

โครงสร้างสถาปัตยกรรมและองค์ประกอบของคลังข้อมูล

จากความหมายของคลังข้อมูลที่ว่า เป็นที่เก็บข้อมูลขององค์กรที่ได้รับการออกแบบเพื่อช่วยในการตัดสินใจของฝ่ายบริหารนั้น ในทางปฏิบัติสิ่งที่เก็บอยู่ในโรงเก็บข้อมูลไม่ได้มีแต่เพียงข้อมูลเท่านั้น หากยังเก็บเครื่องมือสำหรับดำเนินการข้อมูล กระบวนการทำงานกับข้อมูล และทรัพยากรอื่น ๆ ด้วยเช่น ภาพลักษณ์ของเอกสาร ภาพถ่ายแผนที่ เป็นต้น

ข้อมูลในคลังข้อมูลก็คือข้อมูลธุรกรรมของหน่วยงาน ดังนั้นจึงจำเป็นต้องมีเครื่องมือสำหรับแยกข้อมูลออกจากฐานข้อมูลองค์การมาเก็บไว้ ข้อมูลเหล่านี้จำเป็นต้องมีเมตาเดตา (metadata) สำหรับใช้พรรณนา ลักษณะของข้อมูล ต้นกำเนิด รูปแบบ ซิดจำกัดในการใช้ และ ลักษณะอื่น ๆ ของข้อมูลที่ใช้ในการกำหนดว่าจะใช้ข้อมูลนั้นอย่างไร

คลังข้อมูลอาจจะมีข้อมูลเป็นจำนวนมากมายมหาศาล ดังนั้นจึงจำเป็นต้องมีฐานข้อมูลของตนเองในการเก็บและประมวลผลข้อมูล หน่วยงานต้องมีโปรแกรมจัดการฐานข้อมูลและโปรแกรมอื่นๆ สำหรับช่วยในการเปลี่ยนรูปแบบข้อมูล จัดกลุ่มข้อมูล รวมข้อมูล และโยกย้ายข้อมูลจากฐานข้อมูลหนึ่งไปยังฐานอื่นๆ โปรแกรมเหล่านี้ต้องทำงานได้ทั้งกับข้อมูลที่เป็นจำนวน ข้อมูลกราฟิก ข้อมูลภาพลักษณ์ และ ข้อมูลแบบมัลติมีเดีย โปรแกรมเหล่านี้จะต้องสามารถแปลงข้อมูลให้เหมาะสมที่จะนำไปวิเคราะห์ และจัดทำรายงานในรูปแบบต่างๆ ได้

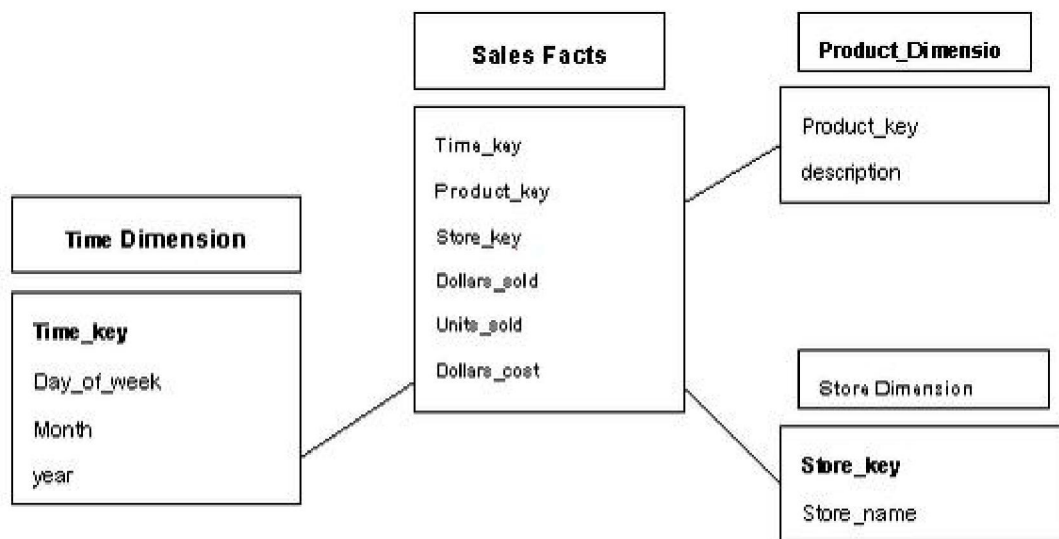
โดยที่คลังข้อมูลมีบริการสำคัญหลายอย่างให้แก่ผู้ใช้ซึ่งอาจจะไม่ใช่ผู้เชี่ยวชาญในด้านคอมพิวเตอร์ ดังนั้นการจัดการคลังข้อมูลจึงจำเป็นต้องจัดการฝึกอบรมให้แก่ผู้ใช้ด้วย นอกจากนี้ยังอาจจะต้องจัดระบบอธิบายการใช้เอาไว้ในระบบด้วยเพื่อให้ผู้ใช้สามารถเรียกคำอธิบายมาใช้เมื่อต้องการได้ ส่วนประกอบสำคัญอีกอย่างหนึ่งก็คือที่ปรึกษาเกี่ยวกับคลังข้อมูลเพื่อช่วยเหลือผู้ใช้ให้สามารถใช้ระบบได้อย่างมีประสิทธิภาพมากขึ้น

ตารางที่ 15.2.2 ส่วนประกอบของคลังข้อมูล

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

ในที่นี้จะขอกล่าวรายละเอียดของฐานข้อมูลสำหรับคลังข้อมูลซึ่งประกอบด้วย

1. **Dimensional Modeling** เป็นชื่อเรียกของเทคนิคในการทำให้ฐานข้อมูลง่ายต่อการทำความเข้าใจ โดยการมองภาพของฐานข้อมูลเป็นลูกบาศก์ที่มี 3,4,5 มิติ หรือมากกว่านั้น ทำให้สามารถจินตนาการการ หั่นหรือแบ่งลูกบาศก์ที่มีลักษณะเหมือนลูกเต๋านี้ได้ นั่นคือสามารถตัดข้อมูลมาวิเคราะห์ดูในช่วงใดก็ได้ และหมุนข้อมูลดูได้จากทุกๆด้านของลูกเต๋า ตัวอย่างเช่น เราขายสินค้า (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



รูปที่ 15.2.2 แสดงภาพ Star Join Schema

2. **Star Join Schema** เป็นชื่อหนึ่งของ dimensional model ซึ่งเป็นชื่อที่ใช้กันมานานเนื่องจาก diagram มีรูปร่างคล้ายดาว ซึ่งมีตารางใหญ่ 1 ตารางอยู่ตรงกลางซึ่งเรียกว่า **fact table** และมีตารางเล็กๆที่มีความสัมพันธ์กับตารางหลักนั้นอยู่รอบๆ เรียกว่า **dimensional table** ซึ่งตารางหลักนี้เป็นตารางเดียวที่ใช้ multiple join เพื่อเชื่อมต่อกับตารางอื่นๆ แต่ตารางอื่นๆที่อยู่รอบๆจะมีเพียงแค่ single join เพื่อเชื่อมเข้ากับตารางหลักเท่านั้น

2.1 **fact table** ข้อมูลที่เก็บอยู่ใน fact table เรียกว่า **grain of fact table** แต่ละเรคคอร์ดใน fact table จะแสดงถึง ยอดขายทั้งหมดของสินค้าหนึ่งที่ขายได้ในสถานที่ที่แห่งหนึ่งในหนึ่งวัน fact table จะเป็นที่เก็บข้อมูลที่วัดได้ของธุรกิจหนึ่งๆ เช่นข้อมูลที่เป็นตัวเลข ข้อมูลที่วัดได้แต่ละตัวนี้มาจากการ intersection กันของทุกๆ dimensions จากรูป 15.2.2 ตัวเลขที่วัดได้ที่อยู่ในตาราง sales fact คือ จำนวนเงิน (number of dollars), จำนวนหน่วยสินค้าที่ขายได้ (number of units sold), ราคาสินค้า (cost)

2.2dimensional table เก็บคำอธิบายของแต่ละ dimension ของธุรกิจเอาไว้ ซึ่งคำอธิบายเหล่านี้จะช่วยในการอธิบายถึงสมาชิกในทุกๆ dimension และใน dimension table จะประกอบด้วยหลายๆ attributes ซึ่ง attribute ที่ดีจะต้องเป็นตัวอักษรและแต่ละ attribute ต้องแยกออกจากกัน