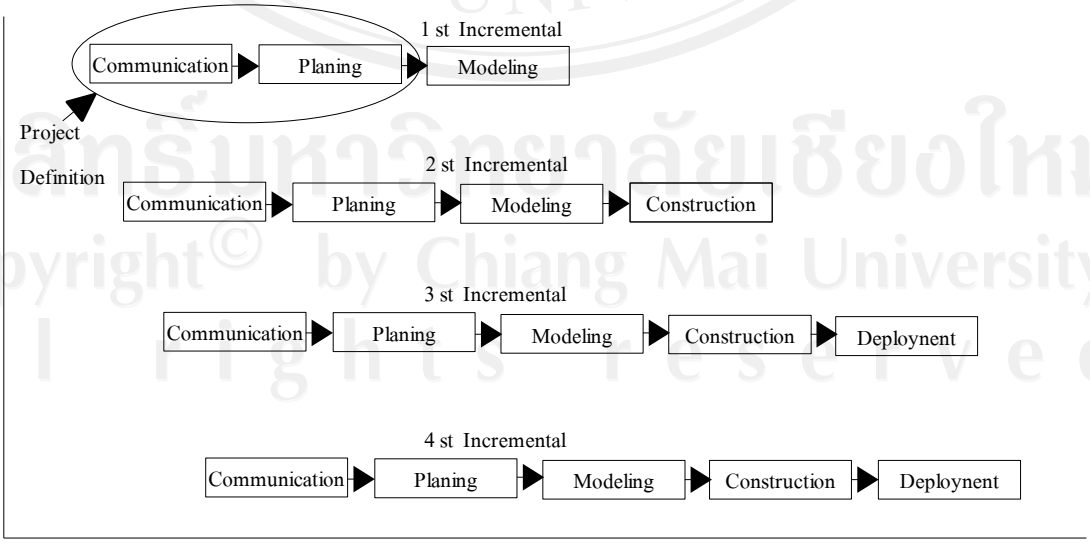
บทที่ 2
ทฤษฎีและเอกสารที่เกี่ยวข้อง

2.1 ทฤษฎีที่เกี่ยวข้อง

2.1.1 การออกแบบเพื่อการจำลองแบบซอฟต์แวร์ (Incremental model)

การออกแบบเพื่อการพัฒนาระบบซอฟต์แวร์แบบประยุกต์ เพื่อนำไปใช้ในงานเฉพาะด้าน (Application Software) มีวงจรการพัฒนาได้หลายรูปแบบ อาทิเช่น วงจรการพัฒนาแบบลำดับขั้นบันได (Waterfall model) แบบอินครีเมนทอลโมเดล (Incremental model) วงจรการพัฒนาแบบลำดับขั้นบันได ถือได้ว่าเป็นแม่แบบของกระบวนการการพัฒนาซอฟต์แวร์(Software development mode) แบบแรกของโลก จึงมีผู้ใช้กันแพร่หลายทั่วโลก แต่พบว่า โครงการพัฒนาซอฟต์แวร์ใหญ่ๆ หลายโครงการที่ใช้แม่แบบ แบบนี้ประสบความล้มเหลวได้ง่าย ปัญหาที่เด่นชัดที่สุด คือ ลูกค้าไม่ทราบว่าตนเองต้องการซอฟต์แวร์แบบไหน และมักจะเปลี่ยนแปลงข้อกำหนดของซอฟต์แวร์บ่อยๆ ในทางปฏิบัติการพัฒนาซอฟต์แวร์ จึงได้มีการประยุกต์และหาแนวทางใหม่ ๆ ที่จะทำให้การพัฒนาซอฟต์แวร์ตรงตามความต้องการของผู้ใช้ มีความยืดหยุ่น สอดคล้องกับวิวัฒนาการของซอฟต์แวร์มากขึ้น สำหรับในการออกแบบที่เป็นการจำลองแบบนั้น จะมีวงจรการพัฒนาระบบซอฟต์แวร์ (SDLC หรือ Software Development Life Cycle) ซึ่งสามารถแสดงได้ดังรูป



รูปที่ 2.1 แสดงขบวนการพัฒนาซอฟต์แวร์แบบอินครีเมนทอล

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

แบบอินครีเมนทอลโมเดล เป็นโครงการการพัฒนาระบบซอฟต์แวร์ โดยจะถูกจัดแบ่งเป็นหลายโครงการย่อย แต่ละโครงการย่อยมีการกำหนดเวลาและแผนงานของตัวเอง เรียกแต่ละโครงการย่อยว่า รอบงาน (Iteration) แต่ละรอบงานจะทำการพัฒนาระบบตามขั้นตอนวิเคราะห์ออกแบบจัดสร้างด้วยตัวเองได้ ผลลัพธ์ที่ได้ในรอบบแรกจะเป็นระบบที่สามารถนำไปประมวลผลได้ แต่ยังไม่ใช่ระบบที่สมบูรณ์ ระบบจะถูกพัฒนาในรอบต่อๆไป และเพิ่มเติมขึ้นตลอดเวลา ตามรอบงานในแต่ละรอบงานที่ทำเสร็จ จนกระทั่งได้ระบบที่สมบูรณ์

อินครีเมนทอลโมเดลสร้างจากข้อเท็จจริงที่ว่า ซอฟต์แวร์ถูกสร้างเพิ่มเติมขึ้นทีละน้อยๆ ให้สมบูรณ์ขึ้น ทั้งในส่วนของรายละเอียดและข้อกำหนดที่มากขึ้น ไปเป็นการออกแบบจนถึงการเขียนโปรแกรม รวมถึงในส่วนของความสามารถในการทำงานซอฟต์แวร์ของตัวต้นแบบ จะยังไม่มีความสมบูรณ์ แต่จะเพิ่มเติมขึ้นในเวอร์ชันถัดไป ดังนั้น ในอินครีเมนทอลโมเดลถูกมองเป็นผลผลิตจากกระบวนการออกแบบ สร้าง ประกอบ ทดสอบระบบซ้ำหลาย ๆ รอบ โดยแต่ละรอบเป็นการเพิ่มความสมบูรณ์ให้กับซอฟต์แวร์ และเมื่อสิ้นสุดรอบการพัฒนา ก็จะได้ระบบที่มีความสมบูรณ์ พอที่นำไปใช้ได้

หลักการในขบวนการพัฒนาซอฟต์แวร์แบบประยุกต์ ตามรูปแบบของวงจรการพัฒนาระบบซอฟต์แวร์แบบอินครีเมนทอลโมเดล ได้แนวคิดมาจากความพยายามแก้ไขข้อจำกัดของแบบลำดับชั้นบันได ซึ่งได้แก่ การติดต่อสื่อสาร การวางแผน การวิเคราะห์และออกแบบ การพัฒนาต้นแบบซอฟต์แวร์ การนำไปใช้งานและปรับปรุงแก้ไข ซึ่งจะประกอบด้วยส่วน 6 ขั้นตอน ดังนี้

(1) การติดต่อสื่อสาร การติดต่อระหว่างผู้พัฒนาซอฟต์แวร์กับลูกค้าหรือผู้ใช้

(2) การวางแผนและพัฒนาระบบ (System Planning)

- กำหนดขอบเขตและวัตประสงค์ของระบบ (Planning)
- วิเคราะห์ความต้องการของระบบ (Design)
- ศึกษาความเหมาะสมและความเป็นได้ของระบบ (Analysis)
- พัฒนาระบบตามความต้องการให้มีส่วนประกอบครบตามที่ได้ศึกษาแล้ว (Prototyping model)

(3) ทำการศึกษาวิเคราะห์ ประเมินระบบถึงความต้องการและหน้าที่ที่สำคัญต่าง ๆ ของระบบที่ได้ทำการจำลองระบบและพัฒนาขึ้นมา (Prototyping Analysis)

(4) พิจารณาจากข้อ 3 เพื่อดำเนินการพัฒนาโครงสร้างซอฟต์แวร์ อันได้แก่

- ออกแบบองค์ประกอบของซอฟต์แวร์ (Component)
- การเขียนโปรแกรม (Coding)
- ทดสอบซอฟต์แวร์ (Testing)

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

(5) ดำเนินการพัฒนาตามข้อ 4 (Development)

(6) เป็นส่วนสุดท้าย เมื่อพัฒนาต้นแบบสมบูรณ์แล้ว ควรปรับปรุงและสนับสนุนให้มีการนำไปใช้อย่างต่อเนื่องและจริงจัง (Operation and Support)

ข้อดีของ Incremental Model

สามารถวางแผนและแบ่งการพัฒนาเป็นรอบ ๆซึ่งในแต่ละรอบก็จะได้ต้นแบบของซอฟต์แวร์ ทำให้ได้ซอฟต์แวร์ที่ตรงกับความต้องการและมีโอกาสใช้ซอฟต์แวร์จริงเร็วกว่าการพัฒนาด้วยกระบวนการอื่น

ข้อเสียของ Incremental Model

หากแบ่งรอบการพัฒนาไม่ดี จะกลายเป็นการพัฒนาแบบไม่มีกระบวนการ ไม่มีขั้นตอนในการพัฒนา (Build and Fix) และได้ระบบที่มีโครงสร้างไม่สมบูรณ์ เปราะบาง ยากต่อการพัฒนาปรับปรุง

2.1.2 การรวมระบบแอพพลิเคชั่นสำหรับองค์กรขนาดใหญ่

การรวมระบบแอพพลิเคชั่น (Enterprise Application Integration หรือ EAI) เป็นกระบวนการที่ใช้ในการเชื่อมโยงแต่ละหน่วยการทำงานขององค์กร ทั้งหน่วยการทำงานทางธุรกิจและสารสนเทศให้สามารถทำงานร่วมกันได้ โดยที่ระบบ Enterprise Application Integration หรือ EAI จะทำการรวบรวมแอพพลิเคชันที่มีอยู่ในระบบที่กระแสงานต่างกันรูปแบบคำสั่ง โครงสร้างของข้อมูล รูปแบบของการสื่อสาร โดยผู้ใช้สามารถเข้าถึงได้แต่ปัญหาที่พบจากการทำ Enterprise Application Integration หรือ EAI คือ ต้นทุนสูง รูปแบบของข้อมูลที่ต่างกัน จะมีปัญหาในการแปลงข้อมูลมาก ซึ่งมีผลต่อความปลอดภัยและการนำข้อมูลกลับมาใช้ใหม่อีกครั้ง สำหรับแนวคิดการรวมโปรแกรมประยุกต์ มองได้ดังนี้

(1) การจัดลำดับการรวมข้อมูล (Data–Level Integration)เป็นแอพพลิเคชันที่ใช้ฐานข้อมูลของคนอื่นร่วมด้วย โดยไม่จำเป็นต้องรับรู้หรือถูกควบคุมโดยแอพพลิเคชันอื่น ที่เป็นเจ้าของฐานข้อมูล ทำให้ง่ายสะดวกและต้นทุนต่ำ แต่ต้องรู้โครงสร้างของฐานข้อมูลแอพพลิเคชันอื่นด้วย ทำให้อาจเกิดความไม่ปลอดภัยเกิดขึ้น

(2) การเชื่อมต่อโปรแกรม (Application interface Integration) จะเชื่อมต่อกันเชิงตรรกศาสตร์ (Logic) และจะต้องมีความเกี่ยวข้องกัน โดยตรงระหว่างกระบวนการ(process)

(3) ลำดับขั้นตอนในการเชื่อมต่อโปรแกรม (Process – level Integration)
สามารถทำการเชื่อมต่อผ่านโปรแกรมสื่อสารระดับสูง ซึ่งมีจะมีตัวกลางที่ทำหน้าที่เป็นตัวจัดการในการติดต่อสื่อสาร และมีองค์ประกอบ ที่ทำหน้าที่เป็นตัวช่วยให้ทำงานร่วมกันแบบประสานและเป็นตัวช่วยจัดการประมวลผลของกระบวนการทางธุรกิจ โดยได้นำเอาแนวทางของ Service Oriented

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

Architecture นำมาใช้ในการแก้ปัญหาดังกล่าว เพื่อแบ่งระดับเซอร์วิสและความเหมาะสมในการเชื่อมต่อ

2.1.3 เอกซ์เอ็มแอล

XML ย่อมาจาก Extensible Markup Language เป็นภาษาที่ได้รับการออกแบบมา เพื่อให้สามารถนิยามความหมายของข้อมูลได้ หรือที่เรียกว่า Data defimition โดยอนุญาตให้ผู้ใช้นิยามแท็กขึ้นมาได้เอง เช่น

<course> online and cd-rom retrieval system</course>

จุดที่ภาษา XML ต่างจากภาษา HTML คือ ภาษา HTML จะประกอบไปด้วยแท็ก สำหรับการใช้งานที่ถูกกำหนดไว้เรียบร้อยแล้ว ผู้ใช้เพียงแต่นำแท็กมาใช้แต่ในภาษา XML ไม่มีแท็กที่ถูกนิยามไว้ก่อน ผู้ใช้ต้องสร้างแท็กขึ้นมาเองเพื่อนำมาอธิบายข้อมูล เพื่อบอกหน้าที่และประเภทของข้อมูลของส่วนต่าง ๆ ของเอกสารนั้นได้โดยชัดเจน สามารถเพิ่มแท็กเข้าไปในเอกสาร ซึ่งจะทำให้โครงสร้างของเอกสารชัดเจนขึ้นและทำให้การประมวลผลเอกสารเป็นไปโดยง่ายและไม่จำเป็นที่จะต้องอาศัยมนุษย์ เพื่อตีความเอกสาร

เราใช้เทคโนโลยี XML ในการพัฒนามาตรฐานเพื่อการกระจายข่าว เนื่องจาก XML เป็นภาษาที่เหมาะกับการแลกเปลี่ยนข้อมูลผ่านเครือข่ายคอมพิวเตอร์ เนื่องจาก XML ไม่ได้ขึ้นอยู่กับโปรแกรมประยุกต์หรือระบบปฏิบัติการใด นอกจากนี้ XML ยังเป็นภาษาที่มีความยืดหยุ่น เนื่องจากผู้ใช้สามารถที่จะกำหนด และตั้งค่าแท็กให้เหมาะกับเอกสารเฉพาะที่ตนต้องการได้อย่างอิสระ และยังสามารถเพิ่มเติม metadata หรือแท็กได้ในภายหลัง โดยไม่มีผลกระทบต่อโปรแกรมที่มีอยู่แล้วด้วย

XML ถือได้ว่า เป็นส่วนหนึ่งของมาตรฐาน SGML (Standard Generalized Markup Language) ที่เป็นข้อกำหนดในการสร้างหรือจัดทำเอกสารในรูปแบบอิเล็คทรอนิกส์ ที่กำหนดโดย W3C (World Wide Web Consortium) ที่มีโครงสร้างและรูปแบบที่เปิดให้แอพพลิเคชันต่าง ๆ สามารถเรียกไปใช้งานได้ จึงทำให้การจัดการข้อมูลหรือเรียกใช้ข้อมูลจากแอพพลิเคชันต่าง ๆ นั้นอยู่ในมาตรฐานเดียวกัน โดยสามารถสร้างแอพพลิเคชั่น เพื่ออ่านและประมวลผล XML ได้อย่างง่ายดาย

เป้าหมายที่สำคัญของ XML จึงเน้นที่จะนำไปใช้งานในอินเทอร์เน็ต เป้าหมายหลักเป็นดังนี้

(1) XML เป็นงานที่นำไปประยุกต์บนอินเทอร์เน็ต โดยใช้ดูเอกสาร XML ได้ง่ายสะดวกและได้ผลดีเหมือน HTML

(2) XML ออกแบบอย่างพิถีพิถัน เพื่อเน้นที่ความจำเป็น กระทัดรัด เข้าใจง่ายและใช้

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

ประโยชน์ได้กว้างขวาง

(3) XML สนับสนุนการประยุกต์เข้ากับงานต่าง ๆ และสนับสนุนโปรแกรมประยุกต์ต่าง ๆ

(4) XML เน้นเรื่องการประมวลผลเอกสาร จึงเหมาะกับการไปใช้งานทางด้านการวิเคราะห์เอกสาร การผลิตเอกสารการแลกเปลี่ยน และการแสดงผล

(5) การเขียนด้วยภาษา XML ทำได้ง่าย

(6) คุณสมบัติของ XML ต้องอยู่ในระดับต่ำสุด เพื่อให้ผู้ใช้อื่นร่วมใช้ได้

(7) XML ควรอ่านได้ด้วยมนุษย์ โดยไม่ต้องอาศัยโปรแกรมหรือเครื่องมือช่วยแปล เพราะบางครั้งควรอ่านและเข้าใจได้ด้วยอักษรธรรมดา

(8) รูปแบบการเขียนโครงสร้างและข้อกำหนดของ XML เป็นไปตามหลักการวิศวกรรมคอมพิวเตอร์ คือ เมื่อเขียนแล้วต้องสามารถใช้โปรแกรมแปลภาษาได้ง่าย โดยทั่วไปเขียนในรูปแบบ BNF (Baches Normal Form) ได้

(9) การเขียน XML ทำได้ตั้งแต่การใช้เครื่องมือทั่ว ๆไปและไม่ต้องการเครื่องมือที่ซับซ้อน อย่างไรก็ดีย่อมต้องมีผู้เขียน XML editor ให้ใช้งานได้ง่ายขึ้น

(10) XML เป็นมาตรฐานที่กำหนดแล้วใช้งานได้ทันที โดยบราวน์เซอร์และอุปกรณ์ต่างๆ พร้อมใช้งานร่วมกัน

(11) เป้าหมายที่สำคัญของ XML อีกอย่างหนึ่ง คือ ใช้เป็นตัวควบคุมข้อมูล (Meta data) ดังนั้นจึงเป็นแนวทางในการขนส่งข้อมูลและสร้างการเชื่อมโยงระหว่างแอบปลิเคชั่นได้ง่าย ด้วยเหตุผลของการใช้งานบนเครือข่ายที่มีพัฒนาการ การจัดการเอกสารจำนวนมาก การสร้างดิจิตอลไลบารี การแลกเปลี่ยนข้อมูลข่าวสารระหว่างกัน การประยุกต์ XML จึงทำได้กว้างขวาง เช่น

- XML สนับสนุน UNICODE ทำให้ใช้ได้หลากหลายภาษา และผสมกันได้หลากหลายภาษา

- การพัฒนา XML Processor ทำให้สามารถดึงเอกสาร XML มาใช้งานได้ง่ายและใช้ร่วมกับโปรแกรมประยุกต์อื่นได้ง่าย เช่น โปรแกรม DB2, Oracle, SAP เป็นต้น

- XML ช่วยทำให้เกิดการรับส่งข้อมูลแบบ EDI โดยทำให้แนวทางการเชื่อมโยงและสร้างความเป็นเอกสารหรือมาตรฐานระหว่างองค์กร

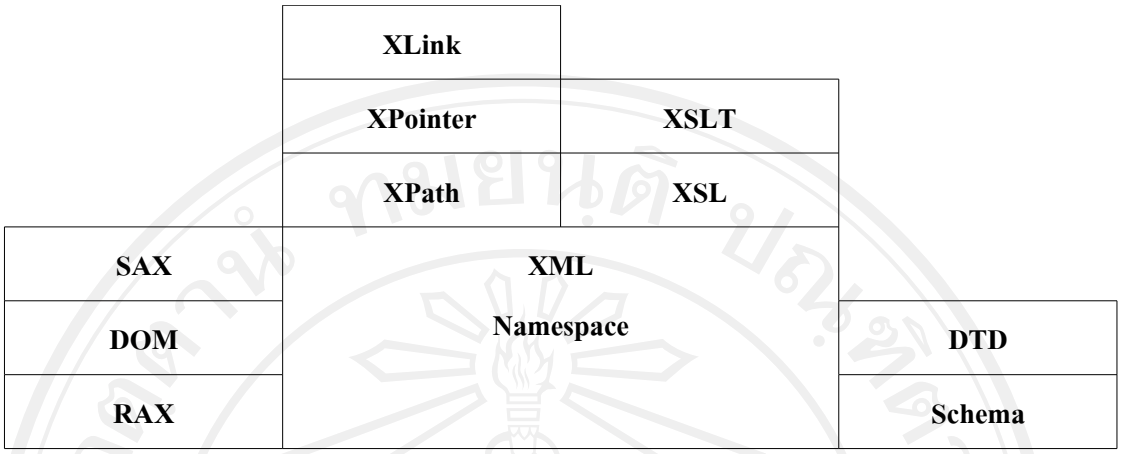
- XML มีสภาพช่วยในการขนส่งข้อมูลไปยังปลายทาง เพื่อให้แปลความหมายและใช้งานได้อย่างเต็มประสิทธิภาพ

- มีการสร้างการประยุกต์ และนำเสนอผลลัพธ์ไปใช้งานจาก XML ได้มาก

- การประยุกต์ใช้ในการดำเนินกิจกรรมบนเครือข่ายมีอีกมากมาย อย่างเช่น eBusiness EDI eCommerce การจัดการ Supply chain, Demand chain management ในการดำเนินการแบบ intranet

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

และ web base application



รูปที่ 2.2 XML forms the core of the family of the XML specifications

มาตรฐานต่าง ๆ ที่เกี่ยวข้องกับภาษาเอกซ์เอ็มแอล (XML) ที่ได้นำเสนอให้แก่ World Wide Web Consortium (W3C) เพื่อให้รับรองมีจำนวนมาก ในที่นี้จะขอยกตัวอย่างแต่พอสังเขปดังนี้

(1) XML คือ รูปแบบข้อมูลที่มีโครงสร้างที่เอื้อต่อการแลกเปลี่ยนข้อมูลในระบบเครือข่ายอินเตอร์เนตที่เป็นมาตรฐาน ได้ถูกกำหนดโดย World Wide Web Consortium (W3C)

(2) Namespaces นำไปใช้สำหรับช่วยในการกำหนดขอบเขต และชื่อ ให้กับองค์ประกอบและเขตข้อมูล (attributes) เพื่อหลีกเลี่ยงการถูกเรียกใช้ซ้ำกันระหว่างเอกสารเอกซ์เอ็มแอล จากแหล่งกำเนิดที่ต่างกัน โดยที่ namespace ในแต่ละชุดเป็นการรวมของชื่อต่าง ๆในองค์ประกอบของเอกซ์เอ็มแอลและคอลัมน์ที่ถูกระบุพร้อมกับ URI (Uniform Resource Identifier) ที่เป็นแหล่งกำเนิดของเอกสารเอกซ์เอ็มแอลนั้น

(3) XSL (Extensible Stylesheet Language) เป็นภาษาที่ใช้ในการจัดรูปแบบของเอกสารให้กับข้อมูลที่มีอยู่ในเอกซ์เอ็มแอล เพื่อใช้สำหรับนำเสนอในเว็บบราวเซอร์หรือสื่อนำเสนออื่น ๆ มีศักยภาพสูงกว่า CSS (Cascading style Sheet) ที่ใช้เป็นการจัดรูปแบบของเอกสารสำหรับ HTML (Hypertext Markup Language) มาก

(4) XSLT (Extensible Stylesheet Language Transformation) ซึ่งถือได้ว่าเป็นเครื่องมืออันทรงพลังที่จะใช้แปลงรูปแบบข้อมูลของเอกซ์เอ็มแอลให้อยู่ในรูปต่างๆ เช่น เฮดทีเอ็มแอลหรือเอกซ์เอ็มแอลหรือรูปแบบใด ๆ ก็ได้ XSLT ใช้ XPath เพื่อสร้างผลลัพธ์ที่ต้องการอธิบายถึงโครงสร้างหรือการแสดงผลที่เป็นเอกซ์เอ็มแอล

(5) XLink เป็นการปรับปรุงการเชื่อมโยง (link) ในเฮดทีเอ็มแอล โดยสนับสนุนการเชื่อม

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

โยงแบบสองทาง แบบหนึ่งต่อหลายแห่งและแบบแยกตามประเภท (ประเภทการเชื่อมโยง เช่น การอ้างอิงถึงส่วนคำอธิบายเพิ่มเติมและพจนานุกรม เป็นต้น) XLink มีพื้นฐานการทำงานจาก Hytime

(6) Xpointer อนุญาตให้สามารถชี้ไปที่จุด หรือตำแหน่งใดก็ได้ ภายในเอกสารที่เป็นเป้าหมาย โดยไม่ต้องกำหนดการเชื่อมโยงแบบ HTML link ไว้ล่วงหน้า มีพื้นฐานการทำงานจาก TEI (Text Encoding Initiative)

(7) XPath (XML Path Language) คือ ประโยคที่เอาไว้ใช้ระบุหรือค้นหาข้อมูลที่อยู่ภายในเอกสาร XML โดยเป็นไปตามมาตรฐานของ World Wide Web Consortium (W3C)

(8) RDF (Resource Description Framework) เป็นคำอธิบายข้อมูลหรือข้อมูลที่ใช้บอกรายละเอียดของข้อมูล ที่ใช้สำหรับเอกสารเอกซ์เอ็มแอลคล้ายกับเป็นส่วนที่ใช้ เพื่อใส่ตัวข้อมูลที่ใช้แสดงหน้าที่และประเภทของข้อมูล (Metadata tags) ของเฮดทีเอ็มแอล เช่น ผู้แต่ง ลิขสิทธิ์ และวันที่จัดพิมพ์ เป็นต้น

(9) XSchema เข้ามาใช้แทนที่ DTD (Document Type Definition) เพื่อใช้สำหรับการกำหนดประเภทข้อมูลของเอกซ์เอ็มแอล เป็นการประยุกต์เทคโนโลยีหลายตัวเข้าด้วยกัน ได้แก่ XML - Data , SOX (Schema for Object - Oriented XML) , DCD (Document Content Dexcription) และ DDML (Document Definition Markup Language or XSchema)

(10) DTD (Document Type Definition) คือแฟ้มข้อมูลหรือแฟ้มข้อมูลหลายแฟ้มที่ใช้งานร่วมกัน ซึ่งได้บรรจุข้อกำหนด และกฎเกณฑ์ของเอกสารชุดที่เป็นข้อกำหนดต่าง ๆ สำหรับการกำหนดรูปแบบองค์ประกอบ

(11) XQL (XML Query Language)เป็นส่วนที่ขยายเพิ่มเติมของเอกซ์เอสแอล สำหรับใช้อ้างถึง เป็นตัวกลั่นกรองส่วนที่เป็นองค์ประกอบและข้อความของเอกสารเอกซ์เอ็มแอล เพื่อกำหนดรูปแบบที่ชัดเจนและให้เข้าใจได้ง่าย เพื่อใช้ในการเข้าถึงองค์ประกอบที่เฉพาะเจาะจงและใช้สำหรับการค้นหาโหนดที่มีคุณลักษณะพิเศษต่างๆนอกจากมาตรฐานในการประยุกต์เอกซ์เอ็มแอล เพื่อการนำไปใช้งานตามที่กล่าวมาแล้ว ยังมีความพยายามในการกำหนดมาตรฐานทั้งในด้าน Application Programming Interface (API) ให้แอปพลิเคชันให้สามารถเข้าถึงเอกสารเอกซ์เอ็มแอลทันที ที่เอกสารนั้นเข้าสู่กระบวนการสืบค้นข้อมูลหรือ parsing โดยมี APIs อยู่ 2 กลุ่มที่ได้รับความสนใจเป็นพิเศษ คือ

- DOM (Document Object Model) เป็นโครงสร้างแบบต้นไม้หรือ tree - based API ที่ใช้ในการประมวลโครงสร้างของเอกสารเอกซ์เอ็มแอล ให้เป็นโครงสร้างแบบต้นไม้ ซึ่งจะประกอบไปด้วยลำต้น กิ่งก้านสาขาและใบ เพื่อให้แอปพลิเคชันนั้นสามารถเข้าหาจุดต่าง ๆ ของโครงสร้างต้นไม้นั้นได้ โดยที่ขอบเขตจะถูกจำกัดด้วยหน่วยความจำที่เรียกใช้ได้ในขณะนั้น

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

- SAX (Simple API for XML) เป็นโครงสร้างตามลำดับเหตุการณ์หรือ event-based API รายงานตามสภาวะการณ์ของการสืบค้นข้อมูล หรือ parsing เช่น ที่เป็นจุดเริ่มต้นและสิ้นสุดขององค์ประกอบต่าง ๆ ไปให้แอปพลิเคชันผ่านทางการร้องขอ โดยปรกติทั่วไปแล้วไม่มีการสร้างโครงสร้างแบบต้นไม้ขึ้นมา เนื่องจากการทำงานเป็นแบบ event-based API การเข้าถึงเอกสารเอกซ์เอ็มแอล สามารถทำได้ง่ายและไม่ซับซ้อน และที่สำคัญคือผู้ใช้สามารถทำการสืบค้นข้อมูล หรือ parsing ของเอกสารที่มีขนาดใหญ่มากกว่าปริมาณของหน่วยความจำ ที่เรียกใช้ได้และทำให้ผู้ใช้สามารถสร้างโครงสร้างข้อมูลเอกสารในรูปแบบที่ตนเองต้องการได้

- PRAX (PHP Record - oriented API for XML) ใน RAX (Record - oriented API for XML) สามารถใช้ภาษาพีเอชพีช่วยในการจัดการกับเอกสารเอกซ์เอ็มแอลได้ และไม่จำเป็นต้องติดตั้งองค์ประกอบอะไรเพิ่ม โดยในภาษาพีเอชพี จะมีฟังก์ชันสำหรับจัดการกับเอกสารเอกซ์เอ็มแอลและนำไปใช้ในกระบวนการต่าง ๆ ของเอกซ์เอ็มแอลได้ เพื่อใช้อธิบายถึงแถวและคอลัมน์จากไวยากรณ์ของ dso_xml การเข้าถึงเอกสารเอกซ์เอ็มแอลจึงทำได้ง่ายและไม่ซับซ้อน ซึ่งโดยทั่วไป PRAX ในใช้การนำเข้าเอกซ์เอ็มแอล การสร้าง RAX object และใช้ในการอ่านข้อมูลหรือ URL เราจะใช้เทคโนโลยี XML ในการพัฒนามาตราฐานเพื่อการกระจายข่าว เนื่องจาก XML เป็นภาษาที่เหมาะกับการแลกเปลี่ยนข้อมูลผ่านเครือข่ายคอมพิวเตอร์ เนื่องจาก XML ไม่ได้ขึ้นอยู่กับโปรแกรมประยุกต์หรือระบบปฏิบัติการใด นอกจากนี้ XML ทั้งยังเป็นภาษาที่มีความยืดหยุ่น เนื่องจากผู้ใช้สามารถที่จะกำหนดและตั้งค่า metadata ให้เหมาะกับเอกสารเฉพาะที่ตนต้องการได้ และยังสามารถเพิ่มเติม metadata ได้ในภายหลังโดยไม่มีผลกระทบต่อโปรแกรมที่มีอยู่แล้วด้วย

2.1.4 คลังข้อมูล

ระบบคลังข้อมูล หรือ Data Ware - house ศูนย์เทคโนโลยีอิเลิกทรอนิกส์และคอมพิวเตอร์แห่งชาติให้ความหมายว่า เป็นระบบการจัดเก็บ รวบรวมข้อมูล ที่มีอยู่ในระบบปฏิบัติการต่าง ๆ ขององค์กร (มักเป็นองค์กรขนาดกลางถึงขนาดใหญ่ เช่น กรมสรรพสากร) โดยข้อมูลเหล่านั้น มักเป็นข้อมูลกระจัดกระจาย นำมารวมไว้เป็นศูนย์กลางข้อมูลขององค์กร และสามารถเก็บข้อมูลย้อนหลังได้หลาย ๆ ปี เพื่อใช้เป็นข้อมูลช่วยสนับสนุนการตัดสินใจ (Decision Support System) หรือนำไปใช้ในการวิเคราะห์ข้อมูลที่ถูกต้อง และมีประสิทธิภาพ โดยการวิเคราะห์ต้องทำได้แบบหลายมิติ (Multidimentsional Analysis)ตลอดจนการวิเคราะห์ทางด้านธุรกิจ เช่น การพยากรณ์ (Forecasting) , What-If Analysis,Data Mining เป็นต้น

(1) องค์ประกอบสำคัญ

- เครื่องมือในการดึงข้อมูล (Data Extract / Cleansing) จากแหล่งข้อมูลที่เก็บอยู่ในระบบปฏิบัติการต่างๆ ขององค์กร เพื่อจัดสร้างคลังข้อมูล

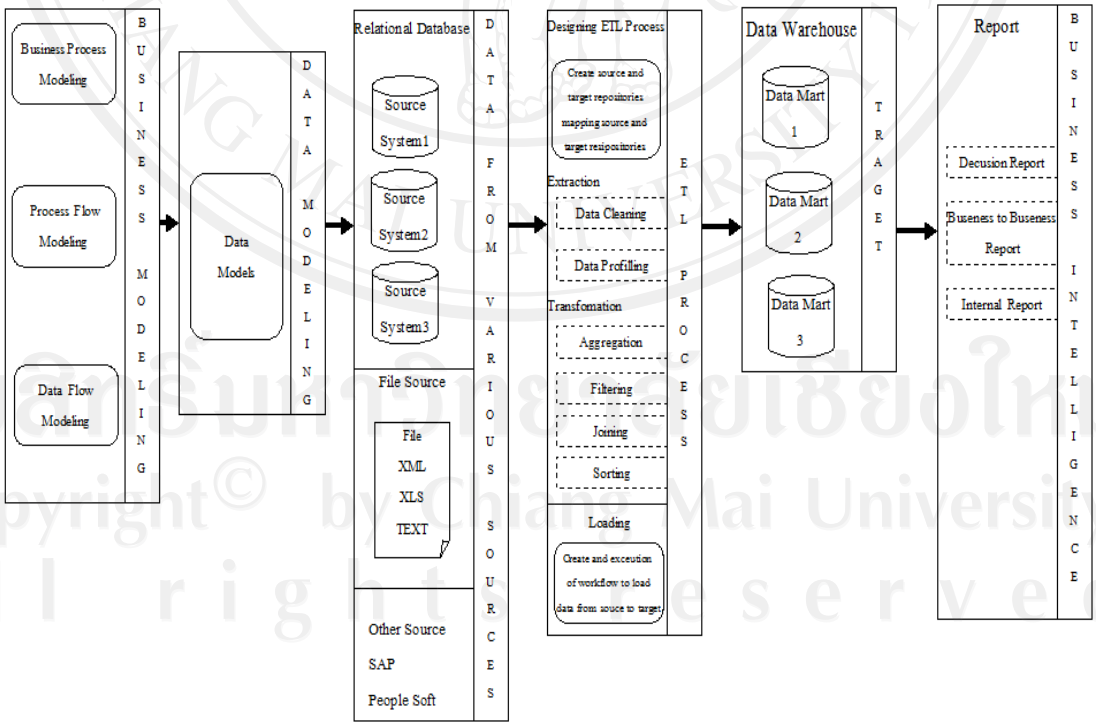
ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

- เครื่องมือในการเข้าไปเรียกค้นข้อมูลเชิงวิเคราะห์ (Front End Tool)

- โครงสร้างของระบบคอมพิวเตอร์ ฐานข้อมูล ระบบเครือข่าย ทีมงานติดตั้งระบบ

สำหรับซอฟต์แวร์ระบบวิเคราะห์ข้อมูล ที่ใช้งานในปัจจุบันนี้เรียกว่า ซอฟต์แวร์ระบบวิเคราะห์ข้อมูลแบบ On Line Analytical Processing (OLAP) ซึ่งปัจจัยที่ต้องคำนึงถึงในการเลือก OLAP ได้แก่

1. ความสามารถในการวิเคราะห์ข้อมูลแบบหลายมิติ (Multidimensional Analysis)
2. สถาปัตยกรรมแบบ Client/Server
3. ความสามารถในการจัดเก็บข้อมูลและความเร็วในการเข้าถึงข้อมูลขนาดใหญ่ (Performance Data Access)
4. เครื่องมือในการพัฒนาระบบ (Application Development Tools)
5. การดึงข้อมูล (transformation) จากแหล่งต่าง ๆ เช่น ระบบคลังข้อมูล , Flat File , Spreadsheet
6. ความสามารถในการจัดลำดับชั้นของข้อมูล (Hierarchy)
7. เครื่องมือในการวิเคราะห์ข้อมูล เช่น Forecasting , Statistic, Data Mining



รูปที่ 2.3 สถาปัตยกรรมคลังข้อมูล

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

(2) สิ่งที่ควรพิจารณาก่อนสร้างคลังข้อมูล

เนื่องจากการลงทุนสร้างคลังข้อมูลขึ้นมาใช้ เพื่อสนับสนุนการทำงานขององค์กรจำเป็นต้องมีค่าใช้จ่ายในการลงทุนมหาศาล ทั้งที่สามารถวัดออกมาเป็นตัวเงินได้ ตัวอย่างเช่น ค่าใช้จ่ายด้านคน ด้านฮาร์ดแวร์ ด้านซอฟต์แวร์และโครงสร้างพื้นฐานอื่น ๆ ที่จำเป็นต้องใช้ส่วนค่าใช้จ่ายที่ไม่เป็นตัวเงิน แต่ก็มีความสำคัญอย่างมาก ได้แก่ กำลังแรงงานที่เสียไปของทรัพยากรบุคคลขององค์กรและเวลาที่ใช้ในการพัฒนา ดังนั้นเมื่อองค์กรตัดสินใจสร้างคลังข้อมูลขึ้นแล้วควรจะต้องประสบความสำเร็จด้วย ทั้งนี้โพ (Poe) ได้เสนอแนวคิด 8 ประการหรือ The Big Eight ที่ควรให้ความสนใจโดยมีรายละเอียดดังนี้

1. ควรมีเป้าหมายที่ชัดเจนร่วมกันของการสร้างระบบนี้ ของคนในองค์กรเหมือนการตอบคำถามที่ว่าทำไมคุณถึงคิดจะสร้างคลังข้อมูล ซึ่งคำตอบที่ได้ขององค์กรที่ได้ก็คือเป้าหมายที่ต้องการโดยควรจะเขียนเป้าหมายนี้ออกมาเป็นลายลักษณ์อักษรที่ชัดเจน เพื่อให้ทีมพัฒนาได้เข้าใจเป้าหมายร่วมกัน

2. ทำความเข้าใจสถาปัตยกรรมของระบบ เพื่อให้ทีมพัฒนาเข้าใจตรงกันในที่นี้ หมายถึง พิมพ์เขียว หรือ Blueprint ที่แสดง E-R model (Entity-Relationship Model) รวมของระบบ หรือ Class Diagram รวมของระบบ ความเข้าใจที่ตรงกันทำให้งานเดินไปได้เร็วขึ้น

3. เทคโนโลยีที่ใช้ ควรอยู่ในวิสัยที่เหมาะสมต่อองค์กร ทั้งในด้านของตัวเงินและความยากง่ายในการเรียนรู้ ทั้งนี้หมายรวมทั้งฮาร์ดแวร์ ซอฟต์แวร์และเครือข่าย อาจมีการทดสอบและฝึกอบรมก่อนการใช้งานจริง

4. ทีมงานต้องมีวิสัยทัศน์ในเชิงบวกต่อการทำงาน เนื่องจากทีมงานพัฒนามักมาจากส่วนของงานด้านเทคโนโลยีสารสนเทศ แต่ในเนื้องานจริง ๆ แล้วเป็นของผู้ใช้ขั้นปลายของหน่วยงานต่าง ๆ ขององค์กร ดังนั้น จึงจำเป็นอย่างยิ่งที่จะให้ผู้ใช้ขั้นปลายที่เป็นเจ้าของงานเข้ามาร่วมทำงานด้วยตั้งแต่ต้นโครงการ

5. ต้องมีความมั่นใจได้ว่าทีมพัฒนาเข้าใจเป็นอย่างดีถึงความแตกต่างกันระหว่างฐานข้อมูลปฏิบัติการและฐานข้อมูลสนับสนุนการตัดสินใจ

6. ควรจัดให้มีการฝึกอบรม โดยการอบรมควรเป็นการฝึกอบรมก่อนเริ่มโครงการ โดยเฉพาะอย่างยิ่งการฝึกอบรมเกี่ยวกับเครื่องมือที่องค์กรจะใช้พัฒนา ทั้งนี้อาจเป็นการฝึกอบรมจากบริษัทผู้ขาย

7. ควรหาบุคลากรที่มีประสบการณ์ในการพัฒนาคลังข้อมูล เพื่อทำหน้าที่เป็นผู้จัดการโครงการหรือถ้าในองค์กรไม่เคยมีประสบการณ์ อาจจ้างที่ปรึกษาที่มีความเชี่ยวชาญและมีประสบการณ์ด้านนี้โดยเฉพาะมาช่วยทีมพัฒนา

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

8. โปรแกรมที่จะใช้นำเสนอข้อมูลในคลังข้อมูล จะต้องสามารถเรียนรู้ได้ง่ายและผู้ใช้ต้องสามารถใช้งานได้อย่างมีประสิทธิภาพ

(3) คุณลักษณะเฉพาะของคลังข้อมูล

จากนิยามของคลังข้อมูล ที่บ่งบอกถึงความแตกต่างกัน ระหว่างคลังข้อมูลกับฐานข้อมูลปฏิบัติการ ซึ่งสามารถสรุปคุณลักษณะของคลังข้อมูลได้ดังนี้

1. การแบ่งโครงสร้างตามเนื้อหา (Subject oriented) หมายถึง คลังข้อมูลที่ได้ถูกออกแบบมา เพื่อมุ่งเน้นไปในแต่ละเนื้อหาที่สนใจ ไม่ได้เน้นไปที่การทำงาน หรือกระบวนการแต่ละอย่างโดยเฉพาะ เหมือนอย่างฐานข้อมูลปฏิบัติการ ในส่วนของรายละเอียดข้อมูลที่จัดเก็บในระบบทั้งสองแบบก็จะแตกต่างกันไป ตามความต้องการใช้งานด้วยเช่นกัน ในคลังข้อมูลไม่จัดเก็บข้อมูลที่ไม่มีส่วนเกี่ยวข้องกับการประมวลผล เก็บเฉพาะข้อมูลที่ใช้สนับสนุนการตัดสินใจ ในขณะข้อมูลส่วนนั้นจะถูกเก็บไว้ในฐานข้อมูลปฏิบัติการรวมกับข้อมูลที่มีส่วนเกี่ยวข้องกับกระบวนการทำงาน

2. การรวมเป็นหนึ่ง (Integration) ซึ่งถือได้ว่าเป็นคุณลักษณะที่สำคัญที่สุดของคลังข้อมูล คือ เป็นการรวบรวมข้อมูลจากหลายฐานข้อมูลปฏิบัติการเข้าด้วยกัน และทำให้ข้อมูลมีมาตราฐานเดียวกัน เช่น กำหนดให้มีค่าตัวแปรของข้อมูลในเนื้อหาเดียวกัน ให้เป็นแบบเดียวกันทั้งหมด

3. ความสัมพันธ์กับเวลา (Time variancy) หมายถึง ข้อมูลในคลังข้อมูลจะต้องจัดเก็บโดยการกำหนดช่วงเวลาเอาไว้ โดยจะสัมพันธ์กับการดำเนินธุรกิจของหน่วยธุรกิจนั้นเพราะว่า ในการตัดสินใจด้านการบริหารจำเป็นต้องมีข้อมูลเปรียบเทียบในแต่ละช่วงเวลา ในแต่ละจุดของข้อมูลจะเกี่ยวข้องกับจุดของเวลา และข้อมูลแต่ละจุด สามารถเปรียบเทียบกันได้ตามช่วงเวลา

4. ความเสถียรของข้อมูล (Nonvolatile) หมายถึง ข้อมูลในคลังข้อมูลจะต้องไม่เปลี่ยนแปลงบ่อย ไม่ว่าจะเป็นการเพิ่มเติมข้อมูลใหม่ หรือเป็นการปรับปรุงแก้ไขข้อมูลเดิมที่บรรจุอยู่แล้ว ผู้ใช้ทำได้เพียงการเข้าถึงข้อมูลเท่านั้น

(4) สถาปัตยกรรมคลังข้อมูล

เป็นโครงสร้างมาตรฐานที่ใช้อธิบาย เพื่อให้เข้าใจแนวคิดและกระบวนการของคลังข้อมูลนั้น ๆ ซึ่งโดยทั่วไปแล้ว คลังข้อมูลในแต่ละระบบอาจจะมีรูปแบบที่ไม่เหมือนกันได้ เพื่อให้เหมาะสมกับองค์กรนั้น ๆ ทั้งนี้ ส่วนประกอบต่าง ๆ ภายในคลังข้อมูลที่สำคัญ ได้แก่

- ชั้นบริหารจัดการฐานข้อมูล (Operational database หรือ External database layer)

ทำหน้าที่จัดการกับข้อมูลในระบบงานปฏิบัติการหรือแหล่งข้อมูลภายนอกองค์กร

- ชั้นการติดต่อสื่อสาร (Information access layer)

เป็นส่วนที่ผู้ใช้ปลายทางติดต่อผ่านโดยตรง ประกอบด้วยฮาร์ดแวร์และซอฟต์แวร์ที่ใช้ในการแสดงผล เพื่อการวิเคราะห์ โดยมีเครื่องมือช่วยเป็นตัวกลางที่ผู้ใช้ ใช้ติดต่อกับคลังข้อมูล

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

โดยในปัจจุบันเครื่องมือที่ได้รับความนิยมเพิ่มขึ้นอย่างรวดเร็ว นั่นคือ Online Analytical Processing Tool หรือ OLAP Tool ซึ่งจะเป็นเครื่องมือที่มีความสามารถ ในการวิเคราะห์ที่ซับซ้อนและแสดงข้อมูลในรูปแบบหลายมิติ

- ชั้นติดต่อข้อมูล (Data access layer) เป็นส่วนต่อประสานระหว่าง Information access layer กับ Operational layer

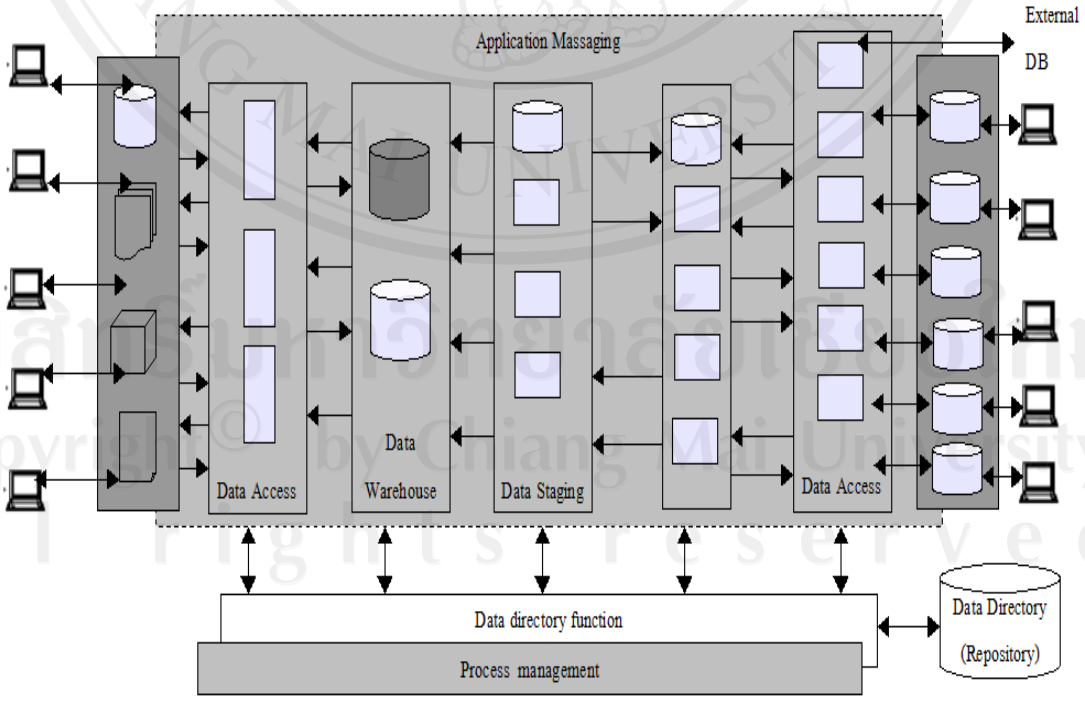
- ชั้นเชื่อมต่อข้อมูล (Data director (Metadata) layer) เพื่อให้เข้าถึงข้อมูลได้ง่ายขึ้นและเป็นการเพิ่มความเร็วในการเรียกใช้ และดึงข้อมูลของคลังข้อมูล

- ชั้นบริหารจัดการกระบวนการทำงาน (Process management layer) ทำหน้าที่จัดการกระบวนการทำงานทั้งหมด

- ชั้นการส่งข้อมูล (Application messaging layer) เป็นเครื่องมือหรือมิดเดิลแวร์ทำหน้าที่ในการส่งข้อมูลภายในองค์กรผ่านทางระบบเครือข่าย

- ชั้นข้อมูล (Data warehouse (physical) layer) เป็นแหล่งเก็บข้อมูลของทั้งข้อมูลด้านสารสนเทศและข้อมูลภายนอกองค์กร ในรูปแบบที่ง่ายแก่การเข้าถึงข้อมูลและยืดหยุ่นได้

- ชั้นปรับปรุงข้อมูล (Data staging layer) เป็นกระบวนการแก้ไขและดึงข้อมูลจากข้อมูลภายนอกองค์กร



รูปที่ 2.4 ส่วนประกอบของคลังข้อมูล

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

(5) เทคนิคในการสร้างคลังข้อมูล

การเคลื่อนที่ของข้อมูลในคลังข้อมูล ข้อมูลที่จะจัดเก็บภายในคลังข้อมูลจะมีการเคลื่อนที่ของข้อมูล (information flow) ได้เป็น 5 ประเภทดังต่อไปนี้

1. Inflow คือ การนำข้อมูลจากฐานข้อมูลอื่น นำเข้าสู่คลังข้อมูลทั้งจากฐานข้อมูล ภายในองค์กรและจากภายนอกองค์กร ในขั้นตอนนี้อาจมีการเปลี่ยนแปลงของโครงสร้างข้อมูล มีการทำ Denormalize การลบหรือการเพิ่มฟิลด์ เพื่อให้ข้อมูลทั้งหมดที่อยู่ในเนื้อหาเป็นแบบเดียวกัน ในขั้นตอนนี้อาจใช้เครื่องมือที่เรียกว่า Data Warehouse Tool

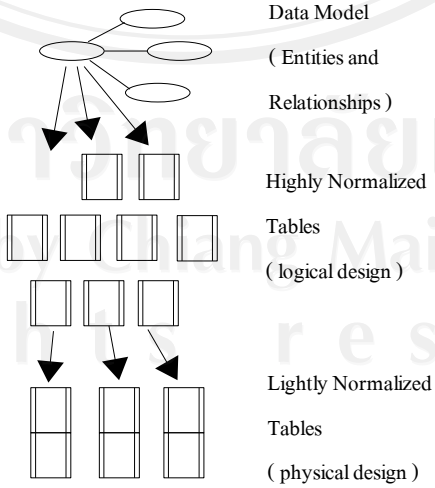
2. Upflow เมื่อข้อมูลที่เราต้องการจัดอยู่ในคลังข้อมูลแล้ว ซึ่งในบางครั้งเราอาจต้องมีการเพิ่มคุณค่าให้กับข้อมูลด้วย เพื่อให้ข้อมูลอยู่ในรูปแบบที่เป็นประโยชน์มากที่สุด ต่อการนำมาใช้ ซึ่งได้แก่ การจัดกลุ่มข้อมูล หาค่าทางสถิติที่ซับซ้อน การจัดข้อมูลให้อยู่ ในรูปแบบเหมาะสมหรือเทมเพลตมาตรฐาน

3. Downflow จะเป็นขั้นตอนของการปรับปรุง เปลี่ยนแปลงข้อมูลเก่าและไม่อยู่ในเนื้อหาที่องค์กรสนใจออกไปจากคลังข้อมูลขององค์กร

4. Outflow เป็นขั้นตอนที่ผู้ใช้เรียกใช้ข้อมูลในคลังข้อมูลผ่านเครื่องมือต่าง ๆ โดยจะมีการเรียกใช้ อาจมีเพียงขอเรียกเป็นครั้งคราว เป็นประจำทุกวันหรือเดือน หรือแม้กระทั่งต้องการแบบทันที

5. Metaflow ข้อมูลที่จะจัดเก็บในคลังข้อมูล จะทำสำเนาข้อมูลไว้อีกชุดหนึ่ง เพื่อระบุแหล่งที่มาของข้อมูลหรือแม้กระทั่งที่อยู่ของข้อมูลนั้น ๆ ในคลังข้อมูลและข้อมูลอื่นที่เกี่ยวข้อง

(6) วิธีการออกแบบฐานข้อมูลสำหรับคลังข้อมูล



รูปที่ 2.5 การออกแบบคลังข้อมูล

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

วิธีการนี้ถูกเสนอ โดย Kimball ในปี 1996 เรียกว่าระเบียบวิธี 9 ขั้นหรือ Nine - Step Methodology โดยวิธีการนี้ เริ่มจากการออกแบบจากส่วนย่อยที่แสดงถึงแต่ละระบบงานขององค์กร หรือเรียกอีกอย่างหนึ่งว่าคลังข้อมูลย่อย (Data mart) โดยเมื่อออกแบบแต่ละส่วนสำเร็จ แล้วจึงนำมารวมกันเป็นคลังข้อมูลขององค์กรในขั้นสุดท้ายซึ่งขั้นตอนทั้ง 9 ขั้นตอน โดยมีรายละเอียดดังนี้

1 . กำหนดคลังข้อมูลย่อย (Data mart) คือ การเลือกว่าจะสร้างคลังข้อมูลย่อยของระบบงานใดบ้าง และระบบงานใดเป็นระบบงานแรก โดยองค์กรจะต้องสร้าง E-R model ที่รวมระบบงานทุกระบบขององค์กรไว้แสดงการเชื่อมโยงของแต่ละระบบงานอย่างชัดเจน และสิ่งที่ต้องคำนึงถึงในการเลือกระบบงานที่จะเป็นคลังข้อมูลย่อยแรกนั้น มี 3 ปัจจัยที่เกี่ยวข้อง ได้แก่ จะต้องสามารถพัฒนาออกมาได้ทันตามเวลาที่ต้องการ โดยอยู่ในกรอบงบประมาณที่ได้กำหนดไว้ และต้องตอบปัญหาทางธุรกิจให้แก่องค์กรได้ ดังนั้นคลังข้อมูลย่อยแรกควรจะเป็นของระบบงานที่นำรายได้เข้ามาสู่องค์กรได้ เช่น ระบบงานขาย เป็นต้น

2. กำหนดตารางข้อเท็จจริง (fact table) ของคลังข้อมูลย่อย คือ กำหนดเนื้อหาหลักที่ควรจะเป็นของคลังข้อมูลย่อย โดยการเลือกตารางหลัก ๆ และกระบวนการที่ได้เกี่ยวกับตารางนั้น ๆ ออกมาจาก E - R model ขององค์กร นั้นหมายถึงว่าจะทำให้เราทราบถึงตารางไดเมนชั่นที่ควรจะมีด้วย

3. กำหนดเขตข้อมูล (Attribute)ที่จำเป็นในแต่ละตารางไดเมนชั่น คือการกำหนดเขตข้อมูลที่จะใช้บ่งบอกหรืออธิบายถึงรายละเอียดของตารางไดเมนชั่นได้ ทั้งนี้เขตข้อมูลที่เป็นกุญแจสำคัญ (primary key) ควรเป็นค่าที่คำนวณได้ กรณีที่มีคลังข้อมูลย่อยมากกว่าหนึ่ง คลังข้อมูลย่อยมีตารางไดเมนชั่นเหมือนกัน นั่นหมายถึงว่า เขตข้อมูลที่มีอยู่ในตารางไดเมนชั่นนั้น จะต้องเหมือนกันทุกประการ นั้นไม่อาจทำการแก้ไขปัญหาการจัดเก็บข้อมูลซ้ำซ้อน อันนำมาสู่ความแตกต่างกันของข้อมูลในชุดเดียวกัน ปัญหานี้จึงเป็นการดีที่จะมีการใช้ตารางไดเมนชั่นร่วมกัน ในแต่ละตารางข้อเท็จจริงที่จำเป็นต้องมีตารางไดเมนชั่นดังกล่าว โดยเราจะเรียกตารางไดเมนชั่นลักษณะแบบนี้ ว่า comformed และเรียกตารางข้อเท็จจริงว่า fact constellation ซึ่งเราสามารถกำหนดข้อดีของการใช้ตารางไดเมนชั่นร่วมกันได้ดังนี้

- แน่ใจได้ว่าในแต่ละรายงานจะออกมาสอดคล้องกัน
- สามารถสร้างคลังข้อมูลย่อยในเวลาต่างๆ กันได้
- สามารถเข้าถึงคลังข้อมูลย่อยโดยผู้พัฒนากลุ่มอื่นๆ
- สามารถรวบรวมคลังข้อมูลย่อยหลายๆ อันเข้าด้วยกัน
- สามารถออกแบบคลังข้อมูลร่วมกันได้

4. กำหนดเขตข้อมูลที่จำเป็นในตารางข้อเท็จจริง โดยเขตข้อมูลหลักในตารางข้อเท็จจริงซึ่งจะมาจากกุญแจสำคัญในแต่ละตารางไดเมนชั่น นอกจากนี้แล้ว ก็ยังสามารถมีเขตข้อมูลที่จำเป็น

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

อื่น ๆ ประกอบอยู่ด้วยเช่น เขตข้อมูลที่ได้จากการคำนวณค่าเบื้องต้น ที่มีความจำเป็นสำหรับการคงอยู่ของเขตข้อมูลอื่น ๆ ในตารางข้อเท็จจริงหรือที่เรียกอีกอย่างหนึ่งว่าการวัด (measure) การกำหนดเขตข้อมูลนี้ ไม่ควรเลือกเขตข้อมูลที่ไม่สามารถนำมาคำนวณได้ เช่น เป็นตัวหนังสือ หรือไม่ใช่ตัวเลขเป็นต้นและไม่ควรเลือกเขตข้อมูลที่ไม่เกี่ยวข้องกับเนื้อหาของตารางข้อเท็จจริงที่เราสนใจด้วย

5. จัดเก็บค่าจากการคำนวณเบื้องต้น ในตารางข้อเท็จจริง คือ การจัดเก็บค่าที่ได้จากการคำนวณให้เป็นเขตข้อมูลในตารางข้อเท็จจริง ถึงแม้ว่าจะสามารถหาค่าได้จากเขตข้อมูลอื่น ๆ ก็ตาม ทั้งนี้เพื่อให้การสอบถามมีประสิทธิภาพมากขึ้น และให้สามารถทำงานด้วยความเร็วที่เพิ่มขึ้น เนื่องจากไม่ต้องคำนวณค่าใหม่ทั้งหมด ถึงแม้ว่าจะเกิดความซ้ำซ้อนของข้อมูลในการจัดเก็บบ้างก็ตาม

6. ต้องเขียนคำอธิบายตารางไดเมนชั่น ทั้งนี้ก็เพื่อทำให้ผู้ใช้สามารถใช้งานคลังข้อมูลย่อยได้อย่างมีประสิทธิภาพเพราะเกิดความเข้าใจอย่างดีในส่วนต่าง ๆ

7. กำหนดระยะเวลาที่ใช้ในการจัดเก็บข้อมูลในฐานข้อมูล โดยอาจเป็นการจัดเก็บเพียงช่วงระยะเวลา 1-2 ปีหรือนานกว่านั้นขึ้นอยู่กับความต้องการขององค์กร เนื่องจากองค์กรแต่ละประเภทมีความต้องการในการจัดเก็บข้อมูลต่างช่วงเวลากัน ทั้งนี้อาจขึ้นอยู่กับความจำเป็น หรือข้อกำหนดที่ใช้ในการดำเนินธุรกิจ มีข้อสังเกตอยู่ 2 ประการ ที่น่าสนใจและสำคัญสำหรับการออกแบบเขตข้อมูล ในเรื่องของการจัดเก็บข้อมูล ดังนี้

7.1 ข้อมูลที่ถูกจัดเก็บไว้นานเกินไป มักเกิดปัญหาการอ่านหรือแปลข้อมูลนั้น ๆ จากแฟ้มหรือเทปเก่า

7.2 เมื่อมีการนำรูปแบบเก่าของตารางไดเมนชั่นนำมาใช้ ซึ่งอาจเกิดปัญหาจากการเปลี่ยนแปลงของตารางไดเมนชั่นอย่างช้า ๆ ได้

8. การติดตามปัญหาการเปลี่ยนแปลงของตารางไดเมนชั่นอย่างช้า ๆ คือ การเปลี่ยนเอาเขตข้อมูลของตารางไดเมนชั่นเก่ามาใช้ แล้วส่งผลกระทบต่อข้อมูลปัจจุบันของตารางไดเมนชั่น โดยสามารถแบ่งประเภทของปัญหาที่เกิดได้ 3 ประเภท ดังนี้

8.1 เกิดการเขียนทับข้อมูลใหม่ โดยข้อมูลเก่า

8.2 เกิดระเบียนใหม่ ๆ ขึ้นใน ตารางไดเมนชั่น

8.3 เกิดระเบียนที่มีทั้งค่าเก่าและใหม่ปนกันไป

9. กำหนดคิวรี่ (query) เป็นการออกแบบด้านกายภาพเพื่อให้ผู้ใช้เกิดความสะดวก ในการใช้งานและสามารถทำงานได้อย่างมีประสิทธิภาพ เมื่อได้ดำเนินการทั้ง 9 ขั้นตอน สำหรับแต่ละคลังข้อมูลย่อยเสร็จแล้ว จึงนำทั้งหมดมารวมกันเป็นภาพรวมของคลังข้อมูลขององค์กรต่อไป

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

(7) การแปลงข้อมูลเข้าสู่คลังข้อมูลย่อย

เมื่อเราได้ออกแบบฐานข้อมูลสำหรับแต่ละคลังข้อมูลย่อยเสร็จแล้ว ขั้นตอนต่อไปที่สำคัญยิ่ง ก็คือ การนำข้อมูลจากแหล่งข้อมูลต่าง ๆ ไปแปลงให้อยู่ในแพลตฟอร์มของฐานข้อมูล ที่เราได้ออกแบบไว้ นั่นก็คือการแปลงข้อมูลหรือ Extraction Transformation and Loading (ETL) นั่นเอง โดยที่คุณภาพของการแปลงข้อมูลเป็นสิ่งสำคัญมาก สำหรับการสร้างคลังข้อมูล ซึ่งจะแตกต่างกันไปตามคลังข้อมูลที่แต่ละองค์กรต้องการ โดยที่การแปลงข้อมูลหมายรวม ตั้งแต่การวิเคราะห์แหล่งข้อมูลกำหนดการส่งข้อมูล การรวบรวมข้อมูลหรือสร้างข้อมูลภายนอก วางแผน และสร้างลำดับงานของการแปลงข้อมูลและตรวจสอบความถูกต้องของข้อมูลที่ได้ สามารถสรุปเป็นขั้นตอนได้ดังนี้

1. วิเคราะห์แหล่งข้อมูล เช่น ปริมาณของข้อมูล จำนวน ประเภทของการเข้าถึงแหล่งข้อมูล รูปแบบข้อมูลหรือแพลตฟอร์ม และภาษาโปรแกรมที่ใช้เป็นต้น

2. ให้ย้ายข้อมูลที่ต้องการจากระบบเดิม มาไว้ในบริเวณที่ใช้ปรับปรุงข้อมูล หรือเรียกบริเวณนี้ว่า staging area เพื่อนำมาเลือกเฉพาะส่วนที่ต้องการแปลงข้อมูล และตรวจสอบความถูกต้องหรือการทำความสะอาดข้อมูล

3. ให้กำหนดกุญแจสำคัญของตารางข้อเท็จจริงและตารางไดเมนชั่น โดยกำหนดกุญแจสำรอง (foreign key) ระหว่างตารางข้อเท็จจริงกับตารางไดเมนชั่น

4. ย้ายข้อมูลที่มีความสมบูรณ์แล้ว จากส่วนจัดการข้อมูลลงสู่เครื่องแม่ข่าย (Server) ของคลังข้อมูลย่อย สร้างข้อมูลที่เกี่ยวกับข้อมูลของแต่ละคลังข้อมูลย่อย โดยทำการเก็บรายละเอียดของข้อมูลการอัปเดต และส่งออกไปไว้ในคลังข้อมูลย่อย

5. ตรวจสอบความถูกต้องของข้อมูล ต้องกระทำโดยตลอดทั้งกระบวนการแปลงข้อมูล ซึ่งทำได้ดังนี้

5.1 ตรวจสอบผลรวมทั้งหมดของจำนวนข้อมูลที่ดึงมาจากแหล่งข้อมูลที่เพิ่มเข้าไป

5.2 ตรวจแก้ข้อมูลในระบบเดิมของแหล่งข้อมูลหรือในลำดับงานของการแปลง ซึ่งควรเก็บข้อมูลในการตรวจแก้ไว้ในข้อมูลที่เกี่ยวข้องกับการแปลงข้อมูลด้วย

5.3 ตรวจสอบค่าของข้อมูลให้ถูกต้องในกระบวนการรวบรวมข้อมูล

5.4 ตรวจสอบผลรวมของข้อมูลหลังจากย้ายข้อมูลลงสู่คลังข้อมูลย่อย

(8) องค์ประกอบทางกายภาพของคลังข้อมูล

ตารางข้อเท็จจริง เป็นตารางศูนย์กลางที่ใช้ในคลังข้อมูลและคลังข้อมูลย่อย ที่จะเก็บค่าคำนวณที่วัดได้และรายละเอียดสำคัญในเชิงธุรกิจ ตารางข้อเท็จจริงจะเก็บค่าคำนวณที่เป็นตัวเลข เพราะค่าที่ได้จะใช้เป็นค่าพื้นฐานในการคำนวณร่วมกัน ระหว่างกุญแจสำคัญของตารางข้อเท็จจริง

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

และกุญแจสำคัญของตารางไดเมนชั่น

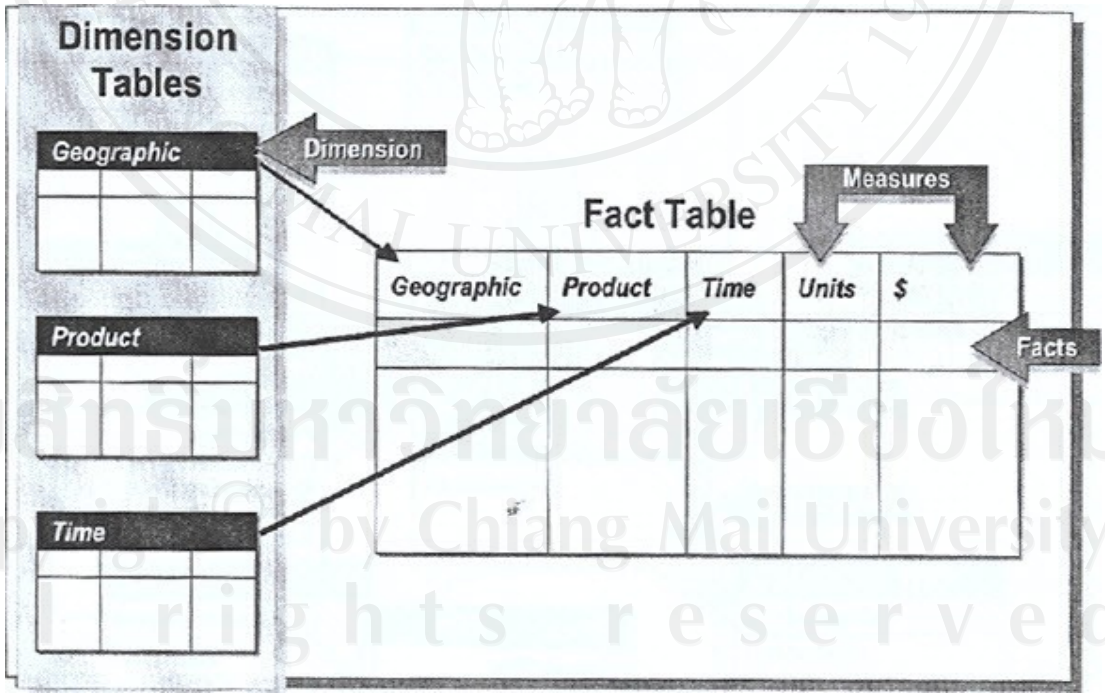
1. ตารางข้อเท็จจริงมีลักษณะดังนี้

- มีความสัมพันธ์แบบ Many To One กับตารางไดเมนชั่น
- เก็บข้อมูลที่เป็นตัวเลขที่สำคัญ
- มีกุญแจสำรองที่สัมพันธ์กับตารางไดเมนชั่นได้หลายตาราง
- มีความสำคัญในแถวมากกว่า ตารางไดเมนชั่น

2. ตารางไดเมนชั่น คือ ตารางในคลังข้อมูล และคลังข้อมูลย่อยที่อธิบายข้อมูลต่าง ๆ ในตาราง จะประกอบด้วยรายละเอียดของข้อมูล เช่น ชื่อและนามสกุลผู้ป่วย ที่อยู่ เป็นต้น

ตารางไดเมนชั่นมีลักษณะดังนี้

- มีความสัมพันธ์แบบ แบบ One to Many กับตารางข้อเท็จจริง
- เป็นรายละเอียดข้อมูลหลักที่ช่วยอธิบายให้ผู้ใช้ระบบเข้าใจ
- มี 1 กุญแจสำคัญ
- สามารถแก้ไขข้อมูลในตารางไดเมนชั่นได้โดยตรง
- มีความสำคัญในแถวน้อยกว่าตารางข้อเท็จจริง



รูปที่ 2.6 องค์ประกอบของตารางในคลังข้อมูล

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

(9) การวิเคราะห์ข้อมูลในคลังข้อมูล

ในการวิเคราะห์ข้อมูลสามารถจำแนกได้ 3 แบบ ดังต่อไปนี้

1. การสร้างรายงาน (Query and Reporting)

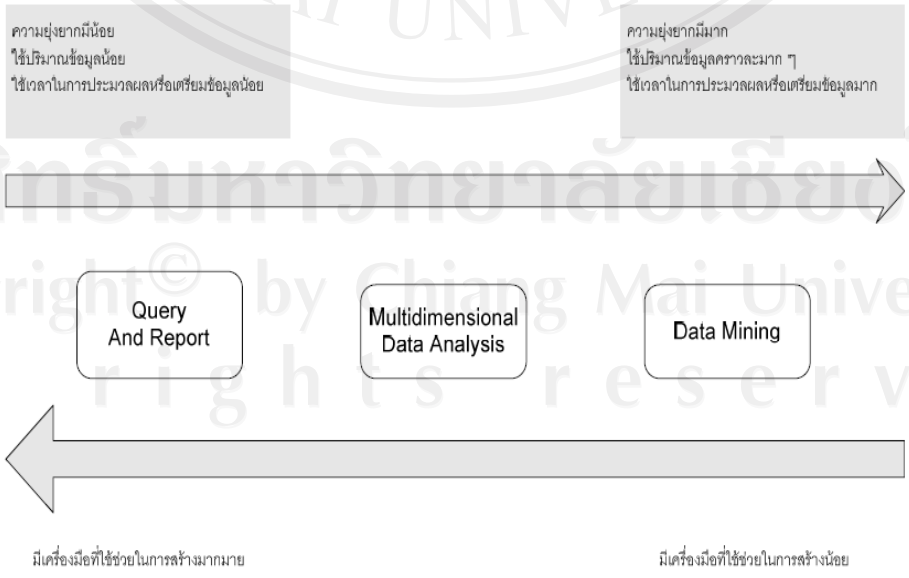
เป็นการใช้โปรแกรม หรือระบบที่เรียกว่า ระบบสร้างรายงาน (Report Generator) เพื่อรับข้อมูลที่เกิดจากการปฏิบัติงานในระบบ มาประมวลผลอย่างใดอย่างหนึ่ง เพื่อให้เกิดผลลัพธ์ เพื่อประโยชน์ในการตัดสินใจตามที่ต้องการ

2. การวิเคราะห์ข้อมูลแบบหลายมิติ (Multidimensional Data Analysis)

เป็นการวิเคราะห์ข้อมูลแบบหลายมิติ โดยใช้โปรแกรมหรือระบบที่เรียกว่าระบบสร้างการประมวลผลเชิงวิเคราะห์แบบออนไลน์ (Online Analytical Processing Generator หรือ OLAP Generator) เพื่อนำข้อมูลสำหรับการวิเคราะห์ที่ได้จัดเตรียมไว้ล่วงหน้า มาประมวลผลอย่างใดอย่างหนึ่ง เพื่อให้เกิดผลลัพธ์หรือมีค่า (Measures) ที่มีหลากหลายมุมมอง (Dimension) เพื่อการเตรียมพร้อมสำหรับการหยิบไปใช้งานได้ทันทีในอนาคต โดยไม่ต้องประมวลผล

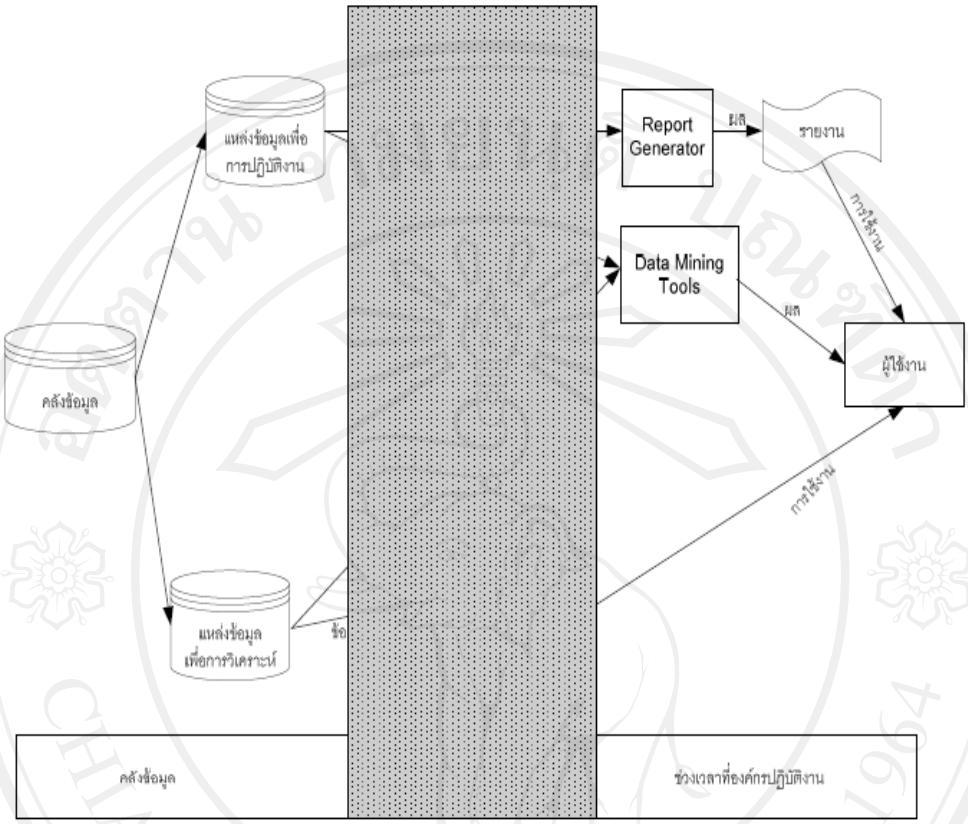
3. การทำเหมืองข้อมูล (Data Mining)

เป็นการใช้โปรแกรมที่มีความชาญฉลาดในระดับหนึ่งที่เรียกว่า Data Mining Tools เพื่อการวิเคราะห์ค้นหาหรือสร้างความรู้ใหม่ (Discovered Knowledge) ซึ่งไม่เคยมีมาก่อน ในการทำ Query and Reporting และ Multidimensional Data Analysis เราสามารถเปรียบเทียบการวิเคราะห์ข้อมูลทั้ง 3 แบบ ในแง่ของความยากง่าย และเครื่องมือที่ใช้สำหรับการวิเคราะห์ข้อมูลได้ ดังรูปที่ 2.7 และ 2.8



รูปที่ 2.7 แสดงการเปรียบเทียบการวิเคราะห์ข้อมูลทั้ง 3 แบบในแง่ต่าง ๆ

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved



รูปที่ 2.8 แสดงการวิเคราะห์ข้อมูลในคลังข้อมูล

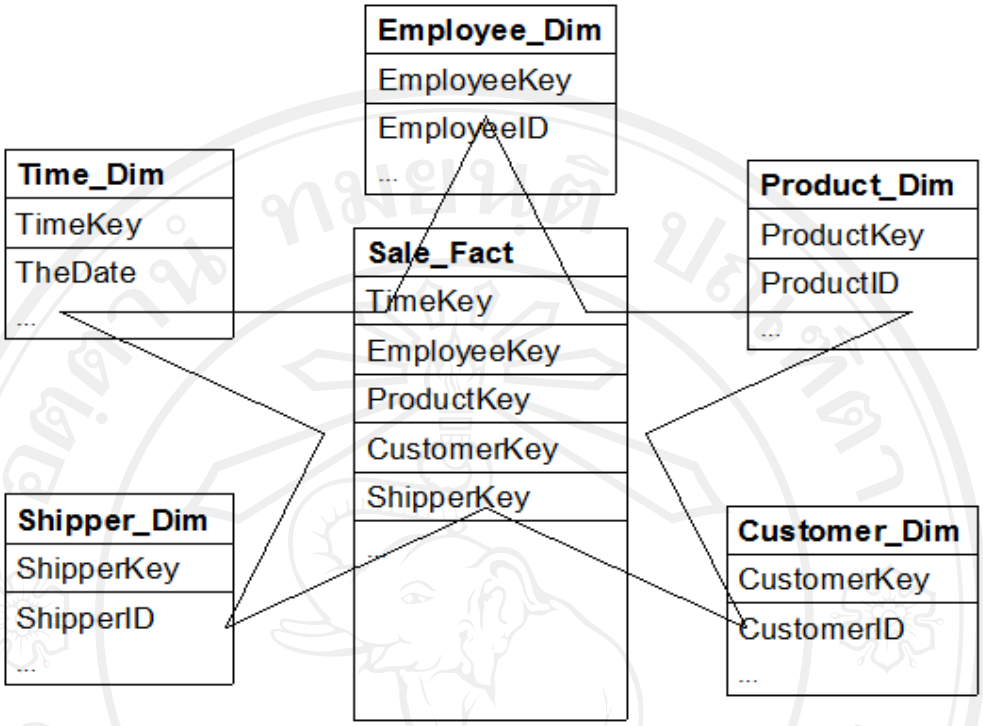
จากรูปที่ 2.8 จะเห็นว่าการเตรียมข้อมูลแบบ Multidimensional เพื่อการใช้งานนั้น มักทำในเวลาที่องค์กรไม่ได้ปฏิบัติงาน เช่น ในเวลาค่ำหรือหลังเลิกงาน ถ้าไม่มีข้อผิดพลาดใด ๆ ในตอนเช้า ผู้วิเคราะห์สามารถนำเอา OLAP ซึ่งเป็นผลจากการประมวลผลดังกล่าวไปใช้งานได้เลย ต่างจากรายงานตามปกติ ซึ่งต้องอาศัยการประมวลผลในเวลาทำงาน หากมีข้อมูลที่ต้องใช้จำนวนมากก็จะเกิดความล่าช้า อย่างไรก็ตาม หากมีการเปลี่ยนแปลงข้อมูลในเวลาทำงาน การออกรายงานตามปกติ จะสามารถแสดงผลจากข้อมูลในปัจจุบัน (Real Time Report) ได้ ซึ่งเป็นสิ่งที่ OLAP ไม่สามารถจะให้ได้

(10) เค้าร่างของคลังข้อมูล

1. Star Schema

เป็นเทคนิคที่ใช้ multidimensional model โดย data warehouse จะมี 1 star schema

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved



รูปที่ 2.9 Star Schema

- ลักษณะของ star schema

star schema มีลักษณะที่มี fact table อยู่ตรงกลางล้อมรอบด้วยหลาย ๆ dimension table ที่เก็บรายละเอียดของ fact ที่ไม่ใช่ normalized center

- ประโยชน์ของการใช้ star schema

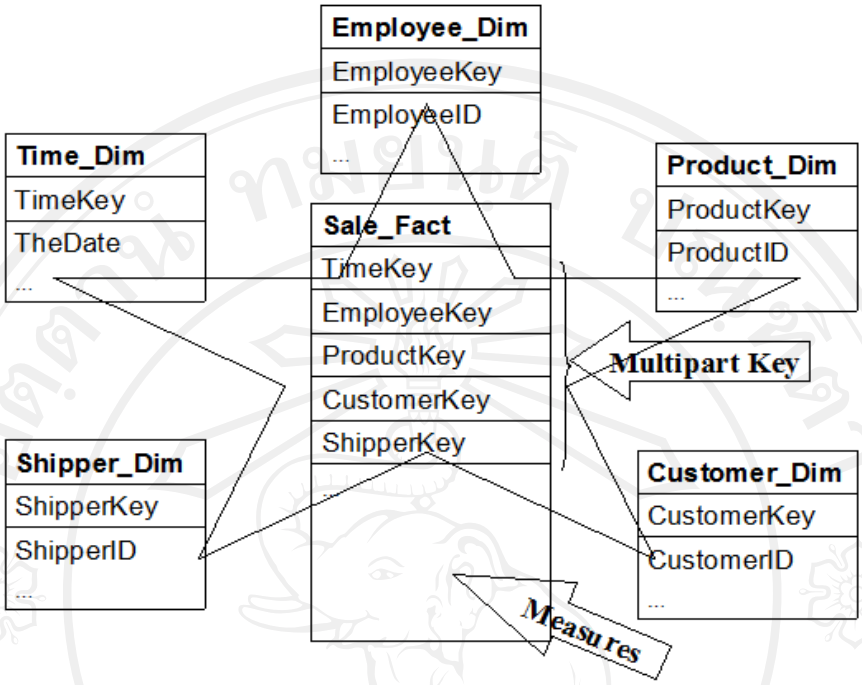
star schema จะประกอบไปด้วย set ของตารางที่ยังไม่ได้ normalized ถ้าใช้ star schema ในการออกแบบฐานข้อมูล เราสามารถเพิ่มประสิทธิภาพของการวิเคราะห์ในส่วนของ dimension ใน relational environment เราสามารถที่จะวิเคราะห์ข้อมูลได้โดยตรงจาก data mart โดยไม่ใช้ OLAP แต่ถ้าใช้จะทำให้

- สามารถวิเคราะห์ได้จำนวนมาก
- มีประสิทธิภาพใน query สูง
- สามารถวิเคราะห์ข้อมูลได้รวดเร็วกว่า Schema อื่น ๆ

- องค์ประกอบของ Star Schema

ส่วนประกอบของ star schema มี dimension table , fact table , dimensional key

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved



รูปที่ 2.10 องค์ประกอบของ Star Schema

- Dimension Tables

จะประกอบด้วนรายละเอียดของข้อมูล เช่น cust_name, title และ address Dimension table มีลักษณะดังนี้

- มี relationship ของ one to many กับ fact table
- ความสำคัญใน row น้อยกว่า fact table
- เป็นรายละเอียดข้อมูลหลักที่ช่วยอธิบาย dimension ให้ผู้ใช้ระบบเข้าใจ
- จะเก็บ multiple column ที่ใช้ในการประกาศ dimension hierarchies
- มี 1 primary (dimensional) key
- ถ้าข้อมูลมีการเปลี่ยนแปลงสามารถแก้ไขข้อมูลใน dimension tables ได้โดยตรง

- Fact table

จะเก็บ Measures ซึ่งจะเป็นตัวเลข เพราะค่าของมันเป็นพื้นฐานในการคำนวณร่วมกันระหว่าง Primary key ของ fact table (เรียก Multipart keys function) และ dimension key ของ dimension table (ได้มาจาก multipart keys ของ fact table) Fact table มีลักษณะดังนี้

- มีความสัมพันธ์แบบ many to one กับ dimension
- มี row มากกว่า dimension table

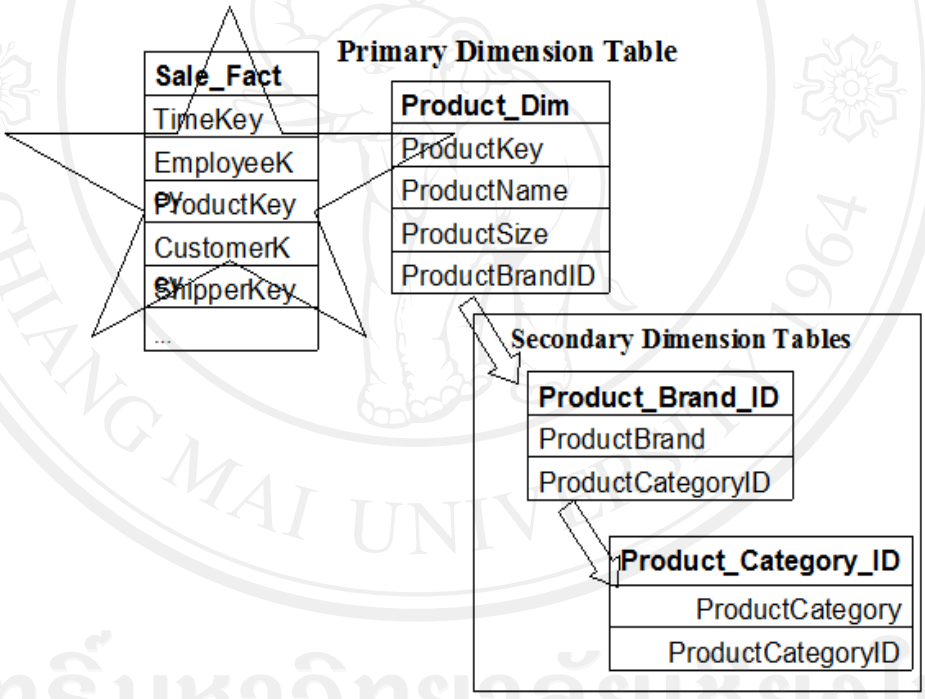
ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

- เก็บข้อมูลที่เป็นตัวเลขที่สำคัญ
- มี multiple foreign keys ที่สัมพันธ์กับ dimension table
- เก็บข้อมูลที่สำคัญ

Dimension key เป็น foreign keys ที่เก็บไว้ใน fact table และ dimension key จะ map กับ primary key ใน dimension table เราสามารถใช้ surrogate keys เป็น Primary key ของ dimension table

2. Snowflake schema

Snowflake schema มีความแตกต่างจาก Star schema ตรงที่ dimension table จะเก็บข้อมูลจะเก็บข้อมูลที่อยู่ในรูป normal form



รูปที่ 2.11 Snowflake schema

จาก star schema เราอาจจะเกิดปัญหาในการ design ได้ snowflake จึงเกิดขึ้นมา เนื่องจากปัจจัยดังนี้

- สภาวะทางธุรกิจ
- การ design ไม่สามารถ implement โดยใช้ star schema

ในบางโอกาส เราสามารถใช้ snowflake schema ได้ แต่ก็ไม่เสมอไป เนื่องจาก snowflake

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

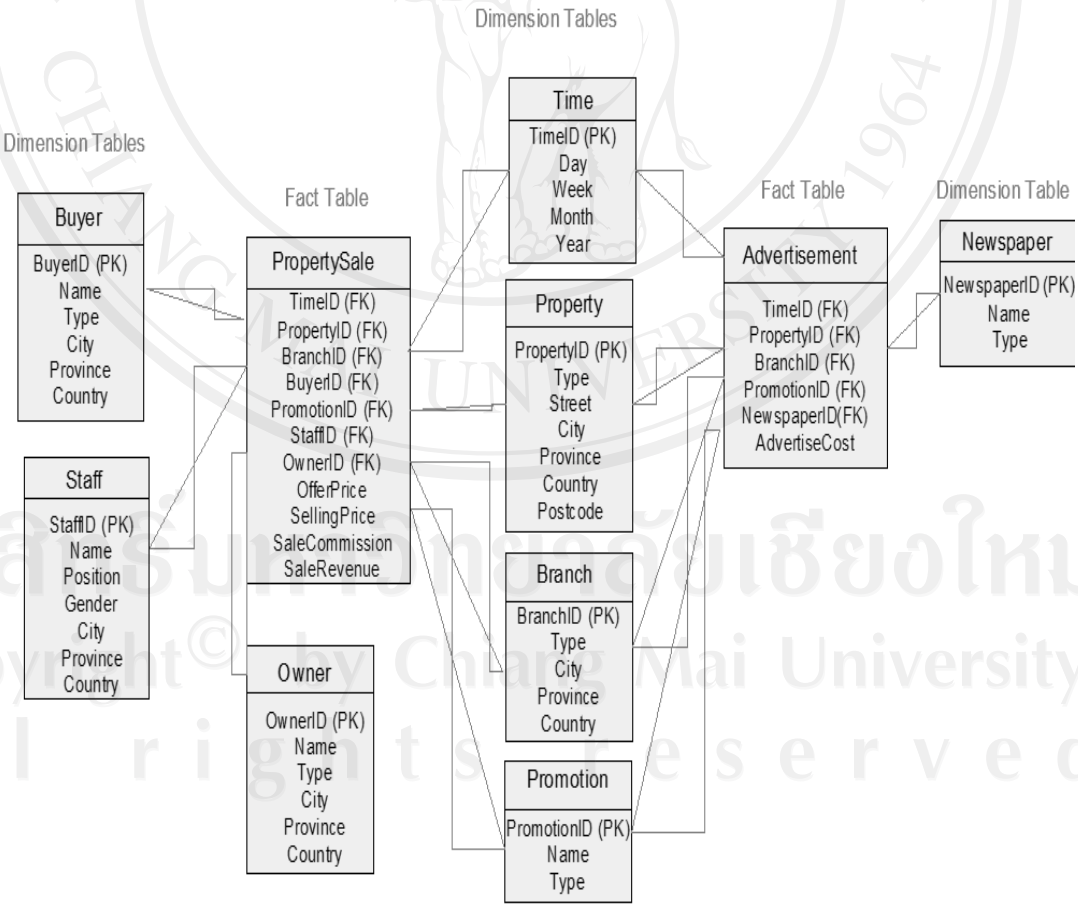
schema ไม่สามารถใช้งานกับกฎ multidimensional model

ลักษณะของ Snowflake schema มีดังนี้

- มีหลาย ๆ ตาราง
- มี primary dimension table เดียวที่สามารถ join กับ fact table
- มี secondary dimension table ที่ join กับ
  - Primary dimension table
  - Secondary dimension table เดียวกันที่อยู่ในระดับที่สูงกว่า
- Primary dimension table มีขนาดเล็กกว่าของ star schema

3. Constellation schema หรือ Galaxy schema

ตารางข้อมูลหลายตารางอยู่รวมกัน โดยมีการเชื่อมโยงไปยังมิติที่ใช้ร่วมกัน ซึ่งมีลักษณะเหมือนกลุ่มดาวหลายดวง เราอาจเรียกว่า เค้าร่างจักรวาล (galaxy schema) ซึ่งมี Fact เป็นแบบหลายตารางและมีการใช้ dimension ร่วมกัน

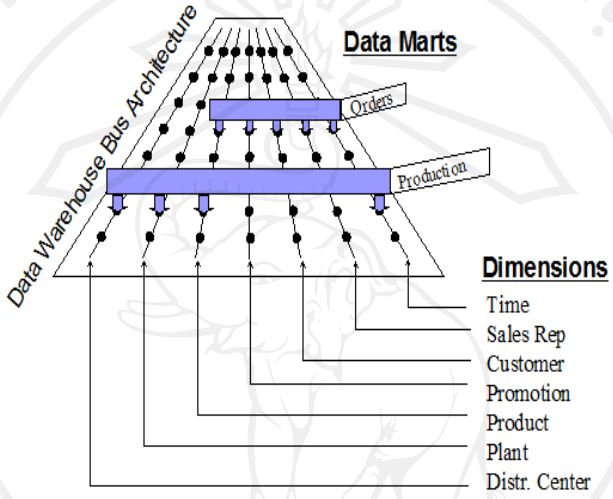


รูปที่ 2.12 Constellation schema

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

ข้อดีของการใช้ dimension ร่วมกัน

- แน่ใจได้ว่าในแต่ละรายงานจะออกมาสอดคล้องกัน
- สามารถสร้างดาต้ามาร์ทในเวลาต่างๆ กันได้
- สามารถเข้าถึงดาต้ามาร์ทโดยผู้พัฒนากลุ่มอื่นๆ
- สามารถรวบรวมดาต้ามาร์ทหลายๆ อันเข้าด้วยกัน
- สามารถออกแบบคลังข้อมูลร่วมกันได้



รูปที่ 2.13 แสดงการ Conformed Dimensions

4. การเลือก Schema

การคัดเลือกรูปแบบที่ใช้ในการออกแบบ Data Warehouse

Star Schema วิธีนี้จะช่วยเพิ่มความสามารถในการ query โดยลดปริมาณ data ที่อ่านจาก disk การวิเคราะห์การ queries ข้อมูลใน dimension table ที่เล็กจะใช้ dimension key จาก index ใน central fact table เป็นการลดจำนวนข้อมูลที่ต้องตรวจสอบ ซึ่งมีข้อดี คือ มีจำนวนของตารางน้อย สร้าง query ง่าย และเร็ว

Snowflake Schema จะใช้ในทางธุรกิจ แต่จะไม่แนะนำให้เลือกวิธีนี้ เนื่องจากวิธีนี้มีความยุ่งยากในการดูแล การเก็บข้อมูล ผู้ใช้ระบบไม่สามารถรู้ได้ว่าข้อมูลอยู่ที่ไหน อีกทั้งการเข้าถึงข้อมูลยังช้าอีกด้วย

Constellation schema หรือ Galaxy schema ใช้ในองค์กรธุรกิจขนาดใหญ่ หรือระดับกระทรวง ที่มีระบบฐานข้อมูลหลากหลายและต้องการประมวลผลที่ซับซ้อนและวิธีนี้มีความยุ่งยากมากที่สุดในที่กล่าวมา

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

2.1.5 การประมวลผลเชิงวิเคราะห์แบบออนไลน์ (Online Analytical Processing : OLAP)

การวิเคราะห์การประมวลผลแบบออนไลน์ คือ เทคโนโลยีที่ใช้ในการจัดการข้อมูลจากคลังข้อมูล เพื่อนำไปใช้ในการวิเคราะห์ และตัดสินใจทางธุรกิจอย่างมีประสิทธิภาพ สามารถค้นหาคำตอบที่ต้องการและสามารถแก้ปัญหาที่มีความซับซ้อน โดยใช้ระยะเวลาสั้น ๆ ในการวิเคราะห์ การประมวลผลแบบออนไลน์ (On - Line Analytical Processing) เป็นกระบวนการประมวลผลข้อมูลที่ทำให้ผู้ใช้สามารถเลือกดูข้อมูล จากจุดการมองที่แตกต่างกันได้ง่าย เริ่มต้นจากการสรุปรวมข้อมูลก่อนที่จะกระทำการสร้างรายงานและการสร้างแผนภูมิจากข้อมูลที่เลือก กระบวนการนี้ข้อมูลจะถูกเก็บในรูปแบบอะเรย์หลายมิติหรือที่เรียกง่าย ๆ ว่า Data Cube เพื่อให้การค้นหาข้อมูลมีประสิทธิภาพมากขึ้น นอกจากนี้ยังมีความสามารถในการบริหารจัดการ วิเคราะห์ข้อมูลได้อย่างรวดเร็ว ความถูกต้อง ตอบสนองการเข้าถึงข้อมูลได้หลาย ๆ มุมมอง และเข้ามามีส่วนช่วยในการวิเคราะห์ข้อมูลเพื่อใช้ในการปรับปรุงกระบวนการทางธุรกิจ ด้วยความสามารถในการวิเคราะห์ที่ซับซ้อน อันได้แก่ การเจาะลึก (Drill - down) ไปถึงรายละเอียดในระดับทรานแซกชัน(Transaction) ผู้ใช้งานสามารถใช้งานข้อมูลตาม Cube ที่ถูกกำหนดไว้ (Cube คือ โมเดลข้อมูลของการวิเคราะห์ การประมวลแบบออนไลน์ ซึ่งเป็นโครงสร้างหลายมิติหรือ Multidimensional Structure เปรียบเสมือนกับรูปลูกบาศก์ที่มีมุมมองหลากหลาย แต่ละมุมมองทำให้เกิดการคิวรีข้อมูลจากคลังข้อมูลได้หลากหลายแบบ)

การสร้างระบบคลังข้อมูลจะเป็นการรวบรวม และจัดเก็บข้อมูลที่มาจากหลาย ๆ แหล่ง ให้สามารถสนองตอบต่อการหาคำตอบของคิวรี จากผลลัพธ์ของการดำเนินงานที่ผ่านมา ตลอดจนการคาดเดาข้อมูลที่เกิดขึ้นในอนาคต ซึ่งเป็นการทำงานแบบ OLAP เช่น Decision Support (DSS)และ Executive Information System (EIS) เป็นต้น การออกแบบการจัดเก็บฐานข้อมูลของ OLAP จะเป็นการกระจายข้อมูลด้วยการทำ De - Normaliztion จากตารางที่เคยทำ Normalize ถึงแม้ว่าจะทำให้ข้อมูลมีความซ้ำซ้อนเกิดขึ้น แต่จะทำให้เกิดปัญหาตามมา เพราะข้อมูลอยู่ในลักษณะของการอ่านอย่างเดียวและเพิ่มประสิทธิภาพการคิวรีที่รวดเร็วยิ่งขึ้น เนื่องจากลดการ JION ของตารางออกไป

(1) ความหมายของการประมวลผลเชิงวิเคราะห์แบบออนไลน์

กระบวนการประมวลผลข้อมูลทางคอมพิวเตอร์ ที่ช่วยให้สามารถวิเคราะห์ข้อมูลในมิติต่าง ๆ Multidimensional Data Analysis) ของข้อมูลได้ง่ายยิ่งขึ้น (การติดตั้ง OLAP ขึ้นใช้งานส่วนใหญ่จะใช้เทคโนโลยีฐานข้อมูลหลายมิติ)

(2) วัตถุประสงค์

โดยทั่วไปมีลักษณะดังต่อไปนี้

- สนองตอบกับการแสดง และนำเสนอข้อมูลได้เป็นอย่างดี

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

- สร้างความยืดหยุ่น และง่ายต่อการทำรายงาน รวมทั้ง ad hoc queries ต่างๆ
- ทำการวิเคราะห์ข้อมูลได้ในทุกมุมมองขององค์ประกอบที่เกี่ยวข้องทั้งหมด
- ช่วยคาดคะเน หรือทำนายผลที่จะตามมาในภาคหน้า

(3) เทคโนโลยีฐานข้อมูลหลายมิติ (Multidimensional Database)

เป็นฐานข้อมูลที่ใช้ในการจัดเก็บและจัดการข้อมูล ให้สามารถแสดงผลตามมิติของข้อมูลต่าง ๆ ได้พัฒนาต่อจากฐานข้อมูลเชิงสัมพันธ์ ให้มีความสามารถมากขึ้นเพื่อรองรับการประมวลผลเชิงวิเคราะห์ หรือ OLAP นั่นเอง ในการแสดงข้อมูล ถ้านำเสนอหลายมิติหรือหลายมุมมองจะทำให้ผู้ใช้เกิดความเข้าใจได้ง่ายกว่า เช่น การนำเสนอข้อมูลจากการสอบถามว่า “ ให้แสดงผลรวมของยอดขายในแต่ละจังหวัดและในแต่ละไตรมาส ” ซึ่งถ้าหากเป็นการจัดการฐานข้อมูลเชิงสัมพันธ์ธรรมดา อาจแสดงได้ในรูปของตาราง ดังต่อไปนี้

จังหวัด	ไตรมาส	ยอดขาย
กรุงเทพฯ	1	750,000
กรุงเทพฯ	2	800,000
กรุงเทพฯ	3	825,000
กรุงเทพฯ	4	700,000
เชียงใหม่	1	550,000
เชียงใหม่	2	500,000
เชียงใหม่	3	650,000
เชียงใหม่	4	700,000
สุราษฎร์ธานี	1	450,00
สุราษฎร์ธานี	2	452,250
สุราษฎร์ธานี	3	425,500
สุราษฎร์ธานี	4	400,000

ตารางที่ 2.1 แสดงฐานข้อมูลเชิงสัมพันธ์แบบธรรมดา

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

จะเห็นว่าการนำเสนอข้อมูล ตามการสอบถามที่ได้รับจากระบบจัดการฐานข้อมูลเชิงสัมพันธ์ธรรมดาจะแสดงเป็น 3 Field และหลาย Record แต่หากแสดงแบบหลายมิติจะดูง่ายขึ้น ดังนี้

จังหวัด / ไตรมาส	กรุงเทพฯ	เชียงใหม่	สุราษฎร์ธานี
1	750,000	550,000	450,000
2	800,000	500,000	452,250
3	825,000	650,000	425,500
4	700,000	700,000	400,000

ตารางที่ 2.2 แสดงข้อมูลแบบหลายมิติ

จะเห็นว่าตารางข้างบนมีการนำเสนอข้อมูลแบบหลายมิติ คือ การแบ่งตามไตรมาส และจังหวัดคล้ายกับตารางเมตริกซ์ 2 มิติ หากเป็นการแสดงข้อมูลตามการสอบถาม “ ให้แสดงข้อมูลรวมของยอกขายในกรุงเทพเฉพาะในไตรมาสที่ 1 ” ผลลัพธ์ที่ได้จากการสอบถามนี้ คือค่ายอดขายค่าเดียว คือ 750,000 บาท จึงไม่จำเป็นต้องแสดงในรูปแบบหลายมิติก็ได้ หากผู้ใช้ต้องการให้แสดงผลข้อมูลที่ก่อนการแสดงผล จะต้องมีการคำนวณหรือรวมข้อมูลหลายตัว หรือฐานข้อมูลมีขนาดใหญ่ประกอบด้วยข้อมูลจังหวัดมาก การแสดงผลจากฐานข้อมูลเชิงสัมพันธ์แบบธรรมดาจะช้าลง

ในการจะพัฒนาคลังข้อมูล จะต้องทราบถึงองค์ประกอบของคลังข้อมูลว่า ส่วนใดบ้างที่ทำหน้าที่ในการเข้าถึง (Access) ข้อมูลที่แตกต่างกันในแของแหล่งข้อมูล การทำให้ข้อมูลไม่มีความซ้ำซ้อน (Cleansing) การแปลงข้อมูลจากแหล่งข้อมูล ให้อยู่ในรูปแบบที่คลังข้อมูลต้องการ และการเก็บบันทึกข้อมูลให้อยู่ในรูปแบบที่ง่าย ต่อการนำไปใช้งาน (จัดกลุ่มตามการสืบค้นหรือรูปแบบที่สามารถเข้าใจได้ง่าย) ซึ่งในกรณีนี้ ถ้าหากเราจัดรูปแบบการแสดงผลจากการจัดการฐานข้อมูลแบบหลายมิติ จะทำให้รวดเร็วขึ้น ให้สังเกตตัวอย่างต่อไปนี้

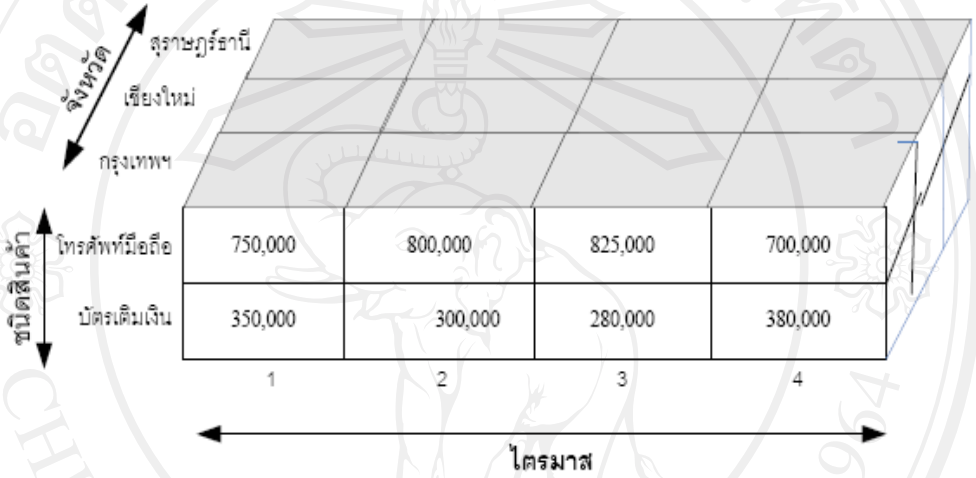
ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

สินค้า	จังหวัด	ไตรมาส	ยอดขาย
โทรศัพท์มือถือ	กรุงเทพฯ	1	750,000
บัตรเติมเงิน	กรุงเทพฯ	1	350,000
โทรศัพท์มือถือ	กรุงเทพฯ	2	800,000
บัตรเติมเงิน	กรุงเทพฯ	2	300,000
โทรศัพท์มือถือ	กรุงเทพฯ	3	825,000
บัตรเติมเงิน	กรุงเทพฯ	3	280,000
โทรศัพท์มือถือ	กรุงเทพฯ	4	700,000
บัตรเติมเงิน	กรุงเทพฯ	4	380,000
โทรศัพท์มือถือ	เชียงใหม่	1	550,000
บัตรเติมเงิน	เชียงใหม่	1	280,000
โทรศัพท์มือถือ	เชียงใหม่	2	500,000
บัตรเติมเงิน	เชียงใหม่	2	250,000
โทรศัพท์มือถือ	เชียงใหม่	3	650,000
บัตรเติมเงิน	เชียงใหม่	3	245,000
โทรศัพท์มือถือ	เชียงใหม่	4	700,000
บัตรเติมเงิน	เชียงใหม่	4	235,000
โทรศัพท์มือถือ	สุราษฎร์ธานี	1	450,000
บัตรเติมเงิน	สุราษฎร์ธานี	1	520,000
...	...	...	...

ตารางที่ 2.3 แสดงฐานข้อมูลเชิงสัมพันธ์แบบธรรมดา

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

จากตารางข้างต้น ข้อมูลที่เพิ่มขึ้นมาคือ ชนิดสินค้า หากมีการสอบถามข้อมูล “ยอดขายในแต่ละจังหวัดของสินค้าแต่ละชนิดและในแต่ละไตรมาส” การแสดงผลจากฐานข้อมูลเชิงสัมพันธ์แบบธรรมดา (ดังตารางข้างบน) จะทำให้ผู้ใช้อ่านข้อมูลได้ลำบากและกระบวนการค้นหาข้อมูลก็จะช้าลงด้วย แต่หากเป็นการแสดงผลแบบหลายมิติ จะมีลักษณะเป็นรูป “ลูกบาศก์หรือลูกเต๋า (Cube)” แทนข้อมูลแต่ละค่า เรียกว่า “เซลล์” แล้วนำมาเรียงต่อกันเป็นตาราง 3 มิติ ได้แก่ มิติของจังหวัด ชนิสินค้า และไตรมาส ดังรูปที่ 2.12



ตารางที่ 2.4 แสดงข้อมูลแบบหลายมิติเป็นแบบลูกบาศก์

(4) การดำเนินการกับการประมวลผลเชิงวิเคราะห์แบบออนไลน์

เนื่องจากฐานข้อมูลแบบหลายมิติถูกค้นคิดขึ้นมา เพื่อรองรับการเรียกใช้ข้อมูลเชิงวิเคราะห์ ซึ่งจะต้องมีการดำเนินการกับข้อมูลในรูปแบบต่าง ๆ ดังนี้

- การรวบรวมให้เป็นหนึ่งเดียว (Consolidation หรือ Roll Up)

การรวบรวมให้เป็นหนึ่งเดียว จะเกี่ยวข้องกับการรวมกันของข้อมูล เกี่ยวข้องกับการจับกลุ่มที่ซับซ้อนของข้อมูลและการเกี่ยวข้องกับความสัมพันธ์ภายในข้อมูล อย่างเช่น สำนักงานขายสามารถรวมยอดขายในเขตพื้นที่และต่างเขตพื้นที่ สามารถรวมยอดขายในเขตภูมิภาคได้ด้วย

Consolidation หรือ Roll Up หมายถึง การการเปลี่ยนแปลงระดับความละเอียดของการพิจารณาข้อมูล จากระดับที่ละเอียดขึ้นไปสู่ระดับที่หยาบมากขึ้น

- การเจาะลึก (Drill – Down)

การประมวลผลเชิงวิเคราะห์แบบออนไลน์ สามารถเข้าไปในทิศทางตรงกันข้าม(Reverse

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

Direction) และแสดงรายละเอียดของข้อมูลโดยอัตโนมัติ ที่เรียกว่า เจาะลึก เช่น การขายผลิตภัณฑ์เฉพาะหรือพนักงานขาย (Sales Reps) สามารถเข้าถึงข้อมูลในพื้นที่การขายได้อย่างง่ายดาย

Drill Down หมายถึง การเพิ่มความละเอียดในการพิจารณาข้อมูล จากระดับที่หยาบไปสู่ระดับที่ละเอียดมากขึ้น

- การแบ่งส่วนและการสุ่ม (Slicing and Dicing)

การแบ่งส่วนและการสุ่ม อ้างอิงถึงความสามารถในการตรวจดูฐานข้อมูล จากจุดตรวจสอบที่แตกต่างกัน ในการแบ่งส่วนออกหนึ่งส่วนของฐานข้อมูลในการขาย อาจจะแสดงรายการขายของสินค้า ประเภทที่อยู่ในขอบเขตทั้งหมด ในการแบ่งส่วนอื่น ๆ อาจจะมีการแสดงรายการขายจากช่องทางการขายของสินค้าแต่ละประเภททั้งหมด มักจะใช้เพื่อวิเคราะห์แนวโน้มและการค้นหารูปแบบ

Slice เป็นการเลือกพิจารณาผลลัพธ์บางส่วนที่เราสนใจ โดยการเลือกเฉพาะค่าที่ถูกกำกับด้วยข้อมูลบางค่าของแต่ละมิติเท่านั้น

Dice เป็นกระบวนการพลิกแกนหรือมิติข้อมูล ให้ตรงกับตามความต้องการของผู้ใช้งาน

(5) ประโยชน์ของการประมวลผลเชิงวิเคราะห์แบบออนไลน์

- ช่วยในการวิเคราะห์ เพื่อเปรียบเทียบข้อมูลในมุมมองต่าง ๆ ซึ่งเป็นการเพิ่มประสิทธิภาพในการตัดสินใจ

- ช่วยให้ผู้ใช้ สามารถคัดเลือกข้อมูลสำหรับตัดสินใจได้อย่างมีประสิทธิภาพ

- ช่วยให้ผู้ใช้แต่ละคนสามารถสร้างข้อมูลตามมุมมองของตนเองได้ เพื่อนำข้อมูลเหล่านั้นไปใช้เฉพาะด้าน

- สามารถสอบถามข้อมูลได้อย่างรวดเร็ว แม้ในฐานข้อมูลขนาดใหญ่และซับซ้อน

2.2 งานวิจัยที่เกี่ยวข้อง

2.2.1 การพัฒนาระบบคลังข้อมูล

เบญจมาศ เต็มอุดมและดร.ภัทรชัย สลิตโรจน์วงค์ ทำการศึกษาวิจัยเรื่อง“ การพัฒนาระบบคลังข้อมูล” โดยใช้แนวคิด 8 ประการหรือ The Big Eight ของโพ(Poe) และใช้ระเบียบวิธี 9 ขั้นหรือ Nine - Step Methodology ของ Kimball มาช่วยในการพัฒนาคลังข้อมูล โดยการรวบรวมข้อมูลจากฐานข้อมูลของระบบงานปฏิบัติงานประจำวันขององค์กร แล้วนำมาแปลงข้อมูลให้อยู่ในรูปที่เหมาะสมในการเก็บ และสะดวกในการใช้งาน ต่อจากนั้นจึงนำข้อมูลนั้นเข้าไปเก็บในคลังข้อมูล

การพัฒนาหรือการสร้างคลังข้อมูลมาใช้ในองค์กร ต้องพิจารณาถึงองค์ประกอบที่จำเป็น

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

ในการสร้างให้เหมาะสมด้วย ทั้งนี้เพื่อให้เกิดความคุ้มค่าในการลงทุนและให้เกิดประโยชน์สูงสุดต่อองค์กร ถึงแม้ว่าเทคโนโลยีคลังข้อมูล จะให้ประสิทธิภาพในการใช้ข้อมูลอย่างมากก็ตาม แต่สิ่งที่ต้องคำนึงถึงด้วยคือ ทรัพยากรที่องค์กรจะต้องทุ่มลงไปในการพัฒนาระบบนี้ มีทั้งที่สามารถวัดเป็นตัวเงินได้และที่ไม่สามารถตีค่าออกมาเป็นตัวเงินได้ นอกจากนี้ ปัญหาในระหว่างการพัฒนาที่อาจจะเกิดขึ้น จนองค์กรไม่สามารถจะพัฒนาระบบนี้จนสำเร็จและนำมาใช้งานได้ เกิดการลงทุนที่สูญเปล่า ดังนั้นจึงต้องมีการวางแผนควบคุมและจัดการให้รอบคอบ

2.2.2 ระบบคลังข้อมูล

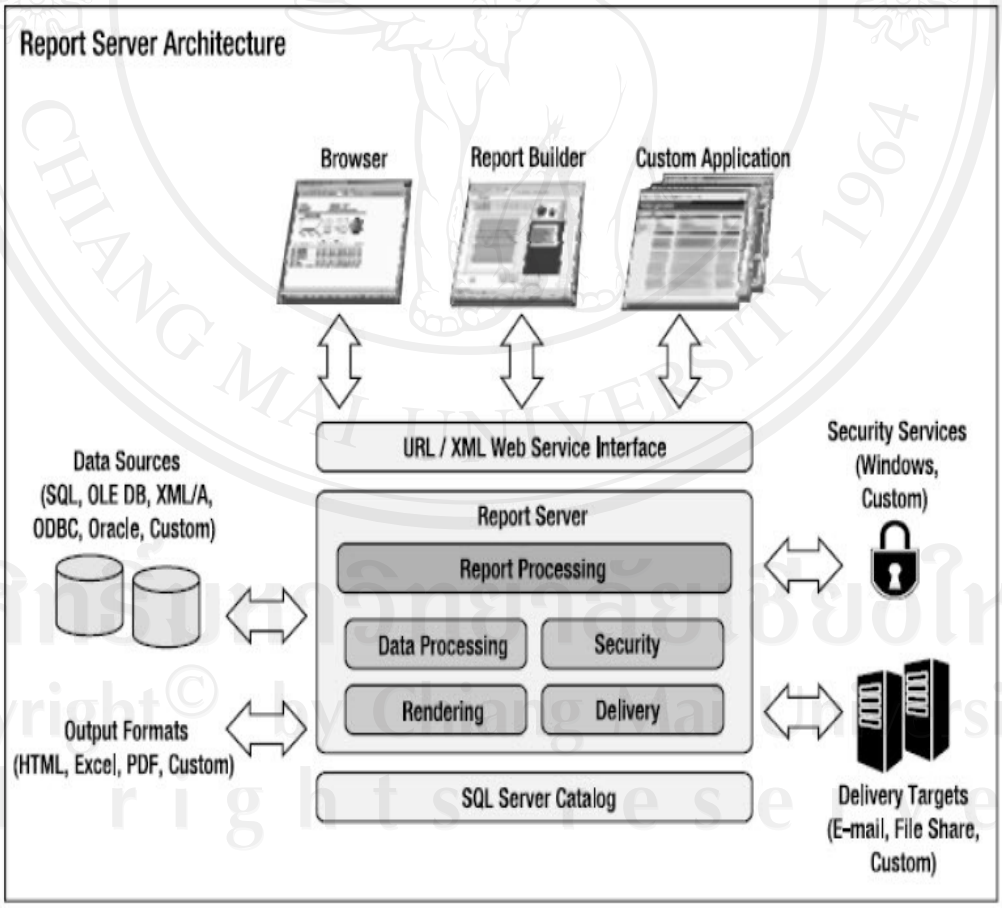
ศูนย์เทคโนโลยีสารสนเทศและการสื่อสาร กระทรวงสาธารณสุข ได้ทำการศึกษาวิจัยเรื่อง " ระบบคลังข้อมูล " คลังข้อมูล คือ ฐานข้อมูลขนาดใหญ่ ที่รวบรวมข้อมูลทั้งจากแหล่งข้อมูลภายใน และภายนอกองค์กร โดยมีรูปแบบและวัตถุประสงค์ของการจัดเก็บข้อมูลแตกต่างจากฐานข้อมูลปฏิบัติการทั่วไป การพัฒนาระบบคลังข้อมูล เริ่มต้นจากการออกแบบฐานข้อมูล ซึ่งวิธีการหนึ่ง เรียกว่า ระเบียบวิธี 9 ขั้นของ Kimball จะเน้นที่การออกแบบจากระบบงานย่อยหรือดาต้ามาร์ทของแต่ละระบบงานในองค์กรก่อน แล้วจึงนำส่วนย่อย ๆ ที่ได้นั้นมารวมเป็นระบบคลังข้อมูลขององค์กรต่อไป ทั้งนี้กระบวนการหนึ่งที่สำคัญมากในการพัฒนาระบบคลังข้อมูล คือ การนำข้อมูลจากแหล่งข้อมูลเข้าสู่ดาต้ามาร์ทของแต่ละระบบ หรือเรียกว่าการแปลงข้อมูล โดยเราจะต้องกำหนดการส่งข้อมูล รวบรวมหรือสร้างข้อมูลภายนอก วางแผนและสร้างรูทีนการแปลงข้อมูล และตรวจสอบความถูกต้องของข้อมูลที่ได้ก่อนนำเข้าสู่คลังข้อมูล เพื่อให้เป็นข้อมูลที่เหมาะสมที่จะนำไปวิเคราะห์ต่อไป

2.2.3 การพัฒนาระบบคลังข้อมูลและรายงานสำหรับธุรกิจการขนส่งและกระจายสินค้า

นางสาวกาญจนาท แสงสิงห์ ทำการศึกษาวิจัยเรื่อง " การพัฒนาระบบคลังข้อมูลและรายงานสำหรับธุรกิจการขนส่งและกระจายสินค้า " การปรับใช้ระบบงานไปสู่ยูนิกซ์หรือเมนเฟรมที่ใช้ในการกระจายสินค้าและพัสดุภายในประเทศของบริษัท เพื่อง่ายต่อการตรวจสอบปริมาณการส่งและรับสินค้าในระดับโลก แต่กลับก่อให้เกิดปัญหาในเรื่องการประสานงานของระบบกระจายสินค้าภายในประเทศ นั่นคือ แผนกปฏิบัติการไม่สามารถออกรายงานไปยังฝ่ายต่าง ๆ ได้อย่างทันท่วงที ซึ่งโดยปกติข้อมูลและระบบที่ใช้สำหรับแผนกปฏิบัติการรายวันนั้น ทำงานอยู่บนระบบยูนิกซ์หรือเมนเฟรม มีข้อจำกัดในเรื่องของการเข้าถึงข้อมูล มีความยุ่งยากซับซ้อน เพื่อให้ได้รายงานแต่ละรายงานที่ต้องการ ในการพัฒนาระบบคลังข้อมูลได้ถูกจัดทำขึ้น เพื่อแก้ปัญหาการทำรายงานของระบบบริหารการกระจายสินค้า โดยนำระบบมาพัฒนาเป็นระบบคลังข้อมูลและรายงานโดยสร้างคลังข้อมูลและรายงานสำหรับแผนกต่าง ๆ เพื่อให้มีข้อมูลที่นำไปวิเคราะห์ และเข้าถึงข้อมูลได้ง่าย โดยการสกัด แปลงรูปและนำเข้าข้อมูลไปไว้ในคลังข้อมูล ได้พัฒนาส่วนของการอ่าน

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

ผลรายงานบนหน้าเว็บและทดลองทำการจัดกลุ่มข้อมูลลูกค้า นอกจากนั้นแล้ว ยังมีส่วนที่ผู้ใช้งานสามารถนำข้อมูลออกไปประมวลผลยังโปรแกรมประยุกต์อื่น ๆ ได้ วิธีแก้ปัญหาการทำรายงานจากระบบกระจายสินค้า โดยการนำเอาข้อมูลจากกิจกรรมการดำเนินงานต่าง ๆ ซึ่งมาในรูปของข้อมูลงานปฏิบัติงาน จากระบบยูนิกซ์หรือเมนเฟรมมาทำการแปลงข้อมูล โ ดยใช้เทคนิคการพัฒนาคลังข้อมูลของ Berson และ Smith ที่เขียนในปี 1997 เพื่อพัฒนาให้เป็นรายงานและคลังข้อมูล โดยใช้เทคโนโลยีการวิเคราะห์งานบริการ (Analysis Service) ของระบบบริหารจัดการฐานข้อมูลไมโครซอฟต์เอสคิวแอล (MS SQL) ในส่วนกลาง โดยใช้เทคนิคการพัฒนาของ Robinson ที่เขียนในปี 2006 เพื่อช่วยลดปัญหาการขาดความต่อเนื่องของระบบงานในแต่ละฝ่าย และสร้างรายงานส่วนกลางที่ผู้ใช้ในระดับล่างสามารถเข้าถึงโดยง่าย และช่วยสนับสนุนผู้บริหารระดับกลางจนถึงระดับสูง (Decision Support) ให้ได้ข้อมูลที่ถูกต้องแม่นยำ และรวดเร็ว



รูปที่ 2.14 โครงสร้างของระบบรายงานในไมโครซอฟต์เอสคิวแอล

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved