บทที่ 3
ขั้นตอนการดำเนินงาน

3.1 การกำหนดเป้าหมาย (Objective)

การทำเหมืองข้อมูลเพื่อสกัดเอาองค์ความรู้ที่มีอยู่ในฐานข้อมูล ให้สอดคล้องกับเป้าหมายหรือนโยบายของบริษัทตัวอย่าง จะเลือกใช้เทคนิคของการวิเคราะห์ข้อมูลทางด้านการตลาด (Market Basket Analysis) ซึ่งเป็นเทคนิคหนึ่งของการทำเหมืองข้อมูล มาช่วยในการลดต้นทุนและเพิ่มยอดขายให้กับบริษัท โดยที่เป้าหมายที่ต้องการคือการวางแผนในการจัดซื้อสินค้าหรือการควบคุมปริมาณของสินค้าคงคลังไม่ให้มีมากเกินไปและมีปริมาณที่เหมาะสม และเพียงพอต่อความต้องการของลูกค้า ในขณะเดียวกันก็วางแผนในการเพิ่มยอดขาย เพื่อเพิ่มโอกาสการทำกำไรให้กับบริษัท

เทคนิคที่ใช้ในการวิเคราะห์การตลาด คือ

- การจัดหมวดหมู่ (Classification) เพื่อหากฎการตัดสินใจ (Decision Rule) เพื่อใช้ในการวางแผนการจัดซื้อและควบคุมปริมาณสินค้าคงคลัง
- การหากฎความสัมพันธ์ (Association Rule) ของสินค้า เพื่อใช้ในการวางแผนการเพิ่มยอดขาย

เครื่องมือที่ใช้

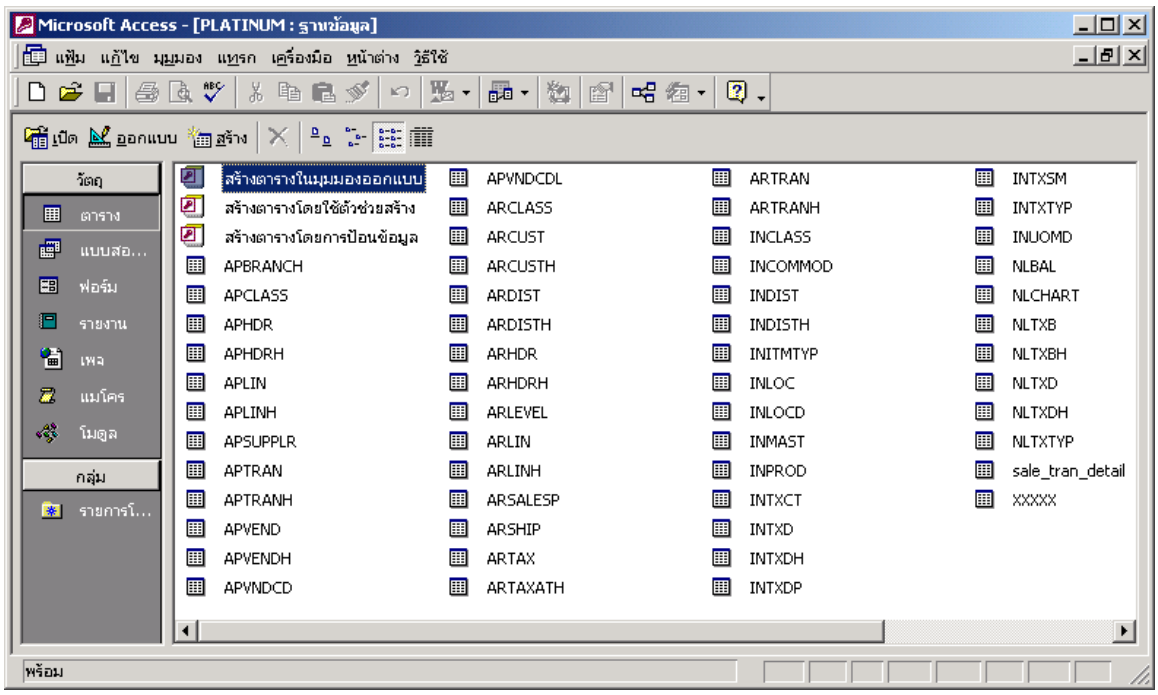
ฐานข้อมูล - Microsoft Access, Microsoft SQL Server

โปรแกรม – Microsoft Analysis Manager, DBMiner

3.2 สร้างฐานข้อมูลสำหรับการทำ Data Mining

3.2.1 การคัดเลือกข้อมูล (Data Selection)

แหล่งของข้อมูลที่ต้องการนำมาทำเหมืองข้อมูลนั้น เป็นข้อมูลที่ได้จากโปรแกรมสำเร็จรูปที่มีชื่อ ว่า PLATINUM ซึ่งใช้ฐานข้อมูลของ Microsoft Access โดยมีตารางที่เกี่ยวข้องกับระบบดังในรูปต่อไปนี้



รูป 3.1 แสดงตารางต่างๆที่ใช้ในโปรแกรมสำเร็จรูป PLATINUM

การคัดเลือกตารางข้อมูลที่ต้องการ จะต้องให้ได้ข้อมูลที่สามารถนำไปสร้างให้อยู่ในรูปแบบที่สอดคล้องกับโปรแกรมสำเร็จรูปที่ใช้เป็นเครื่องมือในการทำเหมืองข้อมูล และได้ตรงตามเป้าหมายที่กำหนดไว้ จากการที่ได้ศึกษาโครงสร้างของฐานข้อมูล สามารถสรุปได้ว่า ข้อมูลต่างๆที่ต้องการอยู่ในตารางดังต่อไปนี้

ตารางที่ 3.1

กลุ่มแฟ้มข้อมูลระบบ Inventory ระบบงานบริหารสินค้าคงคลัง		
1.	**INMAST**	แฟ้มเก็บชื่อสินค้าและรายละเอียดสินค้า
2.	**INPROD**	แฟ้มอ้างอิงกลุ่มของสินค้า เช่น INK คือ หมึก DIS คือ แผ่นบรรจุข้อมูล
3.	**INCOMMOD**	แฟ้มอ้างอิงยี่ห้อผู้ผลิต เช่น HP0 คือ ฮิวเล็ตแพ็คการ์ด
4.	**INUMOD**	แฟ้มอ้างอิงหน่วยนับสินค้า
5	**INITMTYP**	แฟ้มอ้างอิงประเภทของสินค้า
6	**INLOCD**	แฟ้มเก็บรายละเอียดของสาขา

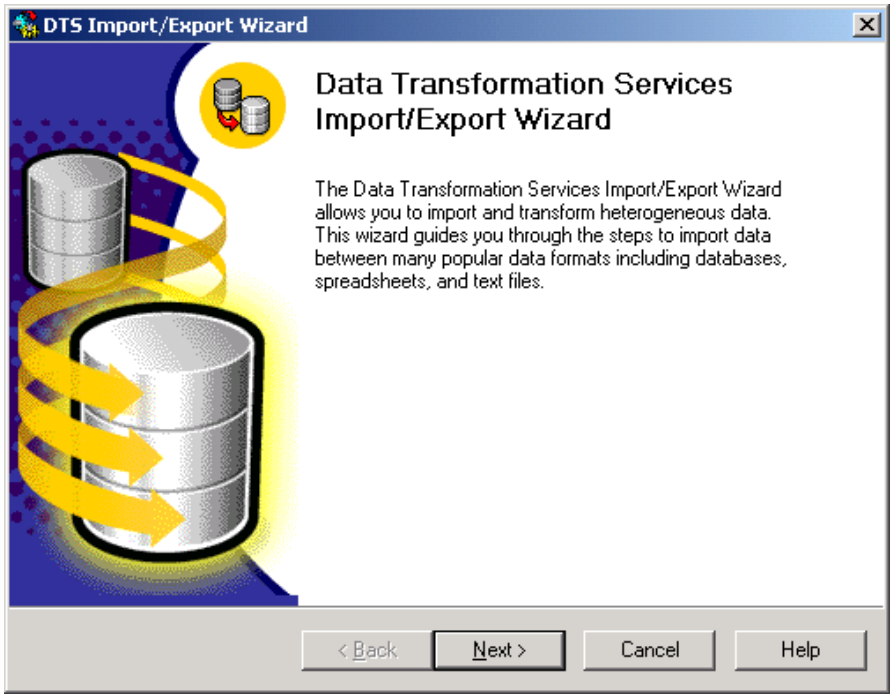
ตารางที่ 3.2

กลุ่มแฟ้มข้อมูลระบบ Account Receivable ระบบงานบัญชีลูกหนี้ การขายสินค้าและบริการ		
1.	**ARCUST**	แฟ้มหลักลูกค้า
2.	**ARCLASS**	แฟ้มอ้างอิงกลุ่มประเภทลูกค้า
3.	**ARHDRH**	แฟ้มประวัติ History Transaction การขายส่วนรายการการขายแต่ละบิล
4.	**ARLINH**	แฟ้มประวัติ Transaction รายการสินค้าที่ทำการขายแต่ละบิล แสดงจำนวน ราคา ส่วนลด

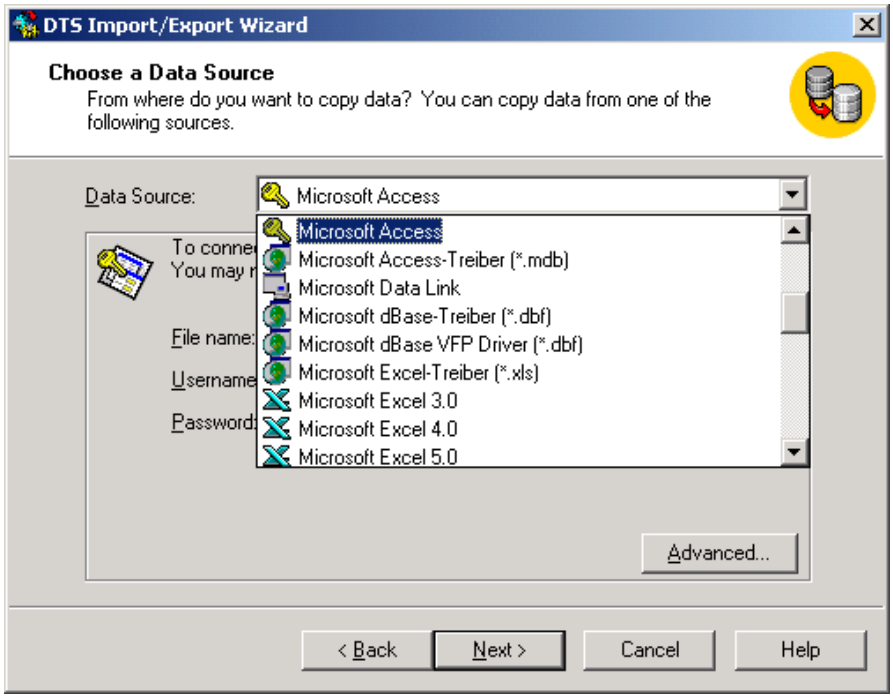
หลังจากได้ทำการคัดเลือกตารางที่ต้องการได้แล้ว ทำการโอนเข้าสู่ฐานข้อมูล Microsoft SQL Server ด้วยโปรแกรม Data Transformation Service ซึ่งตารางต่างๆที่ได้ยังคงมีโครงสร้างของข้อมูลที่เหมือนกับโครงสร้างเดิมทุกประการ

ตัวอย่างของวิธีการโอนตารางข้อมูลจากฐานข้อมูล Microsoft Access ไปยังฐานข้อมูล Microsoft SQL Server

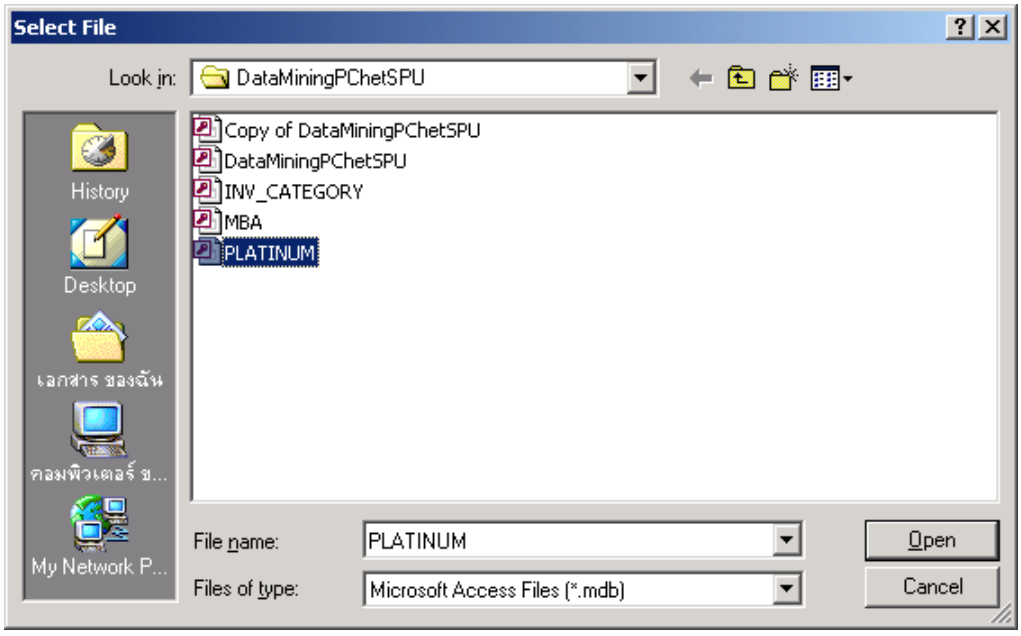
1. เรียกใช้โปรแกรม Import and Export Data จาก Microsoft SQL Server



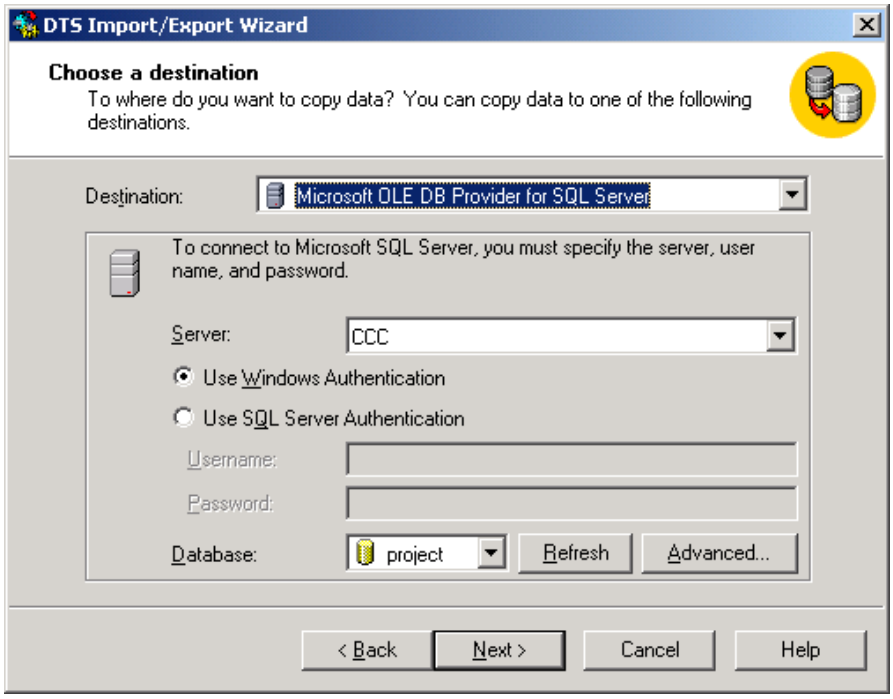
2. เลือกแหล่งข้อมูลที่เป็นต้นทาง ในที่นี้คือ Microsoft Access



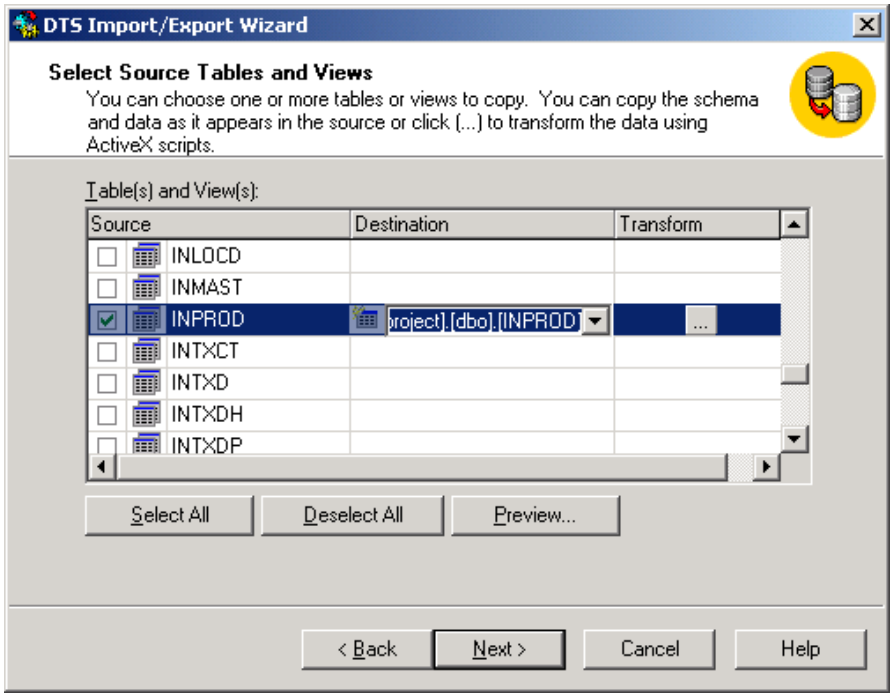
3. เลือกแฟ้มข้อมูลที่มีตารางต่างๆที่ต้องการ



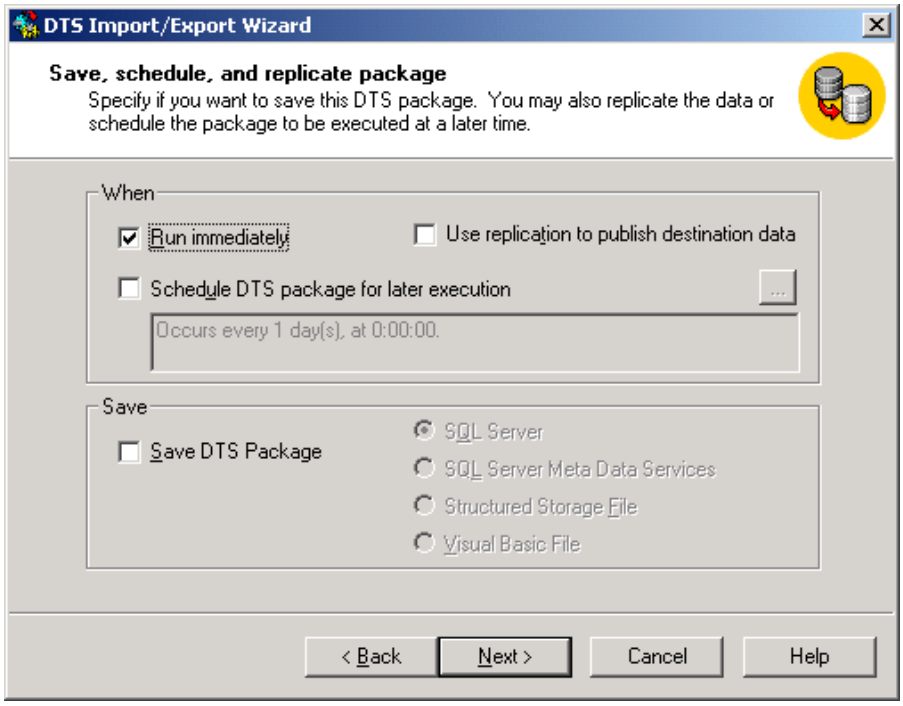
4. เลือกแหล่งข้อมูลที่เป็นปลายทาง ในที่นี้คือฐานข้อมูล Microsoft SQL Server

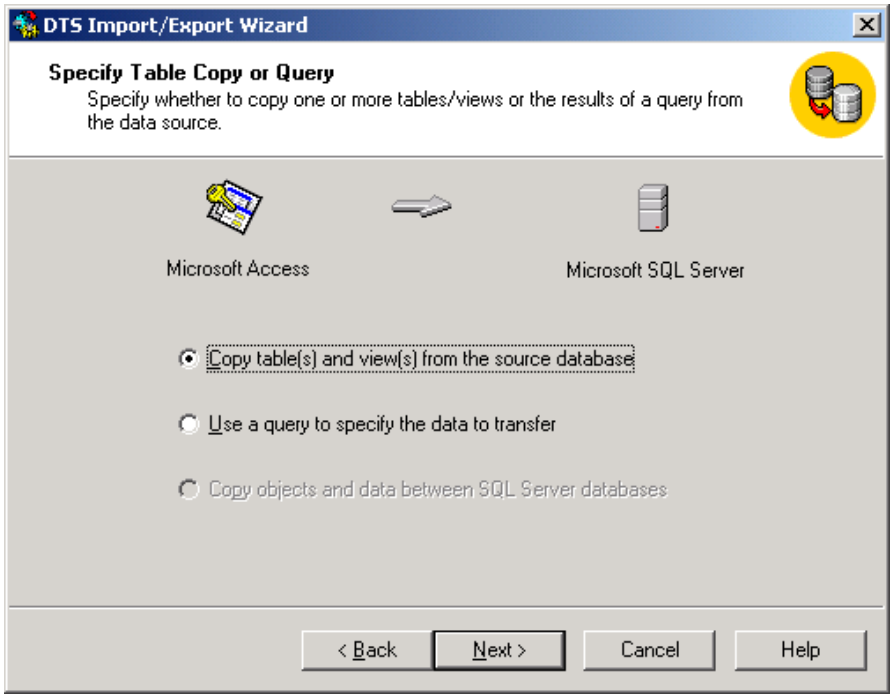


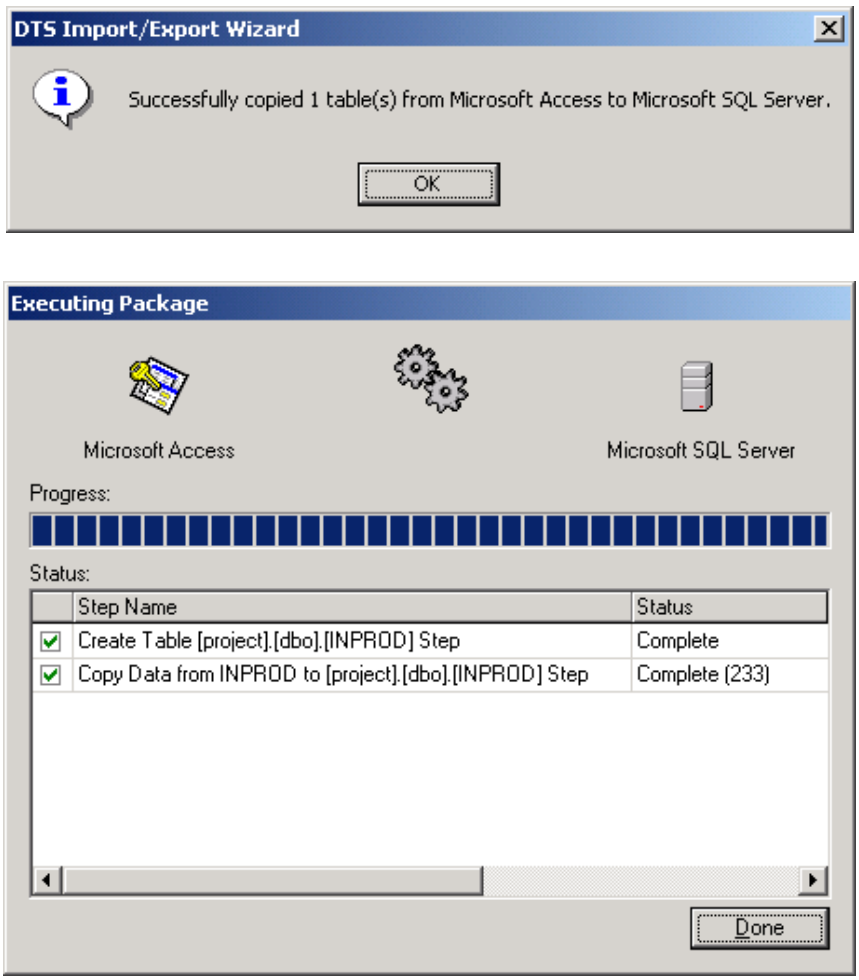
5. เลือกตารางที่ต้องการโอนจากแหล่งข้อมูลต้นทาง และกำหนดชื่อของตารางที่รับมาอยู่ในฐานข้อมูลปลายทาง



6. ภาพต่อไปนี้จะแสดงการโอนข้อมูล และจำนวนแถวของข้อมูลที่ทำการโอนได้ของแต่ละตาราง







ในการโอนข้อมูลระหว่างฐานข้อมูล สามารถเลือกโอนตารางที่ต้องการได้ มากกว่าหนึ่งตารางหรือทั้งหมดในฐานข้อมูลได้

ทำการคัดเลือกข้อมูลจากตารางต่างๆที่ได้ โดยที่ข้อมูลที่ต้องการเป็นข้อมูลที่เกิดขึ้นในระหว่างปี พ.ศ. 2546 เท่านั้น ข้อมูลที่ไม่ต้องการจะทำการลบออกจากตารางด้วยโปรแกรม Query Analyzer ข้อมูลที่เหลืออยู่ในตารางต่างๆจะเป็นข้อมูลที่ต้องการเท่านั้นและยังคงความสัมพันธ์ระหว่างตารางต่างๆ

3.2.2 การทำความสะอาดข้อมูล (Data Cleaning)

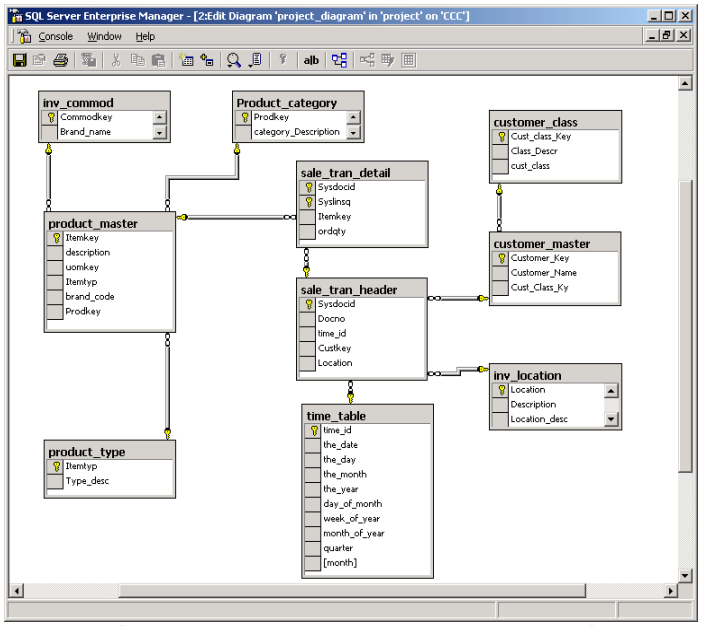
ทำการลดขนาดและปริมาณของข้อมูล ให้เหลือเฉพาะข้อมูลที่สนใจ และมีผลในการนำไปทำเหมืองข้อมูล นอกจากนี้ยังเป็นการทำความสะอาดข้อมูลที่ผิดพลาดจากการใช้ที่ผ่านมา ให้มีความถูกต้องสมบูรณ์ วิธีการทำความสะอาดข้อมูลต่างๆมีดังนี้

- เลือกคอลัมน์สำคัญในแต่ละตารางเพื่อใช้ในการทำเหมืองข้อมูลต่อไป และทำการลบคอลัมน์ที่ไม่ต้องการออกจากตาราง
- เลือกข้อมูลที่เป็นธุรกรรมเกี่ยวกับการขาย (Sale Transactions) และทำการลบข้อมูลที่ไม่เกี่ยวข้อง
- แก้ไขความสัมพันธ์ที่ไม่สมบูรณ์ของข้อมูล เช่น แก้ไขข้อมูลที่มีอยู่ในแฟ้มข้อมูลรองแต่ไม่มีในแฟ้มข้อมูลหลัก
- แก้ไขค่าว่างของข้อมูล (NULL) โดยการแทนค่าด้วยข้อมูลที่ใกล้เคียงซึ่งสอดคล้องกับโครงสร้างของระบบ
- ลบข้อมูลที่มีความซ้ำซ้อน และข้อมูลที่ไม่มีความสัมพันธ์กับข้อมูลในตารางอื่น เช่น ข้อมูลในแฟ้มข้อมูลหลักไม่ได้ถูกใช้โดยแฟ้มข้อมูลรอง

ขนาดของข้อมูลที่ได้ผ่านการทำความสะอาดเพื่อให้มีความสมบูรณ์ในแง่ของข้อมูลและประสิทธิภาพที่จะได้จากการทำเหมืองข้อมูล จะมีขนาดเล็กลง

3.2.3 การรวมข้อมูลให้เป็นหนึ่งเดียว (Data Integration)

เป็นขั้นตอนของการสร้างตารางต่างๆขึ้นมาใหม่ด้วยข้อมูลที่ผ่านการทำความสะอาดแล้ว และทำการสร้างความสัมพันธ์ระหว่างตารางต่างๆที่สร้างขึ้น โดยที่ยังคงความสมบูรณ์และความถูกต้องของข้อมูลและความสัมพันธ์ของข้อมูลเดิม โดยใช้หลักการของการออกแบบฐานข้อมูล ให้อยู่ในรูปแบบของ3rd Normal Form เพื่อลดความซ้ำซ้อนของข้อมูล ซึ่งเป็นสิ่งที่จำเป็นสำหรับฐานข้อมูลที่มีขนาดใหญ่



รูปที่ 3.2 แสดงความสัมพันธ์ต่างๆของตารางที่ได้

รายละเอียดของตารางต่างๆ

ตารางที่ 3.3 ตาราง Product Master

ลำดับ	ชื่อคอลัมน์	ประเภทข้อมูล	ความหมาย
1	ITEM KEY	Nvarchar(18)	รหัสสินค้า
2	DESCRIPTION	Nvarchar(40)	ชื่อสินค้า
3	UOMKEY	Nvarchar(8)	รหัสหน่วยนับ
4	ITEMTYP	Nvarchar(3)	รหัสประเภทสินค้า
5	BRAND_CODE	Nvarchar(8)	รหัสยี่ห้อสินค้า
6	PRODKEY	Nvarchar(8)	รหัสกลุ่มสินค้า

ตารางที่ 3.4 ตาราง Product Category

ลำดับ	ชื่อคอลัมน์	ประเภทข้อมูล	ความหมาย
1	PRODKEY	Nvarchar(8)	รหัสกลุ่มสินค้า
2	CATEGORY_DESCRIPTION	Nvarchar(50)	ชื่อกลุ่มสินค้า

ตารางที่ 3.5 ตาราง Product Type

ลำดับ	ชื่อคอลัมน์	ประเภทข้อมูล	ความหมาย
1	ITEMTYP	Nvarchar(3)	รหัสหน่วยนับ
2	TYPE_DESC	Nvarchar(30)	ชื่อหน่วยนับ

ตารางที่ 3.6 ตาราง Inv commod

ลำดับ	ชื่อคอลัมน์	ประเภทข้อมูล	ความหมาย
1	COMMOD_KEY	Nvarchar(8)	รหัสยี่ห้อสินค้า
2	BRAND_NAME	Nvarchar(30)	ยี่ห้อสินค้า

ตารางที่ 3.7 ตาราง Costomer Master

ลำดับ	ชื่อคอลัมน์	ประเภทข้อมูล	ความหมาย
1	CUSTOMER_KEY	Nvarchar(5)	รหัสลูกค้า
2	CUSTOMER_NAME	Nvarchar(30)	ชื่อลูกค้า
3	CUST_CLASS_KY	Nvarchar(5)	รหัสกลุ่มลูกค้า

ตารางที่ 3.8 ตาราง Customer Class

ลำดับ	ชื่อคอลัมน์	ประเภทข้อมูล	ความหมาย
1	CUST_CLASS_KEY	Nvarchar(5)	รหัสกลุ่มลูกค้า
2	CUST_DEST	Nvarchar(30)	ชื่อกลุ่มลูกค้า (ไทย)
3	CUST_CLASS	Nvarchar(50)	ชื่อกลุ่มลูกค้า (อังกฤษ)

ตารงที่ 3.9 ตาราง Inv Location

ลำดับ	ชื่อคอลัมน์	ประเภทข้อมูล	ความหมาย
1	LOCATION	Nvarchar(5)	รหัสสาขา
2	DESCRIPTION	Nvarchar(30)	ชื่อสาขา (ไทย)
3	LOCATION_DESC	Nvarchar(30)	ชื่อสาขา (อังกฤษ)

ตารางที่ 3.10 ตาราง Time Table

ลำดับ	ชื่อคอลัมน์	ประเภทข้อมูล	ความหมาย
1	TIME_ID	Int(4)	รหัสคาบเวลา
2	THE_DATE	Datetime	วันที่
3	THE_DAY	Nvarchar(20)	วันในสัปดาห์
4	THE_MONTH	Nvarchar(20)	เดือนในปี
5	THE_YEAR	Int(4)	ปี
6	DAY_OF_MONTH	Int(4)	ลำดับของวันที่ในเดือน
7	WEEK_OF_YEAR	Int(4)	สัปดาห์ที่ของเดือน
8	MONT_OF_YEAR	Int(4)	เดือนที่ของปี
9	QUARTER	Nvarchar(5)	ไตรมาส
10	MONTH	Nvarchar(20)	เดือน

ตารางที่ 3.11 ตาราง Sale Header

ลำดับ	ชื่อคอลัมน์	ประเภทข้อมูล	ความหมาย
1	SYSDOCID	Int(4)	หมายเลขจากระบบ
2	DOCNO	Nvarchar(5)	เลขที่เอกสาร
3	TIME_ID	Int(4)	รหัสคาบเวลา
4	CUSTKEY	Nvarchar(5)	รหัสลูกค้า
5	LOCATION	Nvarchar(5)	สาขา

ตารางที่ 3.12 ตาราง Sale Detail

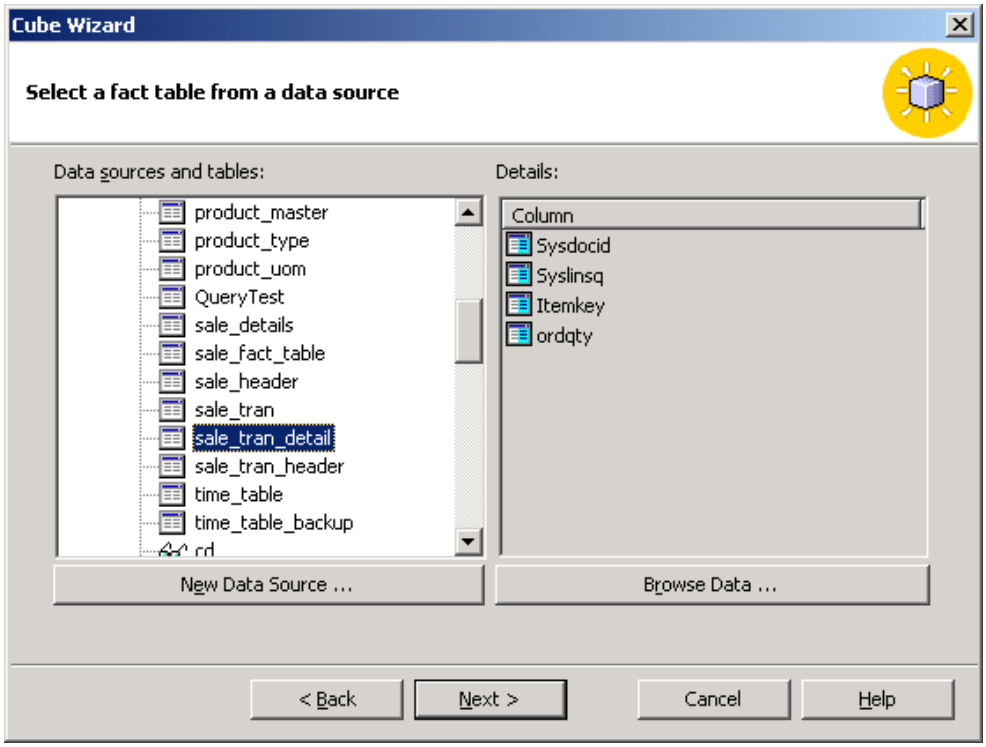
ลำดับ	ชื่อคอลัมน์	ประเภทข้อมูล	ความหมาย
1	SYSDOCID	Int(4)	หมายเลขจากระบบ
2	SYSLINSQ	Int(4)	ลำดับที่
3	ITEMKEY	Nvarchar(18)	รหัสสินค้า
4	ORDQTY	Float(8)	ปริมาณสั่งซื้อ

3.2.4 การเปลี่ยนรูปแบบหรือตัดทอนข้อมูล (Data Reduction)

โดยการนำข้อมูลจากตารางต่างๆที่ได้มาอยู่รวมกันในรูปแบบของลูกบาศก์ข้อมูล (Data Cube Aggregation) และในรูปแบบของมิติต่างๆของข้อมูล (Dimensionality Reduction) ด้วยการใช้โปรแกรม Microsoft Analysis Manager จุดประสงค์ของการเปลี่ยนรูปแบบของข้อมูลให้อยู่ในรูปแบบดังกล่าว เพื่อใช้ในการวิเคราะห์ข้อมูล และใช้กับโปรแกรม DBMiner ที่เป็นเครื่องมือในการทำเหมืองข้อมูลต่อไป

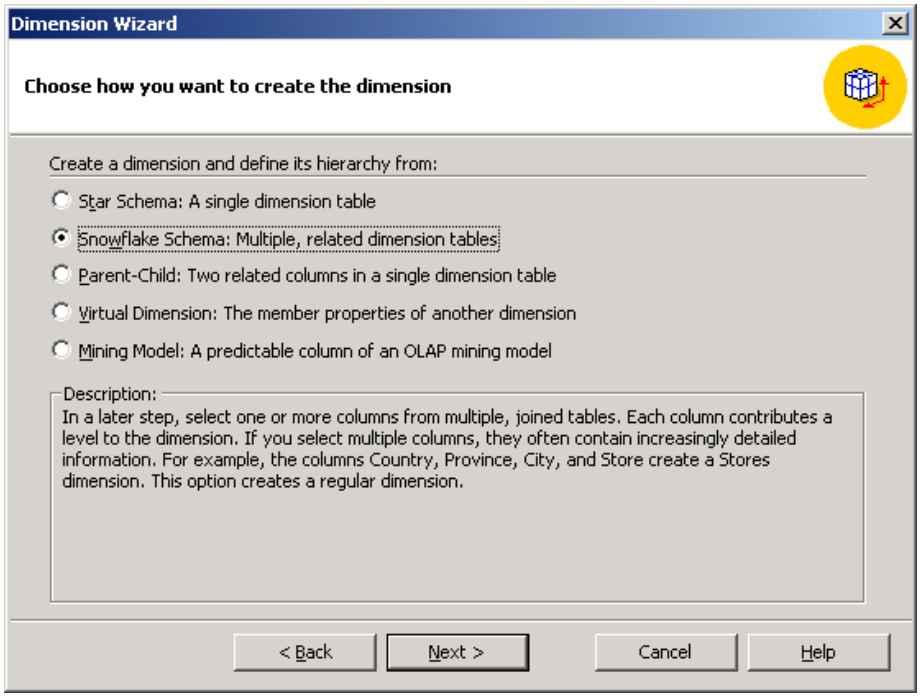
ขั้นตอนในการสร้าง Data cube หรือ OLAP cube ด้วยโปรแกรม Microsoft Analysis Manager ของ Microsoft SQL Server มีดังต่อไปนี้

1. กำหนดตารางหลัก (Fact Table) ของ Data cube ในที่นี้คือตาราง Sale_tran_detail โดยใช้คอลัมน์ Ordqty เป็นหน่วยวัด (Measure)

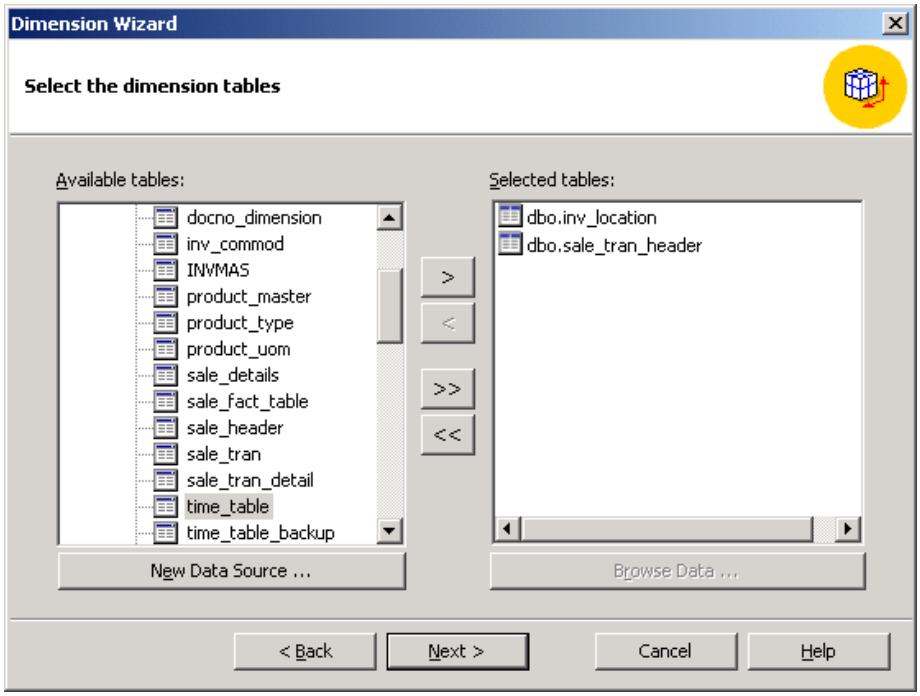


2. สร้างมิติหรือมุมมองต่างของตารางหลัก ซึ่งจะสัมพันธ์กับโครงสร้างของฐานข้อมูลที่ได้ออกแบบรองรับไว้แล้ว
   ตัวอย่างของขั้นตอนในการสร้างมิติต่างๆ มีดังนี้

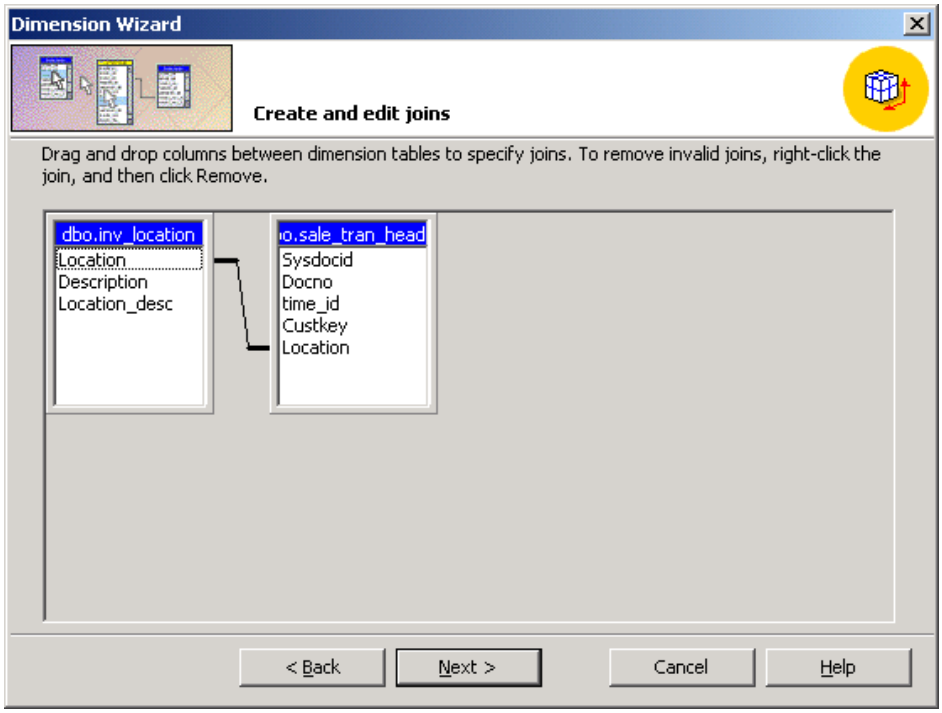
1. เลือกรูปแบบของมิติที่ต้องการซึ่งจะสอดคล้องกับโครงสร้างของฐานข้อมูล



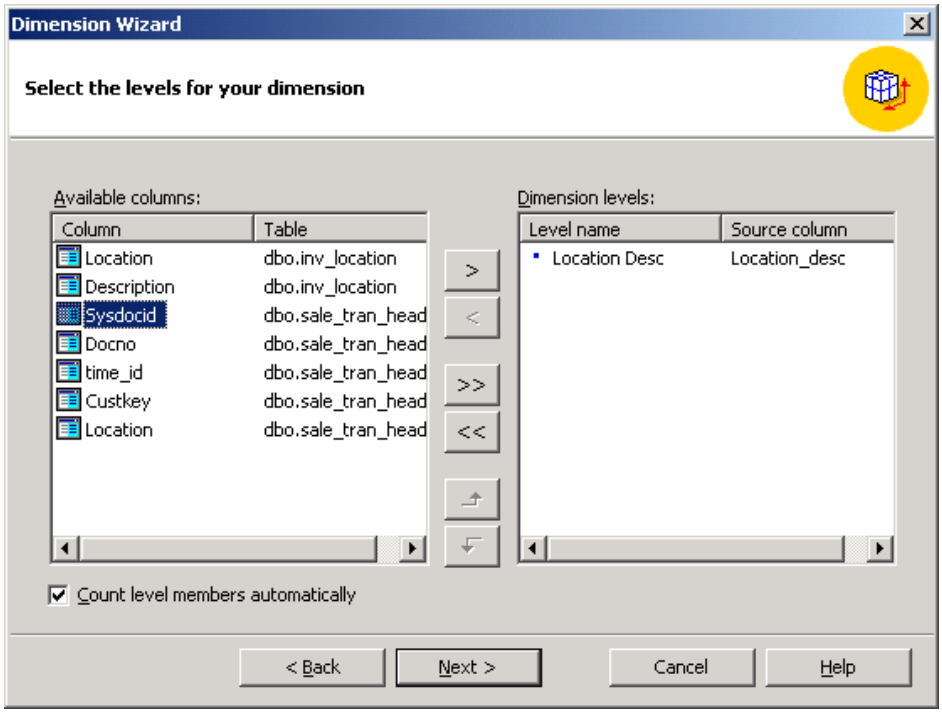
2. เลือกตารางที่ต้องการ



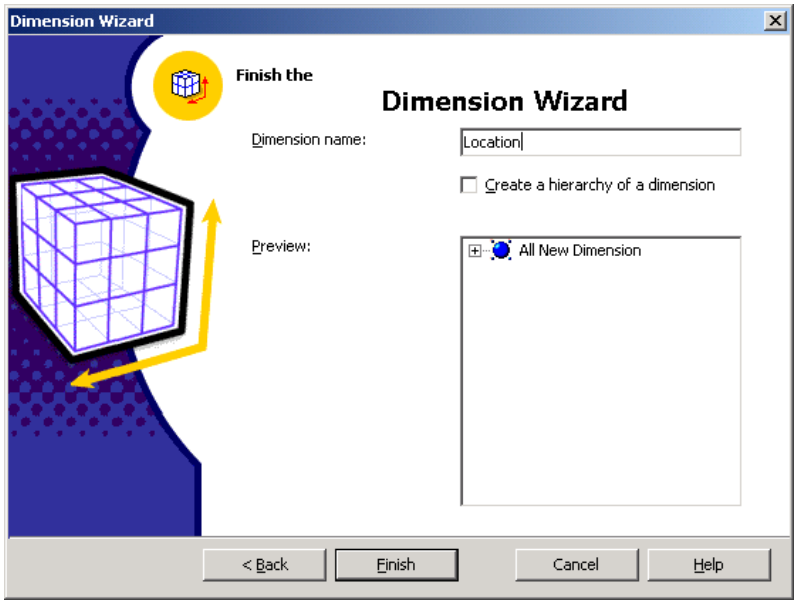
3. สร้างความสัมพันธ์ระหว่างตาราง ถ้ายังไม่มีการสร้างความสัมพันธ์ระหว่างตารางกันในขั้นตอนของการออกแบบฐานข้อมูล



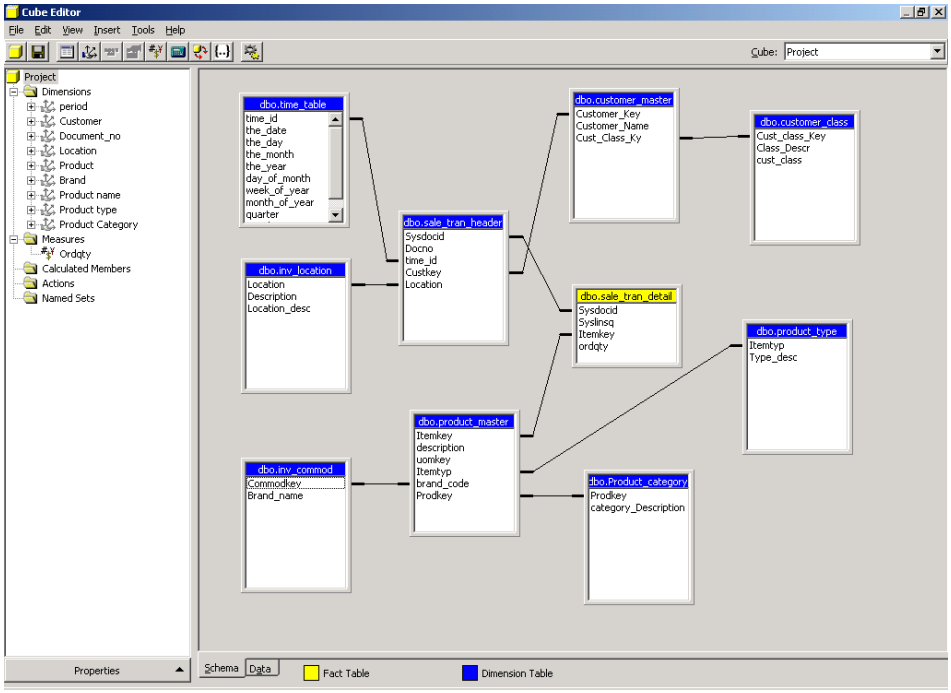
4. เลือกระดับของมิติโดยการเลือกคอลัมน์ที่ต้องการ



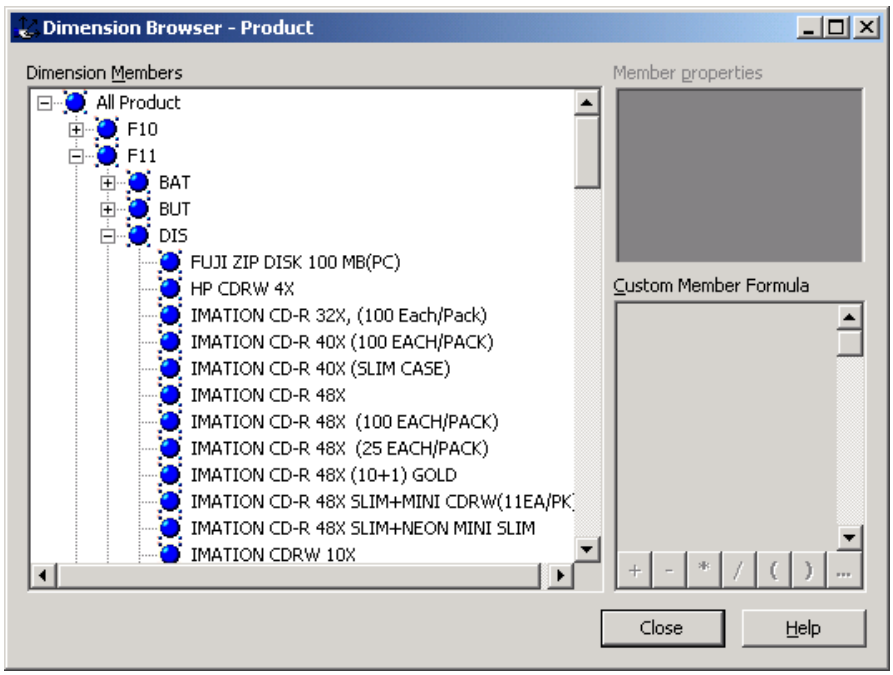
5. ตั้งชื่อให้กับมิติที่สร้าง



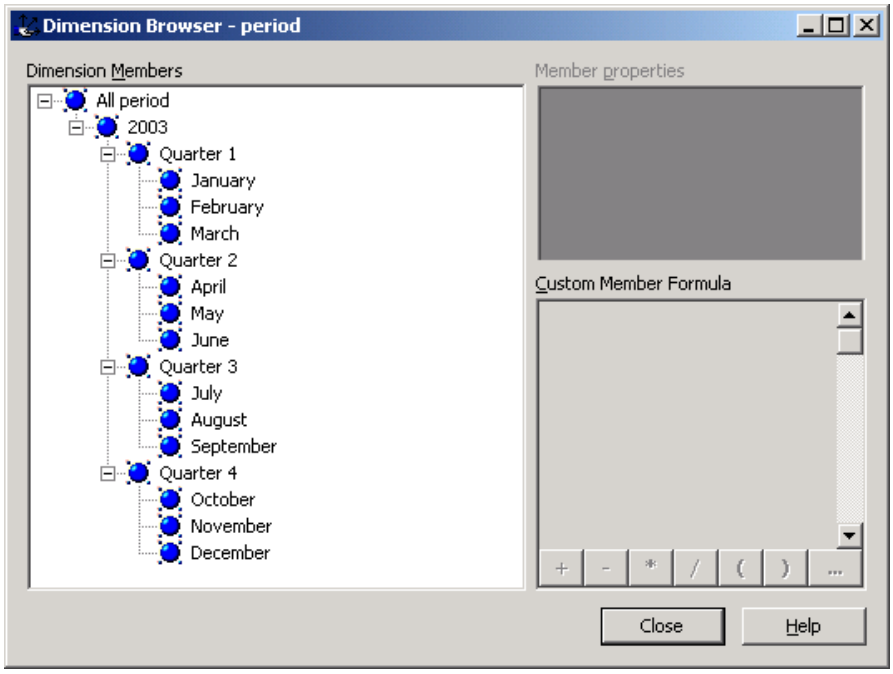
ทำการสร้างมิติต่างที่ต้องการด้วยวิธีการเดียวกันนี้ เพื่อให้ได้มิติหรือมุมมองที่ต้องการ เพื่อใช้ในการวิเคราะห์ข้อมูลในลักษณะของ Online Analysis Processing (OLAP) และใช้ในการทำเหมืองข้อมูลด้วยโปรแกรม DBMiner ต่อไป



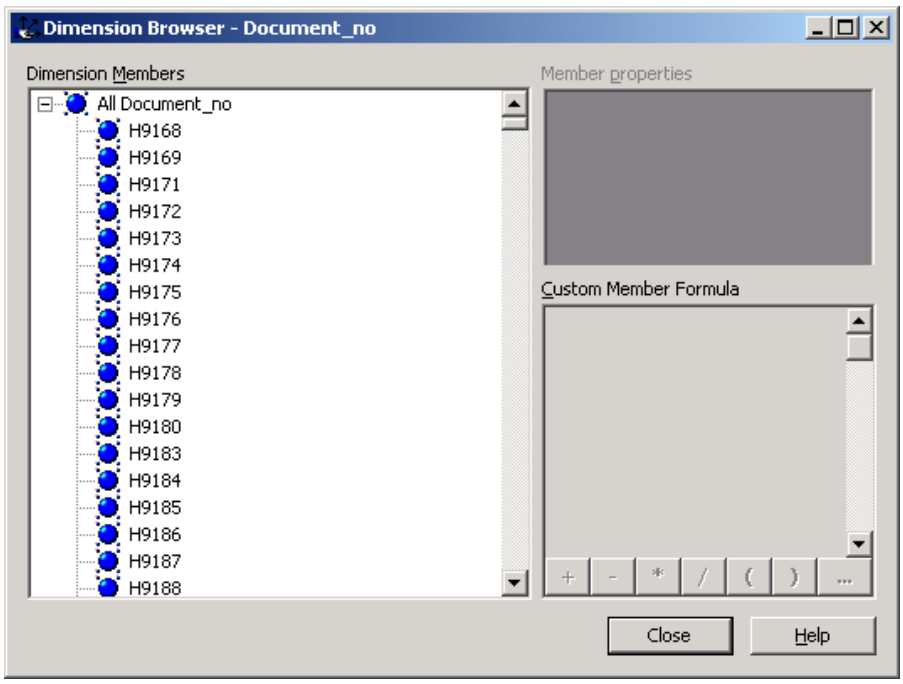
รูปที่ 3.3 แสดงความสัมพันธ์ของ Data Cube



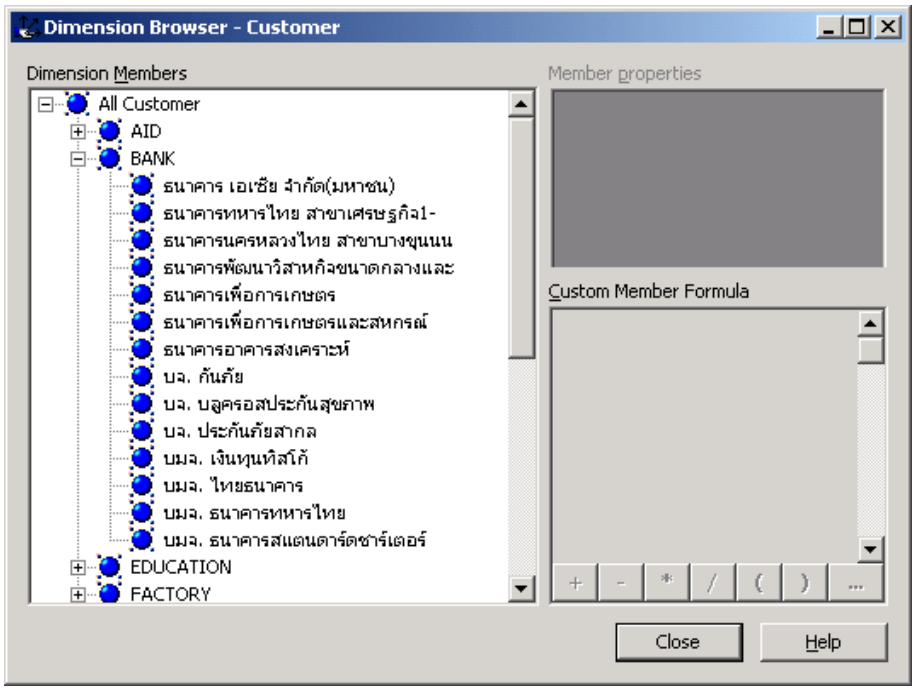
ตัวอย่างข้อมูลในมิติของ Product



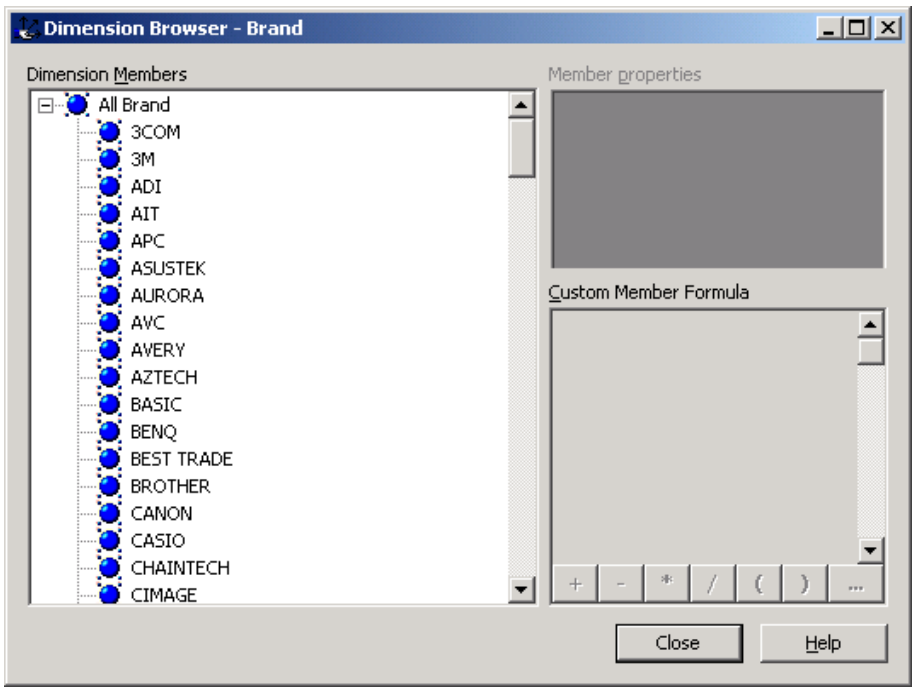
ตัวอย่างข้อมูลในมิติของ Period



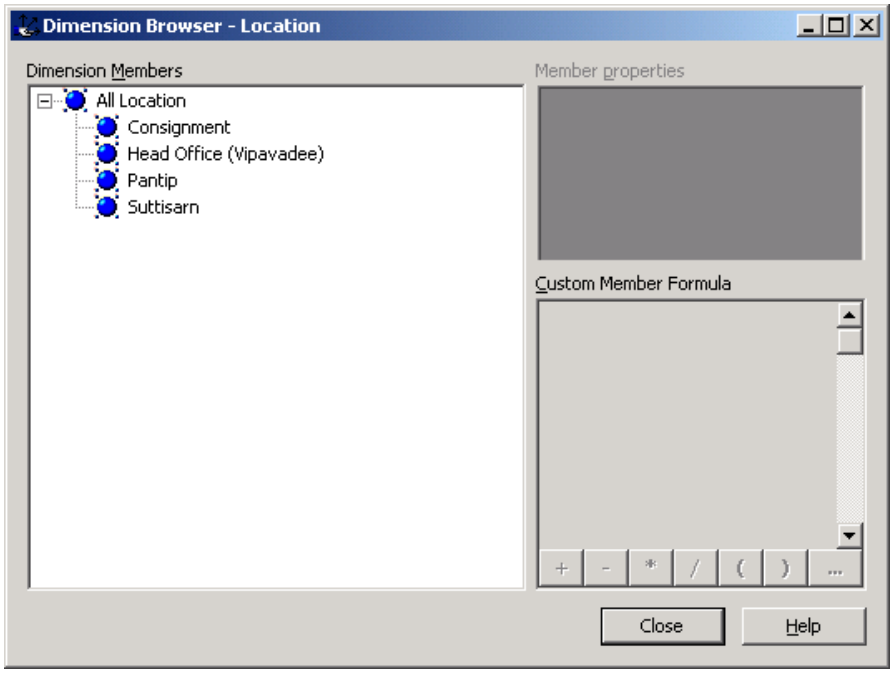
ตัวอย่างข้อมูลในมิติของ Document_no



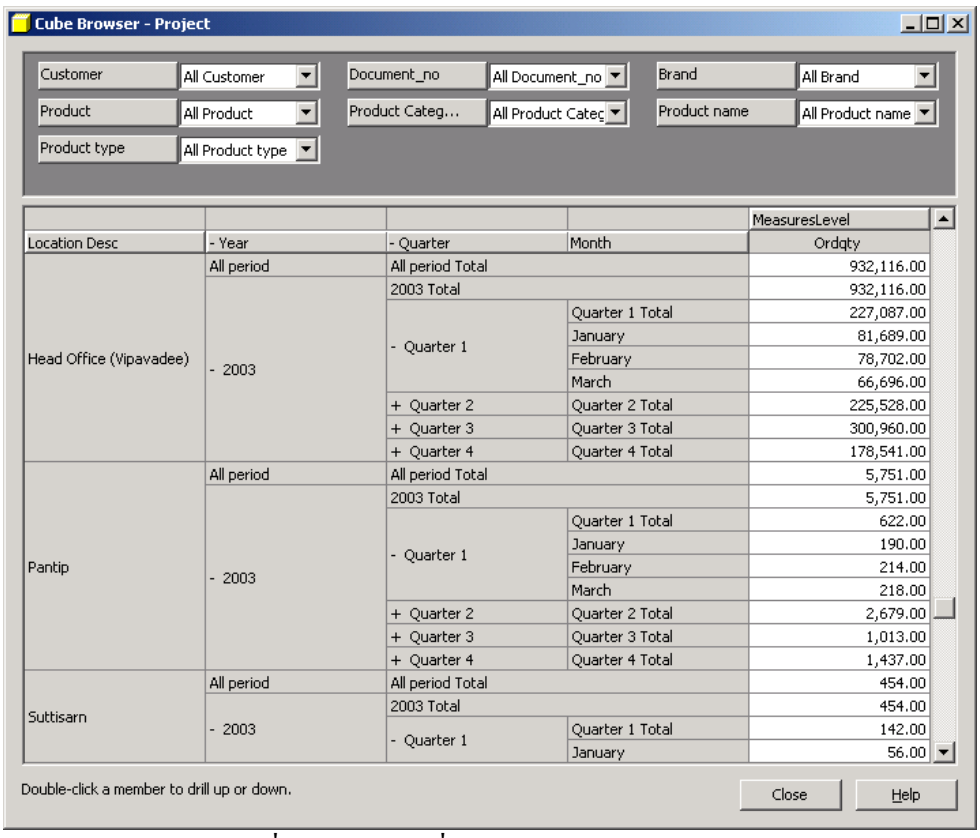
ตัวอย่างข้อมูลในมิติของ Customer



ตัวอย่างข้อมูลในมิติของ Brand



ตัวอย่างข้อมูลในมิติของ Location



รูปที่ 3.4 แสดงผลที่ได้จากการสร้าง Data Cube

Data Cube ที่ได้สามารถแสดงผลเพื่อการวิเคราะห์ข้อมูลในมิติหรือมุมมองต่างๆได้ จากตัวอย่างในรูปที่ 3.4 เป็นการแสดงผลในมิติของ สาขาและคาบเวลา นั่นคือจะแสดงให้เห็นว่าในแต่ละสาขามียอดขายในปี พ.ศ. 2546 และในแต่ละไตรมาสและในแต่ละเดือน เป็นจำนวนเท่าไร