บทที่ 1
บทนำ

1.1 ความสำคัญและที่มา

แนวความคิดของดาต้าแวร์เฮ้าท์ซึ่งนั้นได้รับการพัฒนามาค่อนข้างนานตั้งแต่ช่วงต้นของปี 80 ในอดีตนั้นฐานข้อมูลหนึ่ง ๆ จะถูกนำมาใช้ทั้งในงาน โอเปอเรชั่นนอล และ การวิเคราะห์ข้อมูล ฐานข้อมูลส่วนใหญ่ก็มักจะถูกออกแบบมาเพื่อใช้กับระบบทรานแซคชั่น โพรเซสซิ่ง (transaction processing system) และการวิเคราะห์ข้อมูลก็จะถูกกระทำโดยมนุษย์ โดยทำการดึงข้อมูลมาจากระบบทรานแซคชั่น โพรเซสซิ่ง และเนื่องจากในอดีตนั้นราคาของซอฟแวร์และฮาร์ดแวร์ที่ใช้ในการบำรุงรักษาฐานข้อมูลนั้นมีราคาค่อนข้างสูง ฐานข้อมูลที่ใช้ในการนำเสนอข้อมูลที่ใช้ในการวิเคราะห์นั้นมีจำนวนเรคอร์ดที่ค่อนข้างจำกัด จึงทำให้การมีฐานข้อมูลเดียวที่รองรับการทำงานทั้งสองประเภทที่กล่าวมาจึงได้รับความนิยมค่อนข้างมาก

แต่ปัจจุบันเมื่อความต้องการทางด้านข้อมูลสูงขึ้น จึงทำให้เกิดข้อจำกัดทางเทคโนโลยีที่จะใช้งานฐานข้อมูลระหว่างทรานแซคชั่น และ การวิเคราะห์ข้อมูลพร้อม ๆ กัน ซึ่งโดยภาพรวมแล้ว ลักษณะของการนำโพรเซสซิ่งทั้งสองกระบวนการมาใช้นั้นมีความแตกต่างกันอย่างสิ้นเชิง

ทรานแซคชั่น โพรเซสซิ่ง ดาต้าเบส จะเกี่ยวข้องกับข้อมูลที่เป็นปัจจุบัน ซึ่งจะมีการเปลี่ยนแปลงของข้อมูลอยู่เรื่อย ๆ ดังนั้นจำนวนของข้อมูลที่มีปริมาณเพิ่มขึ้นจะเป็นตัวลดประสิทธิภาพของ โอเปอเรชั่นนอล ดาต้าเบส ซึ่งข้อมูลที่มีความซ้ำซ้อนจะต้องมีการจัดการที่ดี เพื่อป้องกันปัญหาการเปลี่ยนแปลงข้อมูล ซึ่งได้แก่กระบวนการเพิ่มข้อมูล (insert) อัพเดทข้อมูล(update) และ ลบข้อมูล(delete) ข้อมูลที่ซ้ำกัน

แต่สำหรับฐานข้อมูลที่ใช้ในการวิเคราะห์ข้อมูลเพื่อตอบคำถามนั้น มักจะมีการทำงานในลักษณะที่อ่านได้เพียงอย่างเดียว การเปลี่ยนแปลงข้อมูลนั้นเกิดขึ้นน้อยมาก ความซ้ำซ้อนของข้อมูลนั้นไม่ใช่ปัญหาของดาต้าแวร์เฮ้าท์ และมักจะเป็นฐานข้อมูลที่มีขนาดใหญ่เนื่องจากมีการเก็บข้อมูลเป็นเวลา การมีกระบวนการเก็บข้อมูลที่ดีจะมีผลผลทำให้ประสิทธิภาพของการตอบคำถามดีขึ้น

จากที่กล่าวถึงคุณลักษณะของทรานแซคชั่น โพสเซสซิ่งและดาต้าแวร์เฮ้าท์ เป็นเรื่องยากที่จะออกแบบฐานข้อมูลที่สามารถรองรับถึงคุณลักษณะของทั้งสองได้ ดาต้าแวร์เฮ้าท์ซึ่งนั้นจะรวบรวมข้อมูลทางธุรกิจหลาย ๆ แบบเข้าไว้ด้วยกัน เช่น ไฟล์ ฐานข้อมูลแบบเชิงสัมพันธ์ ,สปีรีดชีท ,แฟลท ดาต้าเบส และ เท็กไฟล์ทั่ว ๆไป และดาต้าแวร์เฮ้าท์ซึ่งจะดึงข้อมูลจากฐานข้อมูลที่มีชนิดต่าง ๆ กัน ไปเก็บในโครงสร้างฐานข้อมูลอันใหม่ ทำให้กลุ่มของผู้ใช้งานสามารถใช้งานทุก ๆ ข้อมูลจากฐานข้อมูลเพียงอันเดียว

เพราะฉะนั้นแวร์เฮ้าท์ดาต้าเบส จึงใช้วิธีการหลาย ๆ อย่างในการจัดการกับข้อมูล โดยดาต้าแวร์เฮ้าท์ดึงข้อมูลมาจากระบบ ทรานแซคชั่น โพรเซสซิ่ง แปลงข้อมูลให้อยู่ในรูปแบบที่เหมาะสมแล้วจึงทำการจับคู่กับสกีมาที่ต่างออกไปจากสกีมาเดิม ที่สามารถทำงานได้ดีกับปริมาณข้อมูลขนาดใหญ่และทำให้การคิวรี่มีประสิทธิภาพดีขึ้น

สำหรับการออกแบบดาต้าแวร์เฮ้าท์ ได้มีการใช้โมเดลหลาย ๆ แบบในการแก้ปัญหาของดาต้าแวร์เฮ้าท์ ซึ่งก็มีโมเดลที่โด่งดังมีการใช้งานอย่างแพร่หลายก็คือ โมเดลรูปดาว หรือที่เรียกกันว่า สตาร์ สกีมา

(star schema) ซึ่งใช้หลักการ ดีนอมาไลซ เทเบิ้ล (de-normalized table) และ สโนว์เฟลก สกีมา(Snowflex Schema)ที่แปลงมาจาก สตาร์ สกีมา ใช้การเลือก นอมาไลเซชั่น เฉพาะบาง ไดเมนชั่น เทเบิ้ล (dimension table)

ฐานข้อมูลแบบวัตถุเชิงสัมพันธ์นั้นเป็นการผสมผสานกันระหว่างแนวความคิดในการออกแบบฐานข้อมูลเชิงวัตถุที่มีการใช้งานแบบซับซ้อนและ มีการประกาศโครงสร้างของข้อมูลประเภทใหม่ (user-defined data structure) กับข้อเด่นของฐานข้อมูลเชิงสัมพันธ์ เช่น การนำเสนอข้อมูลอย่างง่าย ๆ ในรูปของตาราง (tabular form) และ สนับสนุนของภาษาเอสคิวแอล(Structure Query Language) สำหรับการประกาศชนิดของข้อมูล(data definition)และการเปลี่ยนแปลงข้อมูล (data manipulation) ทำให้ได้ลักษณะเฉพาะของฐานข้อมูลแบบวัตถุเชิงสัมพันธ์ คือ 1 . การประกาศชนิดของข้อมูลไว้ใช้เอง 2 . สนับสนุนคอมเพล็กออบเจ๊ก (complex object) 3 . การสืบทอดของชนิดของข้อมูล

ผลงานในปัจจุบันที่เกี่ยวกับ ดาต้าแวร์เฮ้าท์ซึ่ง ส่วนใหญ่จะกล่าวถึงเทคโนโลยีแบบรีเลชั่นนอลและมีบางอันที่ใช้เทคโนโลยีแบบออบเจ็ก ซึ่งการออกแบบดาต้าแวร์เฮ้าท์ดาต้าเบสบนเทคโนโลยีแบบทั้งสองนั้นจะแสดงในรูปแบบของโมเดลในรูปดาว แต่เนื่องจากโมเดลรูปดาวที่แต่เดิมจะแสดงให้เห็นในรูปแบบของลอจิคคอล ดาต้า โมเดล(Logical Data Model) ไม่สามารถที่จะแสดงให้เห็นถึงความสัมพันธ์ของแอททริบิวต์ได้ จึงได้มีการพัฒนาโมเดลรูปดาว ให้อยู่ในรูปแบบของ อีอาร์ไดอะแกรม ซึ่งเป็นโมเดลที่นิยมนำมาใช้ในการออกแบบโอเปอร์เรชั่นนอล ดาต้าเบสอยู่แล้วโดยเรียกโมเดลนี้ว่า สตาร์อีอาร์ โมเดล (Micheal K. and II-Yeol Song) แต่สำหรับในฐานข้อมูลแบบวัตถุสัมพันธ์ นั้นสตาร์อีอาร์ โมเดลนั้นจะไม่สามารถแสดง คอมโพสิท แอททริบิทต์ มัลติแวลู แอททริบิวต์ และ ไม่สามารถแสดงการสืบทอดคุณสมบัติของแต่ละ เอ็นทิตี้ไทป์ได้ ซึ่งเป็นส่วนสำคัญของฐานข้อมูลแบบวัตถุเชิงสัมพันธ์ ปริญญานิพนธ์นี้ได้พัฒนาโมเดลเพิ่มขึ้นในส่วนแสดงคอมโพสิท แอททริบิทต์ มัลติแวลู แอททริบิทต์ และ การสืบทอดของแต่ละ เอ็นตีตี้ไทป์ และยังได้แสดงให้เห็นถึงการใช้คำสั่งเอสคิวแอลในการเข้าถึงข้อมูลที่ถูกจัดอยู่ในรูปแบบของเทคโนโลยี ออปเจ็ครีเลชั่นนอล ซึ่งจะแสดงในบทที่ 4 และบทที่ 5

1.2 วัตถุประสงค์ของงานวิจัย

1.2.1 ศึกษาและพัฒนาวิธีการออกแบบฐานข้อมูลเชิงวัตถุสัมพันธ์ เพื่อประยุกต์ใช้งานกับดาต้าแวร์เฮ้าท์จริง โดยใช้โปรแกรม IBM DB2 เวอร์ชั่น 8.1

1.2.2 สร้างต้นแบบบฐานข้อมูลแบบวัตถุเชิงสัมพันธ์ เพื่อใช้งานบนดาต้าแวร์เฮ้าท์ โดยใช้วิธีการออกแบบที่ได้พัฒนาขึ้นมา

1.3 ขอบเขตของงานวิจัย

งานวิจัยนี้ได้ศึกษาถึงระบบฐานข้อมูลแบบเชิงวัตถุที่ใช้งานบนดาต้าแวร์เฮ้าท์ ในกรณีศึกษาที่เหมาะสมสำหรับใช้งานระบบฐานข้อมูลแบบวัตถุเชิงสัมพันธ์ โดยรูปแบบของระบบฐานข้อมูลดังกล่าวจะถูกปรับปรุงโดยการนำเอาแบบจำลองอีอาร์ (ER - Diagram) ของปีเตอร์-เชน ซึ่งมักจะถูกนำมาใช้ในการออกแบบฐานข้อมูลเชิงสัมพันธ์มาประยุกต์กับแนวความคิดในการออกแบบดาต้าแวร์เฮ้าท์

ซึ่งมักจะใช้วิธีการของโมเดลรูปดาว เพื่อที่จะแสดงให้เห็นถึงความสัมพันธ์ของแอททริบิวต์ภายในแต่ละไดเมนชั่น และนอกจากนี้ยังได้พัฒนาฐานข้อมูลนี้โดยใช้แนวความคิดแบบฐานข้อมูลเชิงวัตถุสัมพันธ์

1.4 วิธีการดำเนินงาน

งานวิจัยนี้เริ่มด้วยการศึกษาทฤษฎีพื้นฐานต่างๆที่เกี่ยวข้องกับงานวิจัย ซึ่งก็มีเรื่องหลัก ๆ อยู่ 2 เรื่องด้วยกัน คือ ดาต้าแวร์เฮ้าท์ ฐานข้อมูลแบบวัตถุเชิงสัมพันธ์ ซึ่งมีรายละเอียดดังในบทที่ 2 และ 3 จากนั้นก็เอาความรู้ที่ได้ศึกษาทั้งหมดมาพัฒนาออกแบบ ดาต้าแวร์เฮ้าท์ใหม่ที่ใช้ฐานข้อมูลแบบวัตถุเชิงสัมพันธ์ ซึ่งมีรายละเอียดในบทที่ 4 และ การใช้งานจริงบนดาต้าแวร์เฮ้าท์ โดยใช้ IBM DB2 เวอร์ชั่น8.1 ซึ่งมีรายละเอียดในบทที่ 5