

กระบวนการการรวบรวมข้อมูลและวิเคราะห์หารูปแบบข้อมูลในคลังข้อมูล

เนื่องจากข้อมูลพื้นฐานของฐานข้อมูลในคลังข้อมูลประกอบด้วยเรคคอร์ดจำนวนมากจะต้องมีการออกแบบคลังข้อมูลเพื่อรวบรวมข้อมูลและวิเคราะห์หารูปแบบข้อมูลในคลังข้อมูล ขั้นตอนการออกแบบประกอบด้วย 4 ขั้นตอน

ขั้นตอน 1 เลือก business process ที่ต้องการสร้างซึ่งเป็นขบวนการหลักที่ต้องการทำในองค์กร ซึ่งขบวนการนั้นมีระบบเดิมสนับสนุนอยู่ ข้อมูลในระบบนั้นสามารถนำมารวบรวมเพื่อทำเป็นคลังข้อมูลได้ เช่น ใบส่งของ (order), ใบสั่งของ (invoices), รายการสินค้า (inventory), ยอดขาย (sales) business process ที่เลือกมานั้นต้องเป็นหัวข้อธุรกิจที่สนใจ เพื่อจะได้สามารถทำการออกแบบคลังข้อมูลให้เกี่ยวข้องกับเฉพาะหัวข้อธุรกิจที่สนใจเท่านั้น ส่วนข้อมูลที่ไม่ได้เกี่ยวข้องกับหัวข้อธุรกิจก็จะไม่นำมารวมในคลังข้อมูล

ขั้นตอน 2 เลือก grain ของ business process ข้อมูลที่เป็นข้อมูลพื้นฐานที่เก็บอยู่ใน fact table เรียกว่า grain ใน business process นี้มี grain ที่มีอยู่ทั่วไป เช่น ข้อมูลของการทำ transaction ในแต่ละครั้ง (individual transaction), ข้อมูลของการทำงานในแต่ละวัน, สรุปรายในแต่ละวัน (individual daily snapshots), ข้อมูลจากการสรุปการทำงานในแต่ละเดือน (individual monthly snapshots)

ขั้นตอน 3 เลือก dimension ที่จะถูกนำมาใช้กับแต่ละ record ของ fact table นั้น dimension ที่มีอยู่ทั่วไปเช่น เวลา (time), สินค้า (product), ลูกค้า (customer) ซึ่งแต่ละ dimension จะถูกอธิบายแยกกัน ในลักษณะของ dimensional attribute ซึ่งอธิบายแต่ละ dimension เป็นตัวหนังสือ ซึ่ง attributes เหล่านี้จะถูกเก็บอยู่ในแต่ละตาราง dimension

ขั้นตอน 4 เลือก measured fact (ข้อมูลที่มีการวัด, การประมวลผล หรือการคำนวณไว้แล้ว) ที่จะเก็บอยู่ในแต่ละเรคคอร์ดของ fact table ปริมาณต่างๆหรือ measured fact ที่เพิ่มเข้าไปซึ่งจะมีลักษณะเป็นตัวเลข ได้แก่ ปริมาณที่ขายได้ (quantity sold) และ จำนวนเงินที่ได้รับจากการขาย (dollars sold)

ถ้านักวิเคราะห์ต้องการสอบถามข้อมูล (query) โดยไม่มีการกำหนดเงื่อนไขให้กับบางตาราง dimension แล้วในคลังข้อมูลมีแต่เฉพาะข้อมูลพื้นฐานเท่านั้น นั่นหมายความว่า การสอบถามข้อมูล (query) จะต้องทำการรวมข้อมูลภายในเรคคอร์ดจำนวนมากมาย ถ้าเป็นเช่นนี้จะทำให้การ

ทำ query แต่ละครั้งเกิดการสิ้นเปลืองมาก ดังนั้นจึงต้องทำการรวบรวมหรือข้อสรุป (aggregate) ข้อมูลไว้ล่วงหน้าเพื่อเร่งให้การทำ query สามารถทำได้เร็วขึ้นมีประสิทธิภาพในการทำงานมากขึ้น การรวบรวมหรือข้อสรุป (aggregate) สามารถถูกสร้างได้มากมายตามที่ต้องการในเฉพาะส่วนที่ต้องการเท่านั้น

ในการสอบถามข้อมูล (query) ในคลังข้อมูลจะใช้ Structured Query Language (SQL) เป็นมาตรฐานสำหรับกาสอบถามข้อมูลทั้งหมดในคลังข้อมูลที่เกี่ยวข้องกับ fact table การสอบถามข้อมูลจะต้องมีการใช้ dimension table ในการทำการสอบถามข้อมูลเสมอ เพราะใน dimension table จะเก็บชื่อและคำอธิบายที่การสอบถามข้อมูลต้องการใช้เอาไว้ การสอบถามข้อมูลจะประกอบไปด้วย *Where Clause* 2 ส่วนคือ

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

ที่มา

<http://www.srisangworn.go.th/home/databaselearnx/ms2t3-6.htm>