

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

ในการค้นคว้าอิสระเรื่อง “ระบบสารสนเทศเพื่อใช้สนับสนุนการตัดสินใจของผู้บริหารในด้านการตลาดบริษัทสหพานิชเชียงใหม่จำกัด” นี้ผู้ค้นคว้าได้ศึกษาขั้นพื้นฐานจากเอกสารและงานวิจัยที่เกี่ยวข้องหลายเล่มด้วยกัน ซึ่งพอจะสรุปสาระสำคัญได้ดังนี้

2.1 ระบบคลังข้อมูล

ปัจจุบันนี้องค์กรจะประสบความสำเร็จต้องขึ้นอยู่กับปัจจัยหลายอย่าง และปัจจัยอย่างหนึ่งที่ทำให้องค์กรประสบความสำเร็จ คือ ข้อมูลที่มีอยู่และใช้ประจำวันหรือ Operational Database ซึ่งนับวัน จะมีแต่มากขึ้นจนเป็นปัญหาสำหรับองค์กรที่จะต้องจัดการเก็บข้อมูล ซึ่งข้อมูลเหล่านี้อาจเก็บรวบรวมอยู่ในรูปแบบต่างๆ ไม่ว่าจะเป็นเทปแม่เหล็ก ดิสก์เก็ต หรือในแผ่นซีดี นอกจากนี้ข้อมูลมากมายเหล่านี้ยังไม่เหมาะที่เราจะนำมาใช้ในการวิเคราะห์ทางธุรกิจ หรือช่วยผู้บริหารสำหรับการตัดสินใจ (Decision Support System) ทางธุรกิจได้ เพราะต้องใช้เวลาในการประมวลผลที่ นานพอสมควรและส่งผลกระทบไปถึงระบบการทำงานของเครื่องที่ใช้งานประจำวันอีกด้วย

2.2 วัตถุประสงค์ของการสร้างคลังข้อมูล

เป้าหมายของการสร้างคลังข้อมูลคือการแยกกลุ่มข้อมูลสารสนเทศที่ใช้ในการทำการวิเคราะห์ทางธุรกิจออกจากฐานข้อมูลที่ใช้งานประจำวัน มาเก็บอยู่ใน Relational Database Management Systems (RDBMS) ประสิทธิภาพสูง และทำให้การเรียกใช้ข้อมูลชุดนี้ทำได้ง่าย ยึดหยุ่น จากเครื่องมือที่อยู่บนเครื่องเดสก์ท็อปทั่วไป โดยลด off-loading เพิ่มกลไกการช่วยตัดสินใจ ปรับปรุงเวลาที่ตอบสนอง (response time) รวดเร็วยิ่งขึ้นอย่างมาก และผู้บริหารสามารถเรียกข้อมูลรายละเอียดที่จำเป็น ที่ถูกเก็บมา ก่อนหน้านี้ (historical data) มาใช้ช่วยในการตัดสินใจทางธุรกิจแม่นยำขึ้น

2.3 เป้าหมายในการสร้างคลังข้อมูลมีดังนี้

คลังข้อมูลทำให้สามารถเข้าถึงข้อมูลขององค์กรได้ ผู้จัดการและนักวิเคราะห์ขององค์กรสามารถเชื่อมต่อเข้าไปยังคลังข้อมูล จากเครื่องคอมพิวเตอร์ของตนได้ ซึ่งการเชื่อมต่อสามารถทำได้ทันที ตามความต้องการและด้วยประสิทธิภาพสูง เครื่องมือที่มีให้กับผู้จัดการและนักวิเคราะห์ใช้งานง่าย สามารถออกรายงานได้ด้วยการคลิกปุ่มเดียว

ข้อมูลในคลังข้อมูลมีความถูกต้องตรงกันหมด คำถามเดียวกันต้องได้รับคำตอบที่เหมือนกันเสมอไม่ว่าผู้ถามจะเป็นใคร ถามเวลาใดข้อมูลในคลังข้อมูลสามารถถูกวิเคราะห์จากหัวข้อในธุรกิจประเภทนั้นโดยแบ่งข้อมูล หรือรวมข้อมูลมาวิเคราะห์ตามต้องการคลังข้อมูลเป็นส่วนที่ผลิตข้อมูลจาก OLTP ข้อมูลไม่เพียงแต่ถูกรวบรวมมาไว้ที่ศูนย์กลางอย่างเดียว แต่จะถูกรวบรวมอย่างระมัดระวังจากแหล่งข้อมูลหลายๆแห่งนอกองค์กรด้วย แล้วมาปรับปรุงให้เหมาะสมกับการใช้งานเท่านั้น ถ้าข้อมูลเชื่อถือไม่ได้หรือไม่สมบูรณ์จะไม่ถูกอนุญาตให้นำไปใช้คุณภาพของข้อมูลในคลังข้อมูลเป็นตัวหลักคั่นให้สามารถทำการปรับปรุงเปลี่ยนธุรกิจได้

2.4 ประโยชน์ของระบบคลังข้อมูล

โดยทั่วไปแล้วข้อมูล Operational Database จะเก็บข้อมูลในรูปแบบ Transaction Systems เมื่อมีความต้องการข้อมูลในอันที่จะนำมาใช้ช่วยในการตัดสินใจก็จะประสบปัญหาต่างๆเช่น บุคลากรทางด้าน Information Systems จำเป็นต้องเรียกข้อมูลจากฐานข้อมูลขนาดใหญ่ซึ่งมีข้อมูลมากเกินไปความต้องการส่งผลให้ประสิทธิภาพของ Transaction Operational Database ทำงานได้ช้าลง ข้อมูลจะเป็นรูปแบบข้อมูลตารางเท่านั้น ข้อมูลจะถูกนำเสนอในรูปแบบที่ตายตัว ไม่สามารถเปลี่ยนแปลงตามความต้องการของผู้ใช้ ไม่ตอบสนองความต้องการของการตัดสินใจ เพราะข้อมูลสำหรับการตัดสินใจมีความสลับซับซ้อนสูงมีการรวมตัวกันของข้อมูลจากตารางต่างๆหลายๆตารางข้อมูล ไม่ตอบสนองการสอบถามข้อมูล (Data Queries) สำหรับผู้ใช้ มีข้อมูลย้อนหลังน้อย (Historical Data) ข้อมูลถูกจัดเก็บกระจัดกระจายตามที่ต่างๆ ซึ่งยากต่อการเรียกใช้ หรือขาดความสัมพันธ์ทางธุรกิจอันอาจจะต้องเสียเวลาในการทำให้สอดคล้อง หรือเกิดความซ้ำซ้อนของข้อมูลได้

จากอุปสรรคที่กล่าวมาข้างต้นคลังข้อมูลจึงได้ถูกออกแบบมา เพื่อตอบสนองงานในรูปแบบการตัดสินใจโดยการแยกฐานข้อมูลออกจาก Operational Database และเก็บข้อมูล

ในรูปแบบข้อมูลสรุป (Summary Data) ซึ่งข้อมูลสรุปนี้จะเลือกแต่เฉพาะข้อมูลที่มีความสำคัญต่อการตัดสินใจ หรือเพื่อใช้ในการบริหาร ไปจนถึงการกำหนดแผนงานในอนาคตในระบบคลังข้อมูล ข้อมูลที่ซับซ้อนจะถูกรวบรวมหรือเปลี่ยนแปลงให้ง่ายต่อการจัดเก็บและสามารถเรียกกลับมาใช้ได้อย่างรวดเร็วและถูกต้องโดยข้อมูลต่างๆ เหล่านี้จะถูกนำมาใช้สำหรับการวิเคราะห์และช่วยในเรื่องการตัดสินใจโดยอาศัยเครื่องมือ(tool)ที่อยู่ในเครื่องคอมพิวเตอร์ที่เป็นซอฟต์แวร์มาใช้ในการจัดการทำรายงาน และเพิ่มประสิทธิภาพสำหรับการตัดสินใจให้รวดเร็วยิ่งขึ้น โดยผู้บริหาร นักวางแผน และนักวิเคราะห์ข้อมูลสามารถเรียกหาข้อมูลหรือสอบถาม (query) เพื่อให้ได้รับคำตอบในรูปแบบตารางรายงาน หรือรายงานกราฟ เพื่อมาทำการวิเคราะห์ข้อมูลด้วยตนเอง

2.5 การเปรียบเทียบยอดขายระหว่างช่วงเวลาในอดีตกับปัจจุบันไปจนถึงการทำพยากรณ์ยอดขายในอดีต (Forecasting) การขายยอดขายสูงสุดหรือต่ำสุด

การเปรียบเทียบยอดขาย ต้นทุน กำไร ในรูปแบบตารางรายงาน หรือรายงาน กราฟ ซึ่งเครื่องมือนี้ถือได้ว่าเป็นสิ่งสำคัญในอันที่จะนำองค์กรไปสู่ความสำเร็จในกระบวนการตัดสินใจในปัจจุบันเครื่องมือที่ตอบสนองงานเพื่อช่วยผู้บริหารสำหรับการตัดสินใจมีอยู่มากมายในตลาดทั้งนี้ก็เป็นทางเลือกของผู้ใช้ในการที่จะเลือกเครื่องมือที่มีประสิทธิภาพสูงสุดเพื่อตอบสนองงานของผู้บริหารในกระบวนการตัดสินใจต่อไป

จะเห็นได้ว่าการจัดทำคลังข้อมูลเป็นความท้าทายอย่างหนึ่งของหน่วยงาน ทั้งนี้เพราะหน่วยงานต่าง ๆ มักจะมีข้อมูลธุรกรรมที่ไม่มีความต้องการและความลึกสั้นอยู่มาดั่งได้อธิบายไปบ้างแล้วดังนั้นการจัดทำคลังข้อมูลจะต้องหาทางแก้ปัญหาให้ได้อีกประการหนึ่งก็คือข้อมูลบางส่วนหายไปหรือมีไม่ครบ ยกตัวอย่างบริษัทแห่งหนึ่งต้องการวิเคราะห์ความสนใจของลูกค้าที่ใช้บัตรสมาชิกที่บริษัทออกให้โดยกำหนดจะแยกความสนใจว่ามีความแตกต่างระหว่างเพศหรือไม่ แต่ในการจัดทำระบบประมวลผลธุรกรรมตั้งแต่แรกนั้นนักวิเคราะห์ระบบไม่ได้กำหนดให้เก็บข้อมูลเพศของลูกค้าเอาไว้เพราะเห็นว่าไม่เกี่ยวกับธุรกรรมดังนั้นผู้ใช้จึงไม่สามารถนำข้อมูลมาวิเคราะห์ได้ ในกรณีเช่นนี้ระหว่างการจัดทำคลังข้อมูลก็ต้องจัดให้มีพนักงานที่ทำหน้าที่ศึกษาข้อมูลโดยพิจารณาจากระบบฟอร์มเดิมแล้วนำเพศมาบันทึกเป็นข้อมูลเพิ่มเติมขึ้น

การจัดทำคลังข้อมูลจะมีความสำคัญมากขึ้นในอนาคตเพราะปัจจุบันนี้ผู้ใช้และผู้บริหารของหน่วยงานเริ่มมีเข้าใจความสำคัญของข้อมูลมากขึ้นและเริ่มตระหนักว่าหากนำข้อมูลมาวิเคราะห์ให้เข้าใจสถานภาพหรือเหตุการณ์ที่เกิดขึ้นแล้วจะทำให้หน่วยงานหรือบริษัทสามารถตอบสนองต่อเหตุการณ์ที่เกิดขึ้นได้ดียิ่งขึ้นและจะทำให้หน่วยงานหรือบริษัททำงานบรรลุวัตถุประสงค์และเป้าหมายได้ดียิ่งขึ้นตามไปด้วย

2.6 ความสัมพันธ์ระบบคลังข้อมูลกับระบบฐานข้อมูล

ในปัจจุบันมีการใช้ฐานข้อมูลอย่างกว้างขวางในระบบงานทั่วไป จึงมีการวิจัยและพัฒนาวิธีเก็บข้อมูลจำนวนมากรวมถึงการค้นหาและนำข้อมูลที่ต้องการออกมาจากระบบฐานข้อมูลด้วย แต่เนื่องจากระบบงานข้อมูลทั่วไป ที่นิยมใช้อยู่ในปัจจุบันมีหลักในการเก็บข้อมูลที่เน้นในเรื่องการลดความซ้ำซ้อนรักษาความถูกต้องลดการสูญหายของข้อมูล และลดความผิดพลาดที่เกิดขึ้นจากการแก้ไขข้อมูล

เนื่องจากฐานข้อมูลทั่วไป มีลักษณะดังได้กล่าวมาแล้วจึงมีความสามารถเพียงแค่การเรียกใช้ข้อมูลที่มีอยู่ แต่ไม่สามารถจะนำมาช่วยในการสนับสนุนการตัดสินใจได้ เพราะเมื่อมีการเรียกใช้ข้อมูลจะต้องเรียกใช้ข้อมูลจากฐานข้อมูลขนาดใหญ่ซึ่งมีข้อมูลจำนวนมากและมีการแตกตารางที่นอร์มัลไลซ์ แล้วออกเป็นหลายตาราง จึงไม่รองรับคำถามที่ต้องการจะนำมาช่วยในการสนับสนุนการตัดสินใจ มีการรวมกันของตาราง ต่างๆ ที่ซับซ้อน ซึ่งจะทำให้มีประสิทธิภาพของการค้นหาข้อมูลจากฐานข้อมูลน้อยลง และทำงานช้าลง ไม่สามารถเรียกใช้ข้อมูลที่ต้องการได้ทั้งหมดเพราะมีรูทีนอัตโนมัติ จึงมีความสามารถในการค้นหาข้อมูลแบบที่ไม่ซ้ำซ้อนเท่านั้น นอกจากนี้การเก็บข้อมูลในระบบฐานข้อมูลทั่วไปยังไม่มีการเก็บข้อมูลย้อนหลัง เพื่อใช้ช่วยในการคาดคะเนแนวโน้มที่คาดว่าจะเป็นไปได้ในอนาคต ดังนั้นระบบคลังข้อมูลจึงได้ถูกคิดขึ้นมาเพื่อช่วยให้ผู้ใช้เรียกใช้ข้อมูลที่มีอยู่ได้อย่างมีประสิทธิภาพสูงสุดด้วยวิธีที่สร้างสรรค์เพราะธรรมชาติที่แตกต่างกันระหว่างระบบฐานข้อมูลคลังข้อมูลและระบบฐานข้อมูลทั่วไป ดังนั้นฐานข้อมูลคลังข้อมูลจะต้องมีคุณสมบัติดังนี้ Subject oriented ข้อมูลจะต้องถูกสร้างขึ้นจากหัวข้อธุรกิจที่สนใจ เช่นถ้า บริษัทประกันภัยต้องการใช้คลังข้อมูล ฐานข้อมูลที่ได้จะต้องสร้างขึ้นจากประวัติลูกค้า เบี้ยประกัน และการเรียกร้องแทนที่จะแยกตามชนิดของผลิตภัณฑ์ หรือบริการประกันภัย ประกันชีวิต ข้อมูลที่

สร้างขึ้นจะประกอบด้วยหัวข้อที่เก็บเฉพาะข่าวสารที่จำเป็น สำหรับ กระบวนการตัดสินใจเท่านั้น

Integrated ข้อมูลถูกรวบรวมจากแหล่งต่างๆจากระบบปฏิบัติการ รูปแบบของข้อมูลแพลตฟอร์มที่หลากหลายสร้างขึ้นเป็นฐานข้อมูลที่สอดคล้องเป็นหนึ่งเดียว เช่นค่าของตัวแปรตัวเดียวในแต่ละฐานข้อมูลอาจต่างกัน ฐานข้อมูลหนึ่งอาจใช้ 0 และ 1 อีกฐานข้อมูลหนึ่งอาจใช้ T และ F ดังนั้นฐานข้อมูลที่สร้างใหม่จะต้องได้รับการกำหนดค่าตัวแปรให้เหมือนกันเป็นหนึ่งเดียว Time-variant ข้อมูลซึ่งใช้ตัดสินใจที่เก็บไว้จะต้องมีอายุประมาณ 5 ถึง 10 ปี เพื่อใช้เปรียบเทียบ หาแนวโน้ม และทำนายผลลัพธ์ในอนาคตได้ Non-volatile ข้อมูลจะไม่อัปเดตหรือถูกทำให้เปลี่ยนแปลงง่ายๆ ผู้ใช้สามารถใช้ฐานข้อมูลคลังข้อมูลได้เพียงแค่อโหลดและเข้าถึงเท่านั้นโดยระบบฐานข้อมูลคลังข้อมูลจะแยกกลุ่มข้อมูลสารสนเทศที่ใช้ในการวิเคราะห์ทางธุรกิจออกจากฐานข้อมูลที่ใช้ประจำวัน มาเก็บอยู่ในระบบจัดการฐานข้อมูล (Relational Database Management Systems) ประสิทธิภาพสูงสุด และทำให้การเรียกใช้ข้อมูลชุดนี้ทำได้ อย่างยืดหยุ่น จากเครื่องมือที่อยู่บนเครื่องคอมพิวเตอร์เดสก์ท็อปทั่วไป โดยลด off-loading เพิ่มกลไกการช่วยตัดสินใจ ปรับปรุงเวลาที่ตอบสนองรวดเร็วขึ้นอย่างมาก และผู้บริหารสามารถเรียกข้อมูลรายละเอียดที่จำเป็นที่ถูกเก็บมาก่อนหน้านี้มาใช้ช่วยในการตัดสินใจทางธุรกิจแม่นยำขึ้น

ความแตกต่างอีกประการหนึ่งก็คือผู้ใช้คลังข้อมูลมักจะต้องการจัดกลุ่มข้อมูลด้วยตนเองมากกว่าผู้ใช้ในระบบฐานข้อมูลธรรมดา ยกตัวอย่างผู้ใช้ต้องการวิเคราะห์ผลกระทบของการทำการตลาดแบบต่างๆ อาจต้องการจัดกลุ่มการขายสินค้าแยกตามผลิตภัณฑ์ หรือรูปแบบของการจัดผลิตภัณฑ์ เช่น การห่อรวมสินค้าไว้ในบรรจุภัณฑ์สีต่างๆ หรือการรวมผลิตภัณฑ์ต่างรูปแบบไว้ด้วยกัน ในกรณีต่างๆ เหล่านี้ผู้ใช้ต้องการที่จะเลือกจัดกลุ่มข้อมูลได้ตามใจชอบ นอกจากการนำข้อมูลเข้ามารวมกันแล้ว ผู้ใช้ยังอาจต้องการที่จะแยกแยะข้อมูลในแบบที่ตนเองต้องการได้ ยกตัวอย่างในการจัดทำคลังข้อมูลเกี่ยวกับนักวิจัยและผลงานวิจัยของประเทศ หน่วยงานอาจจัดเก็บข้อมูลเอาไว้เป็นกลุ่มก้อนโดยไม่ได้แยกสาขา แต่ต่อมาผู้ใช้ต้องการนำข้อมูลนักวิจัยมาวิเคราะห์แยกแยะว่าทั้งประเทศมีนักวิจัยสาขาต่างๆ เป็นจำนวนเท่าใด ทำงานวิจัยด้านใดบ้าง ใช้เงินด้านวิจัยไปเท่าใด เป็นต้น โดยปกติแล้วการจัดทำฐานข้อมูลให้สามารถวิเคราะห์แยกแยะข้อมูลในแบบนี้ได้นั้นเป็นเรื่องไม่ยาก แต่ในการออกแบบคลังข้อมูลนั้นจำเป็นต้องเพื่อให้ผู้ใช้หลายคนสามารถแยกแยะข้อมูลตามความ

ต้องการที่แตกต่างกันได้ด้วย ผู้ใช้จำนวนมากในปัจจุบันนี้อาจใช้ซอฟต์แวร์หลากหลายประเภทสำหรับเครื่องคอมพิวเตอร์ส่วนบุคคล ผู้ใช้บางคนอาจจะใช้โปรแกรมสเปรดชีตในการวิเคราะห์ข้อมูล และผู้ใช้งานบางคนอาจต้องการใช้โปรแกรมวิเคราะห์สถิติอื่นๆ ดังนั้นผู้ใช้งานนี้อาจจะมีความต้องการในการนำเข้าข้อมูลจากคลังข้อมูลมาไว้ในแฟ้มข้อมูลที่มีรูปแบบตรงกับโปรแกรมที่ตนต้องการใช้ความต้องการด้านนี้นับว่าสำคัญมากที่สุดในการจัดทำคลังข้อมูล

งานอย่างหนึ่งที่นิยมใช้ฐานข้อมูลกันมากก็คืองานบันทึกข้อมูลธุรกรรมเอาไว้เพื่อประมวลผลข้อมูลธุรกรรมเหล่านี้ได้แก่ ข้อมูลการสั่งซื้อสินค้าของลูกค้า ข้อมูลการซื้อบัตรโดยสารเครื่องบิน ข้อมูลการฝากหรือถอนเงินของลูกค้าธนาคาร แต่เดิมนั้นการบันทึกข้อมูลธุรกรรม เริ่มต้นด้วยการใช้กระดาษแบบฟอร์มสำหรับให้ลูกค้ากรอกข้อมูล จากนั้นจึงนำแบบฟอร์มมาบันทึกข้อมูลลงในฐานข้อมูลของระบบคอมพิวเตอร์ในแบบแบตช์ (Batch) ปัจจุบันนี้การ บันทึกข้อมูลธุรกรรมได้เปลี่ยนไปเป็นระบบออนไลน์เป็นส่วนใหญ่ ในระบบแบบนี้กระบวนการบันทึกข้อมูลมีลักษณะอัตโนมัติมากขึ้นและใช้อุปกรณ์บันทึกข้อมูลที่สามารถเก็บข้อมูลลงในฐานข้อมูลของระบบคอมพิวเตอร์ได้ทันที เช่น การใช้อุปกรณ์ฝากถอนเงินโดยอัตโนมัติทำให้สามารถประมวลผลการฝากถอนเงินและบันทึกข้อมูลที่เกิดขึ้นได้ทันทีหรือในห้างสรรพสินค้าก็มีการใช้เครื่องบริการ ณ จุดขาย สำหรับอ่านรหัสแท่ง แสดงราคาสินค้า แล้วบันทึกข้อมูลการขายไปเก็บไว้ในฐานข้อมูลได้ทันทีการ

2.7 ดำเนินการในลักษณะนี้เรียกกันว่าการประมวลผลธุรกรรมออนไลน์ (On-Line Transaction Processing; OLTP)

1. ลักษณะงานการประมวลผลธุรกรรมออนไลน์และการประมวลผลเชิงวิเคราะห์ออนไลน์ระบบ OLTP โดยทั่วไปจะต้องสามารถดำเนินการกับข้อมูลธุรกรรมได้อย่างมีประสิทธิภาพ งานที่ทำกับข้อมูลได้แก่การปรับค่าของข้อมูลให้เป็นปัจจุบันและการเพิ่มข้อมูลลงไปในฐานข้อมูล ข้อมูลเหล่านี้ อาจจะมีจำนวนมากและเพิ่มขึ้นตลอดเวลา ณ เวลาใดเวลาหนึ่งอาจจะมีการประมวลผลข้อมูลจำนวนนับแสนเรคอร์ดได้ เช่น ณ สนามบินแต่ละแห่งจะมีผู้โดยสารเข้ามารับบัตรที่นั่งของสายการบินต่างๆ เป็นจำนวนนับหมื่นๆ คน คอมพิวเตอร์ของสายการบินจะต้องตรวจสอบการสำรองที่นั่ง ต้องบันทึกเลขที่นั่งและเที่ยวบินรวมทั้งอาจจะต้องปรับเปลี่ยนโยกย้ายข้อมูลจากเที่ยวบินหนึ่งไปอีกเที่ยวบินหนึ่งได้ด้วย หรือในกรณีของศูนย์การค้า และ ซูเปอร์มาร์เก็ต จะมีการบันทึกเรคอร์ดการขายเพิ่มเข้าไปในฐานข้อมูลการ

ขายตลอดเวลา รวมแล้ววันละเป็นหมื่นๆ รายการ การออกแบบระบบ OLTP แบบนี้ จำเป็นต้องหาทางให้ระบบสามารถทำงานได้อย่างถูกต้องรวดเร็วตลอดเวลา เอื้ออำนวยให้ผู้ใช้จำนวนมากสามารถใช้ระบบได้พร้อมกัน อีกทั้งยังต้องสามารถแก้ไขฟื้นฟูสภาพให้กลับคืนดังเดิมได้หากเกิดความขัดข้องเสียหาย

การที่จะจัดทำระบบ OLTP ให้มีความสามารถในการแบบนี้ได้ต้องคำนึงถึงปัจจัยต่อไปนี้

- 1) ขนาดและตำแหน่งของ Rollback segment
- 2) ดัชนี การจัดกลุ่ม และ การคำนวณตำแหน่งที่อยู่ (Hashing)
- 3) การออกแบบข้อมูลธุรกรรมให้เหมาะกับงานประยุกต์
- 4) หน่วยเก็บและเนื้อที่ว่างสำหรับการเก็บข้อมูลใหม่
- 5) ความเข้าใจลักษณะงานประยุกต์และการเขียนคำสั่งสำหรับค้นคืนข้อมูล
- 6) การปรับปรุงสมรรถนะของระบบอย่างต่อเนื่อง

ระบบ OLTP ที่พัฒนาขึ้นโดยใช้เทคนิคด้านฐานข้อมูลตามปกติมักจะไม่สามารถรับกับปริมาณข้อมูลที่เพิ่มขึ้นอย่างมากมาเป็นประจำทุกวันได้ การนำระบบเช่นนี้มาใช้จึงมีความเสี่ยงที่จะเกิดความผิดพลาดเสียหายขึ้น วิธีการแก้ไขก็คือการแยกฐานข้อมูลออกมาเป็นส่วน ๆ ให้เหมาะกับการใช้งาน

ข้อมูลส่วนมากที่จัดเก็บในคลังข้อมูลนั้นปกติจะมีน้อยกว่าข้อมูลในฐานข้อมูลของระบบ OLTP เพราะเป็นข้อมูลที่ได้นำมาจัดกลุ่มให้เหมาะสมแก่การค้นคืนแล้ว ข้อมูลเหล่านี้จะมีลักษณะ consistent กล่าวคือ ข้อมูลทุกรายการที่แสดงเรื่องเดียวกันจะต้องเขียนให้เหมือนกัน สะกดแบบเดียวกัน หรือ มีรหัสเดียวกัน หากข้อมูลมีลักษณะแตกต่างกันแล้วจะวิเคราะห์ข้อมูลได้ยาก หรืออาจทำให้ได้ผลลัพธ์ที่ไม่ถูกต้อง ในหน่วยงานและบริษัทขนาดใหญ่โอกาสที่ข้อมูลทั้งหมดจะ “สะอาด” นั้นเป็นเรื่องที่ยาก ดังนั้นจึงจำเป็นต้องมีผู้ทำหน้าที่กลั่นกรองและควบคุมคุณภาพของข้อมูลด้วย

2.8 ความแตกต่างของคลังข้อมูลกับฐานข้อมูลที่ใช้ประจำวัน

Consistency ทั้ง OLTP และ คลังข้อมูล ต่างก็ให้ความสำคัญในเรื่องข้อมูลควรมีความสอดคล้องกัน สำหรับ OLTP ซึ่งมีการทำ transaction จำนวนมาก ๆ สิ่งที่ต้องการคือการทำ transaction ให้ครบ ไม่มีการสูญหาย ดังนั้นจึงมีความจำเป็นผู้ส่งและผู้รับจะต้องรับรู้และตรวจสอบอยู่ตลอดเวลาว่าขณะนี้มีการทำ transaction เกิดขึ้นหรือไม่ สำหรับคลังข้อมูล จะ

ไม่สนใจทำการทำ transaction แต่ทุกครั้งแต่ละคนจะสนใจว่าการ load data ใหม่เข้ามานั้นทำสำเร็จหรือยัง และการ load data เข้ามาทั้งหมดนั้นถูกต้องหรือไม่

Transaction สำหรับระบบ OLTP นั้น ในแต่ละวันอาจมีการทำ transaction มากมาย ซึ่งการทำ transaction แต่ละครั้งจะใช้ข้อมูลเพียงแค่เล็กน้อยเท่านั้น สำหรับคลังข้อมูลแต่ละวันจะทำแค่เพียง 1 transaction ซึ่ง transaction นี้ต้องใช้ข้อมูลเป็นจำนวนมากมาย

Time Dimension สำหรับ OLTP นั้นจะทำงานอย่างรวดเร็วและทำ transaction อย่างสม่ำเสมอ สถานะของข้อมูลต่างๆมีการเปลี่ยนแปลงอยู่ตลอดเวลา และความสัมพันธ์ระหว่างเอนติตี้ต่างๆก็เปลี่ยนแปลงไปด้วย สำหรับระบบคลังข้อมูลมักจะเก็บข้อมูลในอดีตเพื่อใช้ในการวิเคราะห์ ดังนั้นข้อมูลจะไม่ค่อยมีการเปลี่ยนแปลงตลอดวัน

เป็นที่น่าสังเกตว่าคลังข้อมูลไม่ต้องทำการ normalization เหมือนกับฐานข้อมูลประจำวันที่ต้องทำการ normalization ทั้งนี้เพราะในฐานข้อมูลประจำวัน ข้อมูลจำนวนมากมีการเปลี่ยนแปลงทำให้ทันสมัยอยู่ตลอดเวลา ดังนั้นประเด็นสำคัญจึงอยู่ที่การเปลี่ยนแปลงทำให้ทันสมัยการออกแบบฐานข้อมูลประจำวันจึงต้องทำให้มีความซ้ำซ้อนหรือ redundancy น้อยที่สุดวิธีการที่จะทำให้ได้ตามจุดประสงค์คือการทำ normalization แต่สำหรับข้อมูลในคลังข้อมูลเป็นข้อมูลที่มีการกลั่นกรองมาแล้วใช้ในการวิเคราะห์ตอบคำถามของผู้บริหาร ประเด็นสำคัญจึงไม่อยู่ที่การทำให้ทันสมัย ทำให้ข้อมูลในคลังข้อมูลสามารถ มีความซ้ำซ้อนได้เพราะความซ้ำซ้อนมีข้อดีคือ การตอบคำถามและการออกรายงานสามารถทำได้รวดเร็ว เนื่องจากไม่ต้อง join หลายตารางดังนั้นในคลังข้อมูลจึงไม่มีความจำเป็นต้องทำการ normalization

2.9 โครงสร้างสถาปัตยกรรมและองค์ประกอบของคลังข้อมูล

จากความหมายของคลังข้อมูลที่เป็นที่เก็บข้อมูลขององค์กรที่ได้รับการออกแบบเพื่อช่วยในการตัดสินใจของฝ่ายบริหารนั้น ในทางปฏิบัติสิ่งที่เก็บอยู่ในโรงเก็บข้อมูลไม่ได้มีแต่เพียงข้อมูลเท่านั้น หากยังเก็บเครื่องมือสำหรับดำเนินการข้อมูลกระบวนการทำงานกับข้อมูลและทรัพยากรอื่นๆด้วยเช่น ภาพลักษณ์ของเอกสาร ภาพถ่าย แผนที่ เป็นต้น

ข้อมูลในคลังข้อมูลก็คือข้อมูลธุรกรรมของหน่วยงานดังนั้นจึงจำเป็นจะต้องมีเครื่องมือสำหรับแยกข้อมูลออกจากฐานข้อมูลองค์การมาเก็บไว้ ข้อมูลเหล่านี้จำเป็นจะต้องมีเมตาเดตา(Metadata) สำหรับใช้พรรณนา ลักษณะของข้อมูล ต้นกำเนิด รูปแบบ จิตจำกัคในการใช้ และลักษณะอื่น ๆ ของข้อมูลที่ใช้ในการกำหนดว่าจะใช้ข้อมูลนั้นอย่างไร

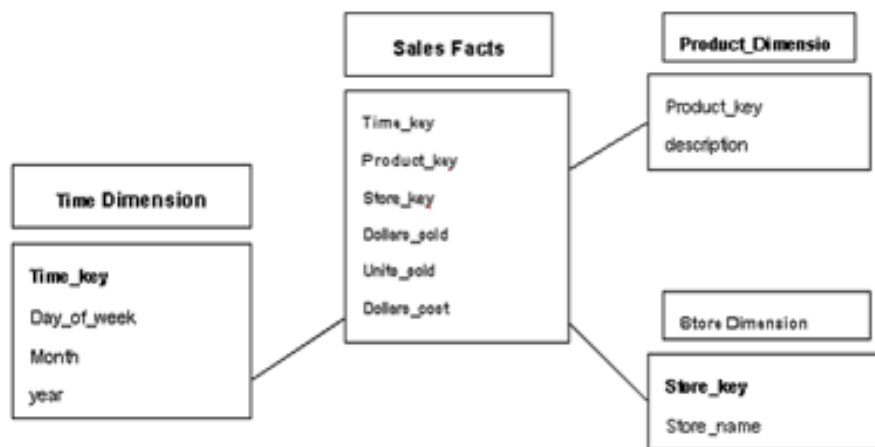
คลังข้อมูลอาจจะมีข้อมูลเป็นจำนวนมากมหาศาล ดังนั้นจึงจำเป็นต้องมีฐานข้อมูลของตนเองในการเก็บและประมวลผลข้อมูล หน่วยงานต้องมีโปรแกรมจัดการฐานข้อมูลและโปรแกรมอื่นๆ สำหรับช่วยในการเปลี่ยนรูปแบบข้อมูล จัดกลุ่มข้อมูล รวมข้อมูล และโยกย้ายข้อมูลจากฐานข้อมูลหนึ่งไปยังฐานอื่นๆ โปรแกรมเหล่านี้ต้องทำงานได้ทั้งกับข้อมูลที่เป็นจำนวน ข้อมูลกราฟิก ข้อมูลภาพลักษณ์ และ ข้อมูลแบบมัลติมีเดีย โปรแกรมเหล่านี้จะต้องสามารถแปลงข้อมูลให้เหมาะสมที่จะนำไปวิเคราะห์ และจัดทำรายงานในรูปแบบต่างๆ ได้ โดยที่คลังข้อมูลมีบริการสำคัญหลายอย่างให้แก่ผู้ใช้ซึ่งอาจจะไม่ใช่ผู้เชี่ยวชาญในด้านคอมพิวเตอร์ ดังนั้นการจัดคลังข้อมูลจึงจำเป็นต้องจัดการฝึกอบรมให้แก่ผู้ใช้ด้วย นอกจากนี้ยังอาจจะต้องจัดระบบอธิบายการใช้เอาไว้ในระบบด้วยเพื่อให้ผู้ใช้สามารถเรียกคำอธิบายมาใช้เมื่อต้องการได้ ส่วนประกอบสำคัญอีกอย่างหนึ่งก็คือที่ปรึกษาเกี่ยวกับคลังข้อมูลเพื่อช่วยเหลือผู้ใช้ให้สามารถใช้ระบบได้อย่างมีประสิทธิภาพมากขึ้น

ตาราง 2.1 ส่วนประกอบของคลังข้อมูล

ส่วนประกอบของคลังข้อมูล
· เครื่องมือสกัดแยกข้อมูล
· ข้อมูลที่สกัดและแยกออกมาแล้ว
· เมตาเดตาสำหรับบรรยายเนื้อหาข้อมูล
· ฐานข้อมูลสำหรับคลังข้อมูล
· เครื่องมือจัดการข้อมูลในคลังข้อมูล
· โปรแกรมสำหรับจัดส่งข้อมูล
· เครื่องมือวิเคราะห์สำหรับผู้ใช้
· วัสดุและหลักสูตรการฝึกอบรม
· ที่ปรึกษาด้านคลังข้อมูล

2.10 รายละเอียดของฐานข้อมูลสำหรับคลังข้อมูล

- 1) Dimensional Modeling เป็นชื่อเรียกของเทคนิคในการทำให้ฐานข้อมูลง่ายต่อการทำความเข้าใจโดยการมองภาพของฐานข้อมูลเป็นลูกบาศก์ที่มี 3,4,5 มิติ หรือมากกว่านั้น ทำให้สามารถจินตนาการการหั่น หรือแบ่งลูกบาศก์ที่มีลักษณะเหมือนลูกเต๋านี้ได้ นั่นคือสามารถตัดข้อมูลมาวิเคราะห์ดูในช่วงใดก็ได้ และหมุนข้อมูลดูได้จากทุกด้านของลูกเต๋าดังอย่างเช่นเราขายสินค้า (product) ในหลายๆที่ (market) และในช่วงเวลาต่างๆกัน (time) เราสามารถสร้าง Dimensional Modeling ได้โดยให้ label คือ product market และ time อยู่บนแต่ละด้านของลูกบาศก์ที่เป็น 3 มิติ แต่ละจุดภายในลูกบาศก์เกิดจากการตัดของ coordinate ซึ่งมี label อยู่ที่ขอบของลูกบาศก์ ดังนั้นจุดต่างๆภายในลูกบาศก์คือผลลัพธ์ทางด้านธุรกิจที่พิจารณาจากทั้ง 3 เรื่องคือ สินค้า ที่ขายสินค้า เวลา พร้อมๆกัน



รูป 2.1 แสดงภาพ Star Join Schema

- 2) Star Join Schema เป็นชื่อหนึ่งของ dimensional model ซึ่งเป็นชื่อที่ใช้กันมานาน เนื่องจาก diagram มีรูปร่างคล้ายดาว ซึ่งมีตารางใหญ่ 1 ตารางอยู่ตรงกลางซึ่งเรียกว่า fact table และมีตารางเล็กๆที่มีความสัมพันธ์กับตารางหลักนั้นอยู่รอบๆ เรียกว่า dimensional table ซึ่งตารางหลักนี้เป็นตารางเดียวที่ใช้ multiple join เพื่อเชื่อมต่อ

กับตารางอื่นๆ แต่ตารางอื่นๆที่อยู่รอบๆจะมีเพียงแค่ single join เพื่อเชื่อมเข้ากับตารางหลักเท่านั้น

- (1) Fact table ข้อมูลที่เก็บอยู่ใน fact table เรียกว่า grain of fact table แต่ละเรคคอร์ดใน Fact table จะแสดงถึง ยอดขายทั้งหมดของสินค้าหนึ่งที่ยขายได้ในสถานที่ที่แห่งหนึ่งในหนึ่งวัน fact table จะเป็นที่เก็บข้อมูลที่วัดได้ของธุรกิจหนึ่งๆ เช่นข้อมูลที่เป็นตัวเลข ข้อมูลที่วัดได้แต่ละตัวนี้มาจากการ transaction กันของทุกๆ dimensions จากรูป 2.3 ตัวเลขที่วัดได้ที่อยู่ในตาราง sales fact คือ จำนวนเงิน (number of dollars) จำนวนหน่วยสินค้าที่ขายได้ (number of units sold) และ ราคาสินค้า (cost)
- (2) Dimensional table เก็บคำอธิบายของแต่ละ dimension ของธุรกิจเอาไว้ ซึ่งคำอธิบายเหล่านี้จะช่วยให้การอธิบายถึงสมาชิกในทุกๆ dimension และใน dimension table จะประกอบด้วยหลายๆ attributes ซึ่ง attribute ที่จะต้องเป็นตัวอักษรและแต่ละ attribute ต้องแยกออกจากกัน

2.11 กระบวนการการรวบรวมข้อมูลและวิเคราะห์หารูปแบบข้อมูลในคลังข้อมูล

เนื่องจากข้อมูลพื้นฐานของฐานข้อมูลในคลังข้อมูลประกอบด้วยเรคคอร์ดจำนวนมาก จะต้องมีกระบวนการแบบคลังข้อมูลเพื่อรวบรวมข้อมูลและวิเคราะห์หารูปแบบข้อมูลในคลังข้อมูล ขั้นตอนการออกแบบประกอบด้วย 4 ขั้นตอน

1) เลือก Business process ที่ต้องการสร้างซึ่งเป็นขบวนการหลักที่ต้องการทำในองค์กร ซึ่งขบวนการนั้นมีระบบเดิมสนับสนุนอยู่ข้อมูลในระบบนั้นสามารถนำมารวบรวมเพื่อทำเป็นคลังข้อมูลได้เช่น ใบส่งของ (order) ใบสั่งของ (invoices) รายการสินค้า (inventory) ยอดขาย (sales) business process ที่เลือกมานั้นต้องเป็นหัวข้อธุรกิจที่สนใจ เพื่อจะได้สามารถทำการออกแบบคลังข้อมูลให้เกี่ยวข้องกับเฉพาะหัวข้อธุรกิจที่สนใจเท่านั้น ส่วนข้อมูลที่ไม่ได้เกี่ยวข้องกับหัวข้อธุรกิจก็จะไม่นำมารวมในคลังข้อมูล

2) เลือก Grain ของ business process ข้อมูลที่เป็นข้อมูลพื้นฐานที่เก็บอยู่ใน fact table เรียกว่า grain ใน business process นี้มี grain ที่มีอยู่ทั่วไป เช่น ข้อมูลของการทำ transaction ในแต่ละครั้ง (individual transaction) ข้อมูลของการทำงานในแต่ละวัน สรุปในแต่ละวัน ข้อมูลจากการสรุปการทำงานในแต่ละเดือน

3) เลือก Dimension ที่จะถูกนำมาใช้กับแต่ละ record ของ fact table นั้น dimension ที่มีอยู่ทั่วไปเช่น เวลา (time) สินค้า (product) ลูกค้า (customer) ซึ่งแต่ละ dimension จะถูกอธิบายแยกกันในลักษณะของ Dimensional attribute ซึ่งอธิบายแต่ละ dimension เป็นตัวหนังสือซึ่ง attributes เหล่านี้จะถูกเก็บอยู่ในแต่ละตาราง dimension

4) เลือก Measured fact (ข้อมูลที่มีการวัด การประมวลผล หรือการคำนวณไว้แล้ว) ที่จะเก็บอยู่ในแต่ละเรคคอร์ดของ fact table ปริมาณต่างๆหรือ measured fact ที่เพิ่มเข้าไปซึ่งจะมีลักษณะเป็นตัวเลข ได้แก่ ปริมาณที่ขายได้ (quality sold) และจำนวนเงินที่ได้รับจากการขาย (dollars sold)

ถ้านักวิเคราะห์ต้องการสอบถามข้อมูลโดยไม่มีการกำหนด เงื่อนไขให้กับบางตาราง Dimension แล้วในคลังข้อมูลมีเฉพาะข้อมูลพื้นฐานเท่านั้น นั่นหมายความว่า การสอบถามข้อมูล จะต้องทำการรวมข้อมูลภายในเรคคอร์ดจำนวนมากมาย ถ้าเป็นเช่นนี้จะทำให้ การทำ query แต่ละครั้งเกิดการสิ้นเปลืองมาก ดังนั้นจึงต้องทำการรวบรวมหรือข้อสรุป (aggregate) ข้อมูลไว้ล่วงหน้าเพื่อเร่งให้การทำ query สามารถทำได้เร็วขึ้นมีประสิทธิภาพในการทำงานมากขึ้น การรวบรวมหรือข้อสรุปสามารถถูกสร้างได้มากมายตามที่ต้องการในเฉพาะส่วนที่ต้องการ เท่านั้น

ในการสอบถามข้อมูลในคลังข้อมูลจะใช้ Structured Query Language (SQL) เป็นมาตรฐานสำหรับกาสอบถามข้อมูลทั้งหมดในคลังข้อมูลที่เกี่ยวข้องกับ fact table การสอบถามข้อมูลจะต้องมีการใช้ dimension table ในการทำการสอบถามข้อมูลเสมอเพราะใน dimension table ออจะเก็บชื่อและคำอธิบายที่การสอบถามข้อมูลต้องการใช้เอาไว้การสอบถามข้อมูลจะประกอบไปด้วย Where Clause 2 ส่วนคือ

- การ join ระหว่าง fact table และ dimension table
- เขตของข้อกำหนดหรือเงื่อนไข (criteria) สำหรับคอลัมน์ที่อยู่ภายใน dimension table