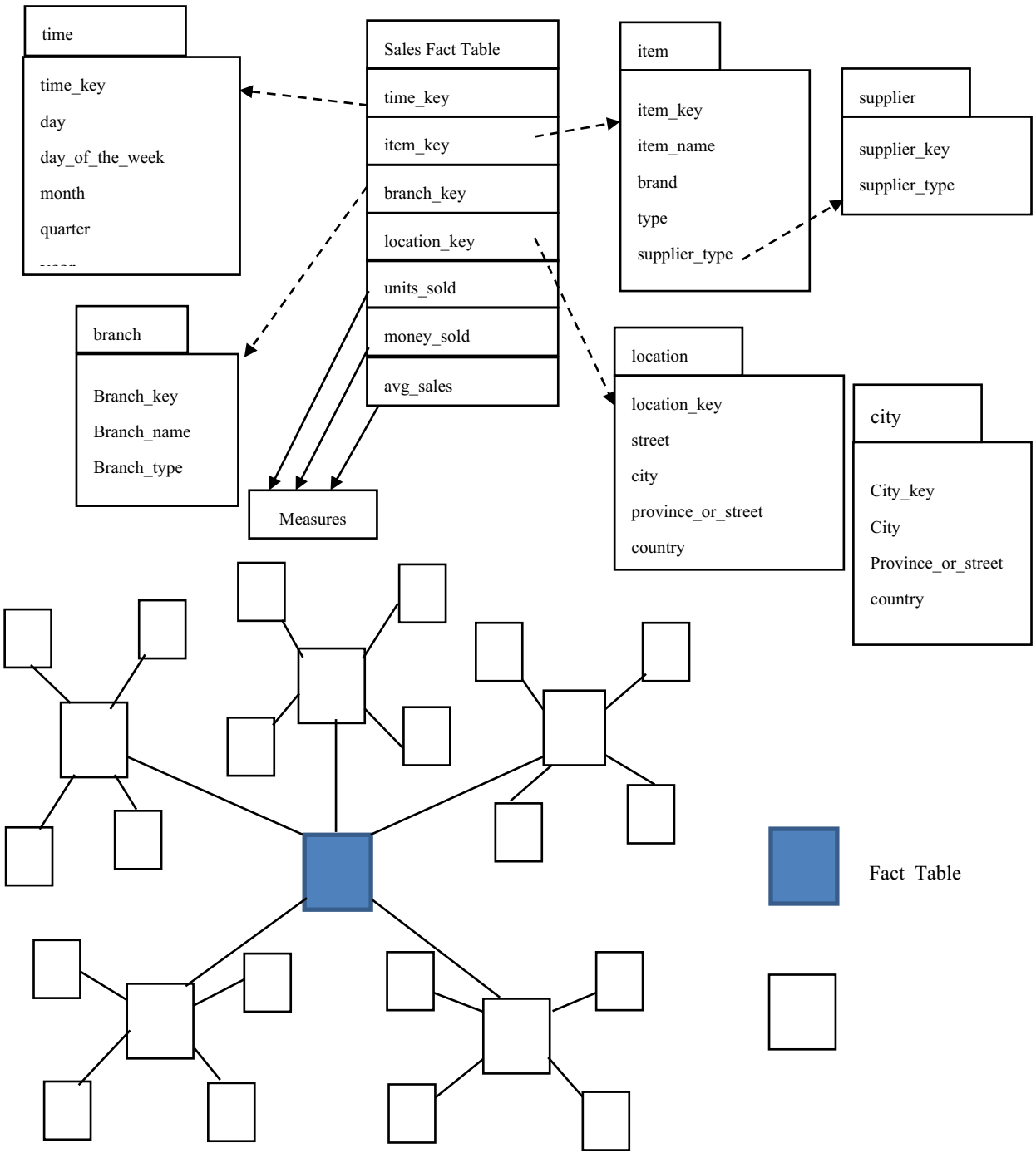
2. โครงสร้างแบบดาว (Star Schema) เป็นเทคนิคที่ใช้ multidimensional โดย data warehouse จะมีตั้งแต่ 1 star schema ขึ้นไปคือแบบจำลองข้อมูลเชิงมิติ ที่มีตารางข้อเท็จจริงขนาดใหญ่เพียงหนึ่งเดียวอยู่ตรงกลาง และมีตารางมิติจำนวนหนึ่งเชื่อมโยงเพื่อกำหนดมุมมองที่จะมีต่อข้อมูลที่ใช้เพื่อการวัดค่าในตารางข้อเท็จจริง โดยจำนวนมุมมองที่มองได้จะเท่ากับจำนวนของตารางมิติที่เชื่อมต่อโดยตรงกับตารางข้อเท็จจริง Star Schema จะประกอบไปด้วย set ของตารางที่ยังไม่ได้ normalized ถ้าใช้ star schema ในการออกแบบฐานข้อมูลเราสามารถเพิ่มประสิทธิภาพของการวิเคราะห์ dimension ใน relational environment สามารถที่จะวิเคราะห์ข้อมูลได้โดยตรงจาก data mart โดยไม่ใช้ OLAP



ภาพ 3 โครงสร้างแบบดาว (Star Schema)

3. โครงสร้างแบบเกล็ดหิมะ(Snowflake Schema)

หมายถึงแบบจำลองข้อมูลเชิงมิติ ที่มีตาราง ข้อเท็จจริงขนาดใหญ่เพียงหนึ่งเดียวอยู่ตรงกลางและมีตาราง
มิติจำนวนหนึ่งเชื่อมโยงอยู่เพื่อกำหนดมุมมองที่จะมีต่อข้อมูลที่ใช้เพื่อการวัดค่าในตารางข้อเท็จจริงนั้นตาราง แกนมิติ (Dimension Table) มีหลายระดับ และมีคีย์ที่โยงไปยัง ตารางแกนมิติ (Dimension Table) อื่นอีก ดังนั้น โครงสร้างแบบนี้จะซับซ้อนมากขึ้น รวมทั้งมีผลให้การการสืบค้นยากขึ้นด้วย ลักษณะของข้อมูลจะมีความเป็น Normalized



ภาพ 4 โครงสร้างแบบเกล็ดหิมะ (Snowflake Schema)

6. ตัวอย่างขององค์กรที่ใช้คลังข้อมูล

6.1 บริษัทเอวอน (Avon) เอวอน เป็นบริษัทจำหน่ายตรงเครื่องสำอางค์ที่ใหญ่ที่สุดในโลก มีสินค้ากว่า 400 ผลิตภัณฑ์ จำหน่ายใน 135 ประเทศ มีรายได้กว่าปีละ 5 พันล้านเหรียญสหรัฐ เอวอนได้เริ่มทดลองโดยสร้างมาร์ทข้อมูลของตลาด โดยภายใน 4 เดือนที่เริ่มใช้งาน ได้ฝึกอบรมผู้ใช้ราว 100 คน และจากการใช้พบว่า มาร์ทข้อมูลสามารถให้ข้อมูลการตลาดที่เป็นประโยชน์และรวดเร็ว และส่งผลต่อการดำเนินงาน จากนั้นเอวอนจึงได้สร้างคลังข้อมูลให้ครอบคลุมทั้งองค์กร ซึ่งส่งผลลัพธ์ได้เป็นที่น่าพอใจ กล่าวคือ ผู้ใช้สามารถดึงข้อมูล และวิเคราะห์ข้อมูลที่ต้องการได้อย่างรวดเร็ว ไม่ต้องเสียเวลาในการออกรายงานต่างๆ ให้กับผู้ใช้เหมือนแต่ก่อน และจากการที่ผู้บริหารตัดสินใจได้ดีขึ้น ทำให้เอวอนมีรายได้เพิ่มขึ้น

6.2 บริษัทเทเลแบรนด์ (Telebrands) เทเลแบรนด์เป็นบริษัทที่จำหน่ายสินค้าตามทีวีที่คุ้นเคยกันดีว่า “As Seenon TV” นั่นเอง เทเลแบรนด์มีสินค้าหลายร้อยรายการ ทำทั้งขายส่งและขายตรง บริษัทได้สร้างมาร์ทข้อมูลการขายและการตลาดสำหรับการขายปลีก และสร้างคลังข้อมูลสำหรับการขายส่ง ปัจจุบัน ผู้บริหารของบริษัทสามารถมองเห็นบริษัทในมุมมองใหม่ๆสามารถติดตามการโฆษณาที่ปล่อยออกไป ติดตามร้านค้าปลีก และติดตามตัวแทนขาย เป็นต้น ปัจจุบันผู้บริหารมีความสามารถในการตัดสินใจที่สำคัญๆ คลังข้อมูลทำให้บริษัทมีข้อมูลที่ทันสมัยพร้อมที่จะให้ผู้บริการเรียกดูได้ทุกเวลา

6.3 บริษัท 3เอ็ม (3M) 3M เป็นบริษัทที่ผลิตและจำหน่ายสินค้าหลากหลาย ตั้งแต่เครื่องใช้สำนักงาน เครื่องไฟฟ้า จนถึงอุปกรณ์รักษาความปลอดภัย บริษัทมีสินค้ากว่า 50,000 รายการ มี 50 หน่วยธุรกิจ มีรายได้ 15 พันล้านเหรียญสหรัฐต่อปี และจากพันธกิจของบริษัทที่ว่า “หนึ่งหน้า หนึ่งเสียงของหนึ่งลูกค้า” หรือ “ One face, one voice to everyone in contact with the company” ทำให้ 3M ต้องเข้าใจลูกค้าของตนเป็นอย่างดี บริษัทได้สร้างคลังข้อมูลที่ใหญ่และท้าทายที่สุดโครงการหนึ่งของโลก และผลของโครงการเป็นที่น่าพอใจอย่างมากทำให้บริษัทมีรายได้เพิ่มขึ้น สามารถลดค่าใช้จ่ายด้านเทคโนโลยีสารสนเทศที่ไม่ต้องพัฒนาระบบสารสนเทศหลากหลายที่มักจะซ้ำซ้อน และที่สำคัญช่วยให้พันธกิจของบริษัทประสบผลสำเร็จ คลังข้อมูลช่วยให้บริษัททราบแนวโน้มในการซื้อสินค้าของลูกค้า อุปนิสัยในการจับจ่ายใช้สอย ความชอบพฤติกรรม และการขายข้ามสินค้า (Cross Sale)

บทสรุป

คลังข้อมูล คือฐานข้อมูลขนาดใหญ่ที่รวบรวมข้อมูลทั้งจากแหล่งข้อมูลภายในและภายนอกองค์กร เพื่อสนับสนุนการวางแผน การตัดสินใจและการบริหารงานของผู้บริหาร ทำให้ผู้บริหารสามารถเรียกใช้ข้อมูลที่ต้องการได้อย่างรวดเร็ว สามารถทำการวิเคราะห์ข้อมูลต่างๆ นำมาเพิ่มความสามารถขององค์กรทั้งในด้านการตัดสินใจ การเตรียมพร้อมรับมือกับปัญหาต่างๆ รวมถึงสามารถคาดการณ์ในอนาคต และนำไปวิเคราะห์เพื่อให้องค์กรสามารถบรรลุเป้าหมายได้อย่างมีประสิทธิภาพมากที่สุดซึ่งในมหาวิทยาลัยอีสเทิร์นเอเชียยังมีข้อมูลที่มีความจำเป็นที่ช่วยในการทำงานหลากหลายเช่น คลังข้อมูลเกี่ยวกับการเข้าอบรมของอาจารย์ คลังข้อมูลการผลิตผลงานวิจัย คลังข้อมูลการผลิตผลงานวิชาการในแต่ละปีการศึกษา เป็นต้นถ้ามีการจัดทำคลังข้อมูลดังกล่าวแล้วนั้นจะช่วยวิเคราะห์ด้านความเชี่ยวชาญในการนำมาพัฒนาด้านการเรียนการสอนหรือการบริการวิชาการต่อหน่วยงานภายนอกต่อไปได้

เอกสารอ้างอิง

กิตติพงศ์ กลมกล่อม. (2546). *การออกแบบและพัฒนาคลังข้อมูล*. กรุงเทพฯ: เคทีพี คอมพ์ แอนด์ คอนซัลท์.

ชนวัฒน์ ศรีสอ้าน. (2550). *ฐานข้อมูลคลังข้อมูลและเหมืองข้อมูล*. กรุงเทพฯ: เอเชียดิจิตอล การพิมพ์.

Andy Wicks. (2013). *Data Warehouseing*. ค้นจาก https://www.youtube.com/watch?v=zTsjSXnvs

GirijaNarasimhan. (2513). *Part1-Create Dimension Star Schema*. ค้นจาก https://www.youtube.com/watch?v=IbkccIE5trU

Inmon, W. H. & Rechard, D. H. (1994). *Using the Data Warehouse*. New York: John Wiley & Sons.

Jared Hillam. (2010). *Benefits of a Data Warehouse*. Retrieved from https://www.youtube.com/watch?v=KGHbY_Sales

Michael Deskis. (2012). *Facts and Dimensions*. Retrieved from https://www.youtube.com/watch?v=AkYvngDGTHY

Shirley Becker. (2002). *Data Warehousing Web Engineering.Hershey*. PA: IRM Press.

Studyblue. (ม.ป.ป.). *สถาปัตยกรรมคลังข้อมูล*. ค้นจาก https://www.studyblue.com/#file/view/10640500

Youtube. (2011). *How is a datawarehouse used to make decisions*. Retrieved from https://www.youtube.com/watch?v=reyn-SX17MM

Youtube. (2013). *Data Warehousing-Concepts-Introduction*. Retrieved from https://www.youtube.com/watch?v=VICYl-6hW5U