

คู่มือปฏิบัติการ Data warehouse: Northwind-Star Schema

วัตถุประสงค์ของแบบฝึกหัด

- เพื่อเตรียมข้อมูลสำหรับใช้ในคลังข้อมูล
- เพื่อให้ทราบวิธีการนำเสนอข้อมูลจากฐานข้อมูล(OLTP) เข้าสู่คลังข้อมูล (Data-warehouse)

สิ่งที่ได้จากแบบฝึกหัด

- สามารถสร้างโครงสร้างฐานข้อมูลสำหรับเก็บข้อมูลในคลังข้อมูลได้
- สามารถสร้างและ Execute SSIS Package สำหรับดึงข้อมูลจากฐานข้อมูล OLTP เข้าสู่คลังข้อมูลได้

คำแนะนำ

ในคู่มือนี้ จะทำการดึงข้อมูลจาก Northwind จากฐานข้อมูล OLTP เข้าสู่ Northwind_Star ที่เป็นคลังข้อมูล (Data-warehouse)

ขั้นตอนในการดึงข้อมูลจาก Northwind สู Northwind_star

1. การเตรียมข้อมูลและโครงสร้างคลังข้อมูล

- 1.1 Attach ฐานข้อมูล Northwind
- 1.2 สร้างฐานข้อมูล Northwind_star
- 1.3 สร้างโครงสร้างตารางสำหรับคลังข้อมูล Northwind_star

2. สร้าง SSIS Package สำหรับดึงข้อมูลจาก Northwind เข้าสู่ Northwind_star (คลังข้อมูล)

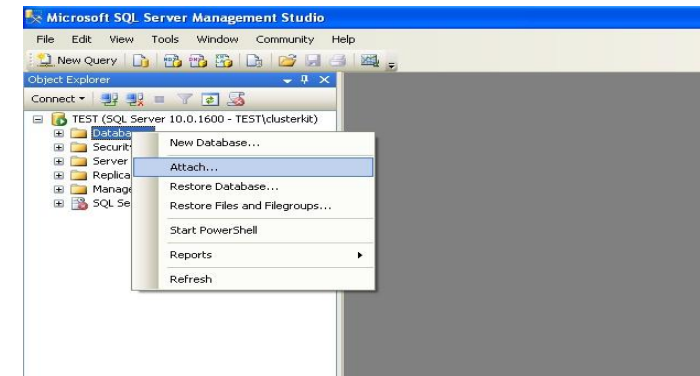
- 2.1 สร้าง Data Flow สำหรับดึงข้อมูลเข้าสู่ Dimension table
- 2.2 สร้าง Data Flow สำหรับดึงข้อมูลเข้าสู่ Fact Table
- 2.3 สร้าง Execute SQL Task สำหรับลบข้อมูลใน Dimension table และ Fact Table ก่อนดึงข้อมูลเข้ามาใหม่

3. Execute SSIS Package เพื่อดึงข้อมูลจาก Northwind เข้าสู่ Northwind_star (คลังข้อมูล)

ขั้นตอนที่ 1: การเตรียมข้อมูลและโครงสร้างคลังข้อมูล

1.1 Attach ฐานข้อมูล Northwind

1. เปิดโปรแกรม MS SQL Server Management Studio จากเมนู MS SQL Server 2008
2. คลิกขวาที่โฟลเดอร์ Databases แล้วเลือก Attach ดังรูปที่ 1



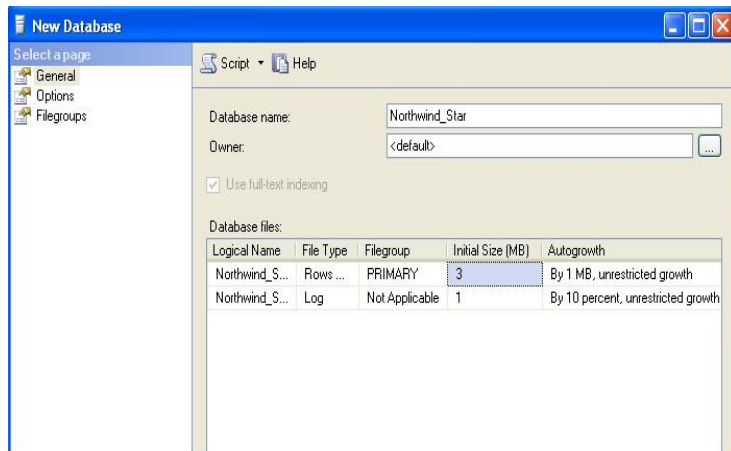
รูปที่ 1 การ Attach ฐานข้อมูลโดยใช้ MS SQL Server Management Studio

3. จะปรากฏหน้าต่าง “Attach Databases” แล้วให้กดปุ่ม Add พร้อมกับระบุไฟล์ฐานข้อมูล NORTHWIND.MDF

4. กดปุ่ม OK

1.2 สร้างฐานข้อมูล Northwind_star

5. คลิกขวาที่โฟลเดอร์ Databases อีกครั้งแล้วเลือก New Databases... จะปรากฏหน้าต่างดังรูปที่ 2 กำหนดค่าต่างๆ ดังตารางข้างล่าง แล้วกดปุ่ม OK



รูปที่ 2 แสดงหน้าต่างสำหรับ New Database

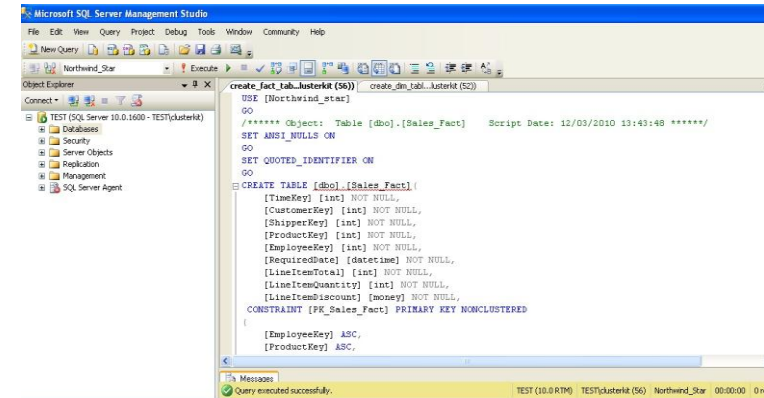
กำหนดค่าต่างๆ สำหรับ Databases ดังต่อไปนี้

General Page:	Name	--	Northwind_star
Option Page:	Collation Name	--	Thai_CL_AS

1.3 สร้างโครงสร้างตารางสำหรับคลังข้อมูล Northwind_star

1.3.1 เปิดไฟล์ create_dim_tables.sql โดยการ Drag Drop ไฟล์เข้าไปที่ Management Studio หรือ กดดับเบิลคลิกที่ไฟล์ก็ได้ (ในกรณีที่ Default program ที่จะใช้เปิดไฟล์ .sql เป็นโปรแกรม SQL Server Management Studio)

1.3.2 กดปุ่ม Execute



รูปที่ 3 การสร้างโครงสร้างตาราง Dimension Table โดยทำการ Execute คำสั่ง SQL

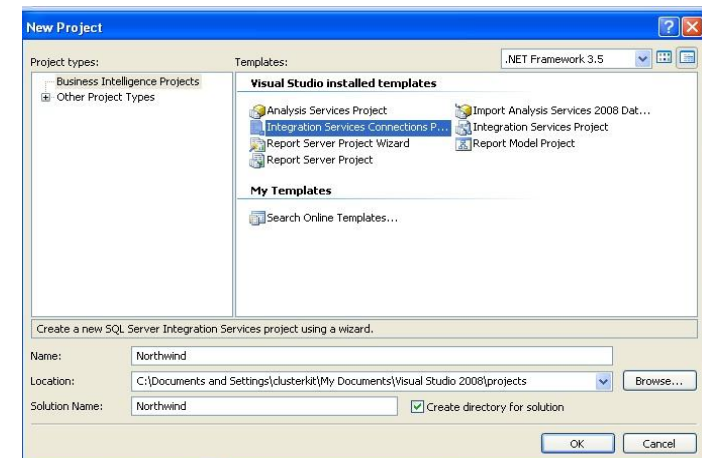
7. สร้างโครงสร้างตาราง Fact Table (กระทำเช่นเดียวกับข้อที่ 6)

- เปิดไฟล์ create_fact_table.sql โดยการ Drag Drop ไฟล์เข้าไปที่ Management Studio หรือ กดดับเบิลคลิกที่ไฟล์ก็ได้ (ในกรณีที่ Default program ที่จะใช้เปิดไฟล์ .sql เป็นโปรแกรม SQL Server Management Studio)

- กดปุ่ม Execute

ขั้นตอนที่ 2 การสร้าง DTS Package สำหรับดึงข้อมูลจาก Northwind เข้าสู่ Northwind_star

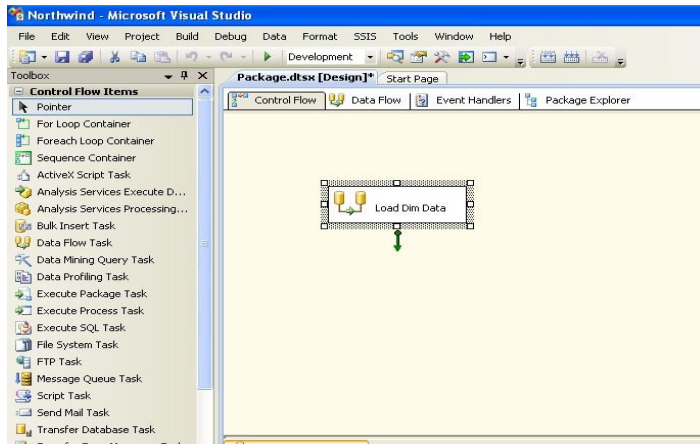
8. เปิดโปรแกรม SQL Server Business Intelligence Development Studio จากเมนู MS SQL Server 2008 แล้วคลิกที่เมนู File เลือก New Project แล้วกำหนดรายละเอียดดังรูปที่ 4



รูปที่ 4 แสดงรายละเอียดการสร้าง Project ใหม่

2.1 สร้าง Data Flow สำหรับดึงข้อมูลเข้าสู่ Dimension table

9. คลิกที่ Tab Control Flow หลังจากนั้น Double-clicks ที่ Data Flow Task จากเมนูย่อย Control Flow Items ใน Tool box ด้านซ้ายมือ ดังรูปที่ 5 แล้วเปลี่ยนชื่อเป็น “Load Dim Data”

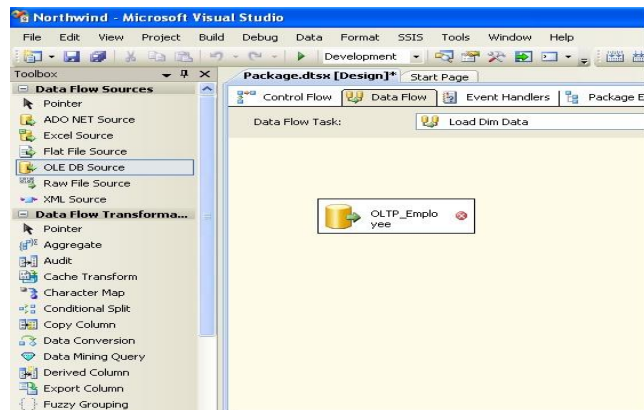


รูปที่ 5 แสดงหน้าต่างการสร้าง DataFlow

10. Double-clicks ที่ Data Flow “Load Dim Data” จะเข้าสู่ Tab Data Flow

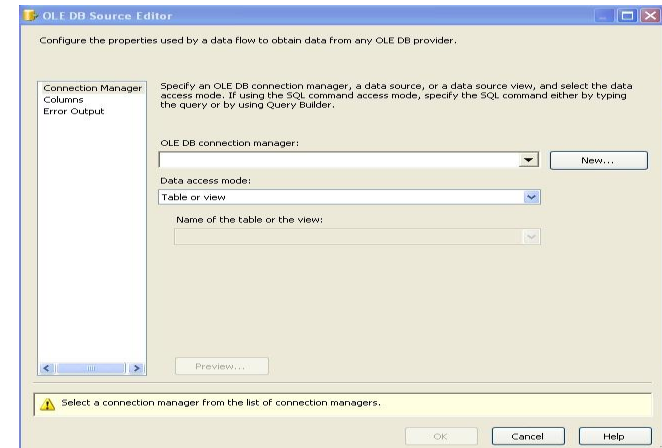
2.1.1 สร้างการดึงข้อมูลเข้าสู่คลังข้อมูลในตาราง Employee_Dim

11. Double-clicks ที่ OLE DB Source จากเมนูย่อย Data Flow Sources ใน Tool box ด้านซ้ายมือ ดังรูปที่ 6 แล้วเปลี่ยนชื่อ OLE DB Source เป็น OLTP_Employee



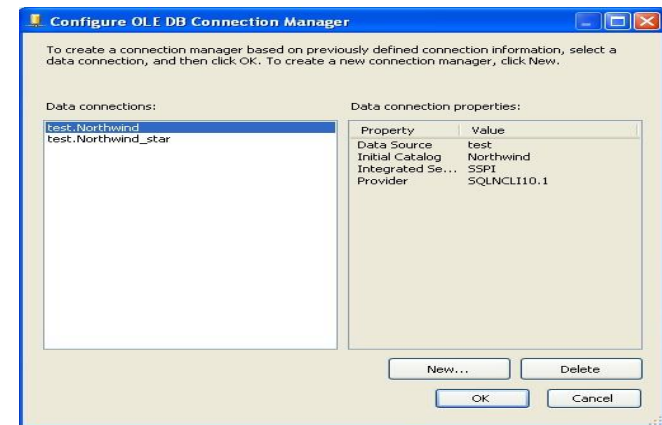
รูปที่ 6 การสร้าง OLE DB Source

12. Double-clicks ที่ กล่อง OLTP_Employee จากรูปที่ 7 จะปรากฏหน้าต่างดังรูปที่ 8 เพื่อทำการกำหนดค่าคุณสมบัติต่างๆ ของแหล่งข้อมูล ดังขั้นตอนต่อไปนี้



รูปที่ 7 หน้าต่างสำหรับกำหนดคุณสมบัติต่างๆ ของ OLE DB Source

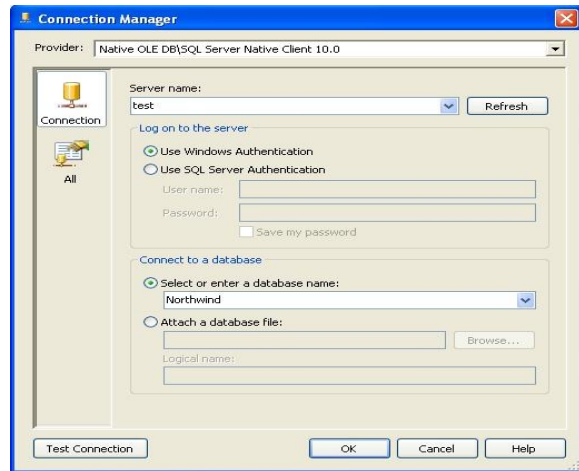
12.1 สร้าง OLE DB Connection Manager ขึ้นมาใหม่ โดยกดปุ่ม New จากรูปที่ 9 จะปรากฏหน้าต่างดังรูปที่ 8



รูปที่ 8 แสดง Connection ไปยังฐานข้อมูลที่เคยถูกสร้างไว้แล้ว

12.2 ทำการสร้าง Connection ไปยังฐานข้อมูล Northwind เนื่องจากยังไม่ปรากฏในลิสต์ จึงต้องกดปุ่ม New จากรูปที่ 8 ซึ่งจะเข้าสู่หน้าต่าง “Connection Manager”

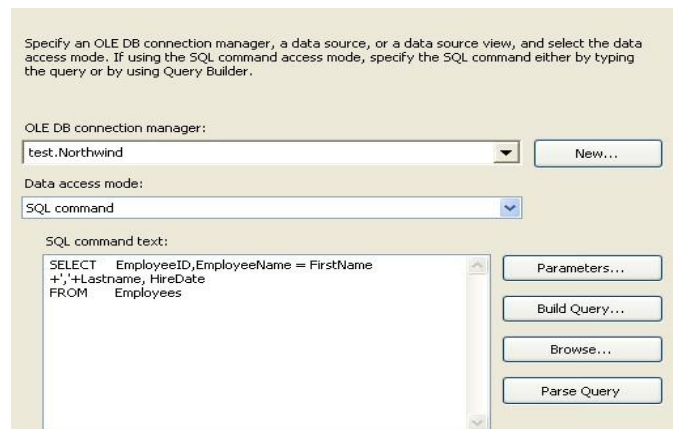
12.3 กำหนด Properties ต่างๆ แล้วกดปุ่ม OK



รูปที่ 9 Properties ต่างๆ ของ Connection Manager

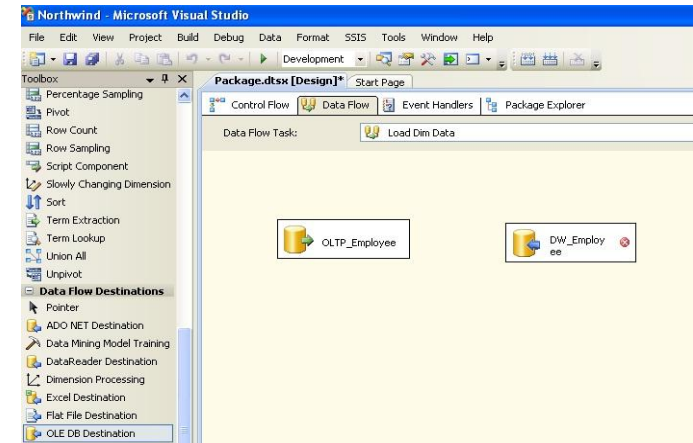
12.4 กลับสู่หน้าต่างในรูปที่ 9 ให้เลือกฐานข้อมูลที่ได้ทำการสร้างใหม่จากรูปที่ 8 แล้วกดปุ่ม OK

12.5 กลับสู่หน้าต่าง OLE DB Source Editor ดังรูปที่ 7 โดยกำหนด Properties ดังรูปที่ 10 โดยเลือกที่ Data access mode เป็น SQL Command แล้วใส่ sql command ลงไป เพื่อจะให้ ชื่อจริง และนามสกุล รวมเป็น column เดียวกัน แล้วกดปุ่ม OK



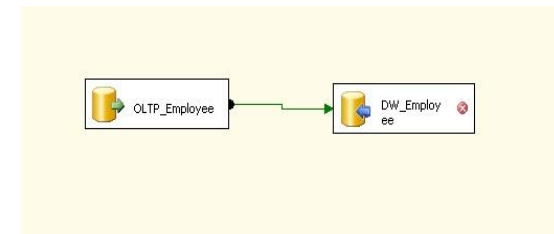
รูปที่ 10 กำหนดค่าของ OLE DB Source เป็นตาราง Employee

13. Double-clicks ตรง OLE DB Destination จากเมนูย่อย Data Flow Destination ใน Tool box ด้านซ้ายมือ ดังรูปที่ 10 แล้วเปลี่ยนชื่อ OLE DB Destination เป็น DW_Customer



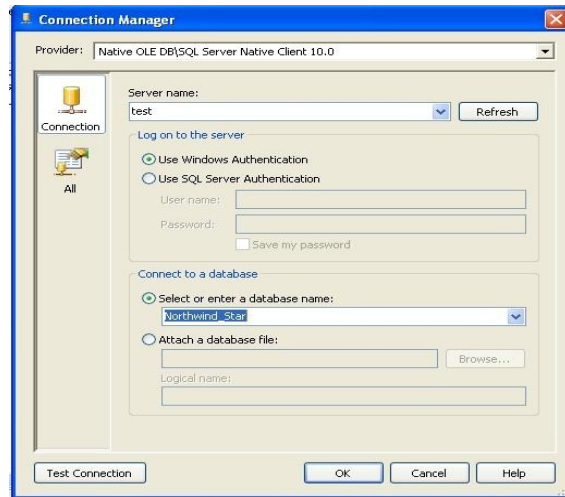
รูปที่ 11 การสร้าง OLE DB Destination

14. จากรูปที่ 13 ให้คลิกที่ OLTP_Employee จะปรากฏลูกศรสีเขียวและสีแดง ให้คลิกลาก(Drag) ลูกศรสีเขียวมาวาง(Drop) ยังกล่อง DW_Employee ดังรูปที่ 14



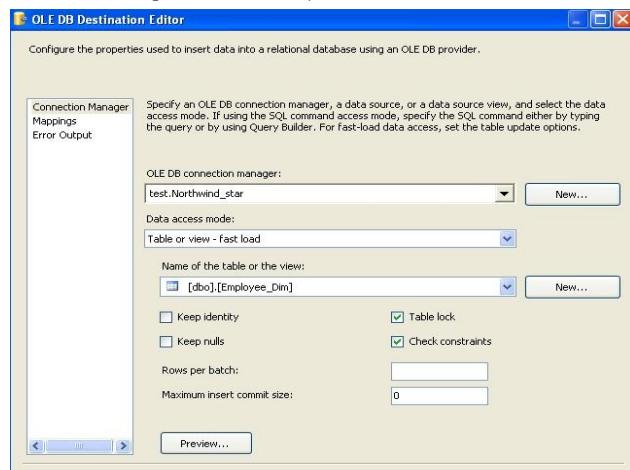
รูปที่ 12 การ Link ดึงข้อมูลจาก Source ไปยัง Destination

15. Double-clicks ที่ DW_Employee เพื่อกำหนดค่าให้กับแหล่งข้อมูลปลายทางในช่อง OLE DB Connection Manager โดยทำเหมือนกับขั้นตอนในข้อ 8 โดยเปลี่ยนจากฐานข้อมูล Northwind เป็น Northwind_star ในขั้นตอนที่ 12.3 ดังรูปที่ 13

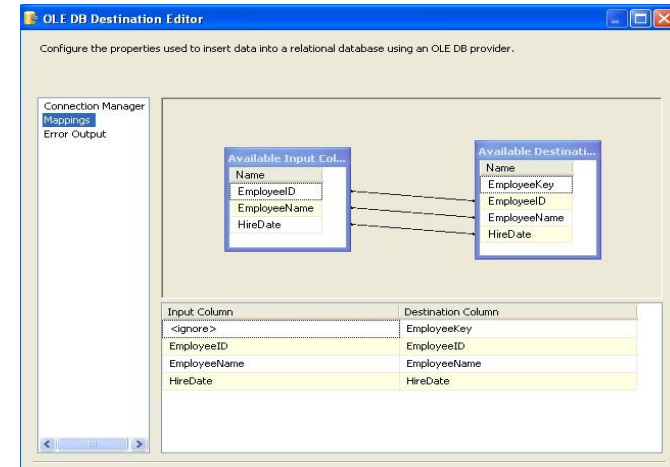


รูปที่ 13 สร้าง Connection สำหรับฐานข้อมูล Northwind_Star

16. ทำการกำหนดค่า Connection Manager สำหรับทำการเชื่อมต่อไปยังตาราง Employee_Dim ในฐานข้อมูล Northwind_star ดังรูปที่ 14 และทำการ Mapping ระหว่าง Source(OLTP_Employee) ไปยัง Destination (DW_Employee) ดังรูปที่ 15 แล้วกดปุ่ม OK



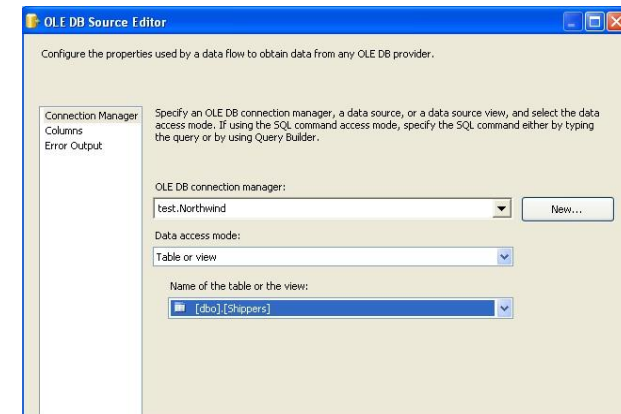
รูปที่ 14 การกำหนดค่า Connection Manager สำหรับ DW_Employee



รูปที่ 15 ค่า Mappings จาก Source (OLTP_Employees) ไปยัง Destination (DW_Employee_Dim)

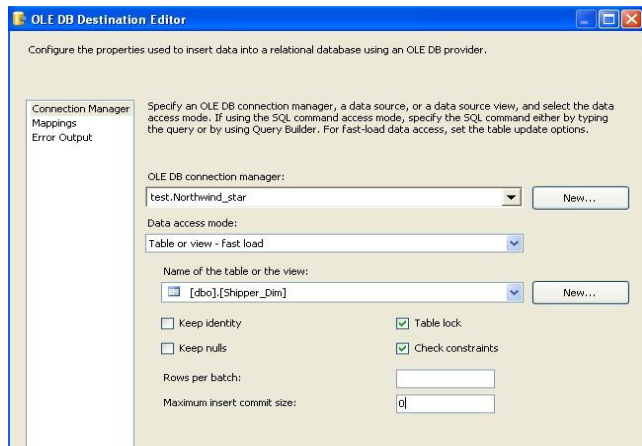
2.1.2 สร้างการดึงข้อมูลเข้าสู่คลังข้อมูลในตาราง Shipper_Dim

17. สร้าง OLTP Datasource ชื่อ OLTP_Shipper โดยทำเหมือนขั้นตอน 11 พร้อมกำหนดค่า Connection Manager ของ OLTP_Shipper ดังรูปที่ 16 แล้วกดปุ่ม OK

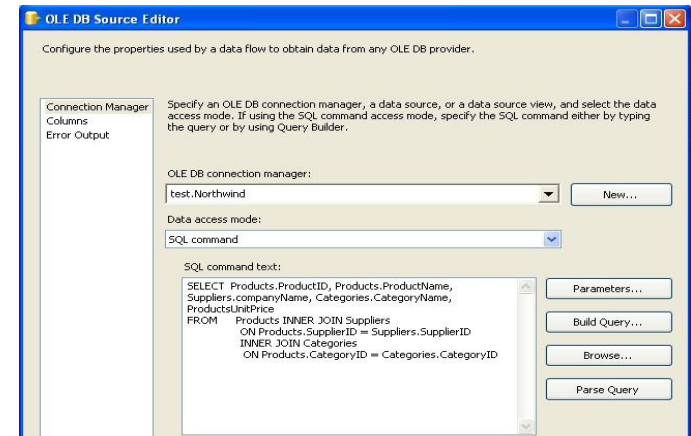


รูปที่ 16 การกำหนดค่า Connection Manager สำหรับ OLTP_Shipper

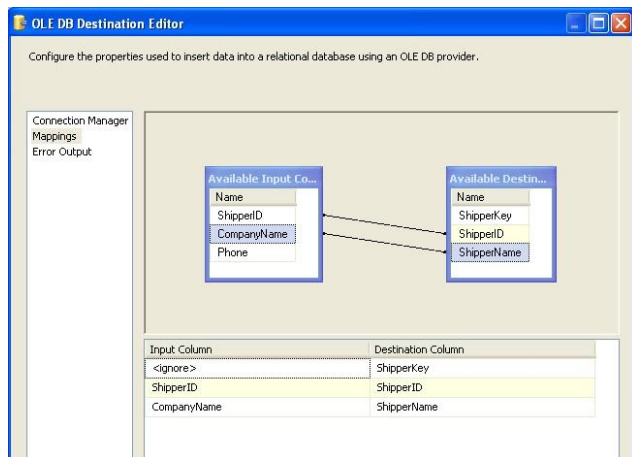
18. สร้าง OLTP Destination ชื่อ DW_Shipper พร้อมทั้งสร้าง Link ระหว่าง OLTP_Shipper ไปยัง DW_Shipper โดยทำเหมือนขั้นตอนในข้อ 11-14 โดยกำหนดค่า Connection Manager ดังรูปที่ 17 และ Mappings ของ DW_ShipperDim ดังรูปที่ 18 แล้วกดปุ่ม OK



รูปที่ 17 การกำหนดค่า Connection Manager สำหรับ DW_Shipper



รูปที่ 19 การกำหนดค่า Connection Manager สำหรับ OLTP_Product



รูปที่ 18 ค่า Mappings จาก Source (OLTP_Items) ไปยัง Destination (DW_Shipper)

2.1.3 สร้างการดึงข้อมูลเข้าสู่คลังข้อมูลในตาราง Product_Dim

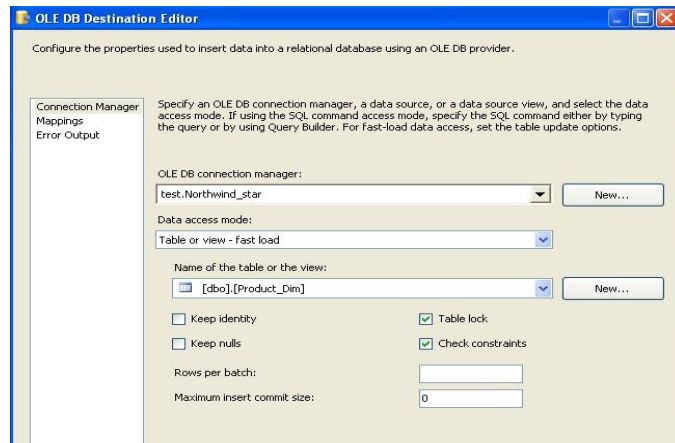
19. สร้าง OLTP Datasource ชื่อ OLTP_Product โดยทำเหมือนขั้นตอน 9 พร้อมกำหนดค่า Connection Manager ของ OLTP_Product ดังรูปที่ 19 โดยเลือก Data access mode เป็น SQL Command แล้วใส่ sql Command ลงไป แล้วกดปุ่ม OK

SQL Command

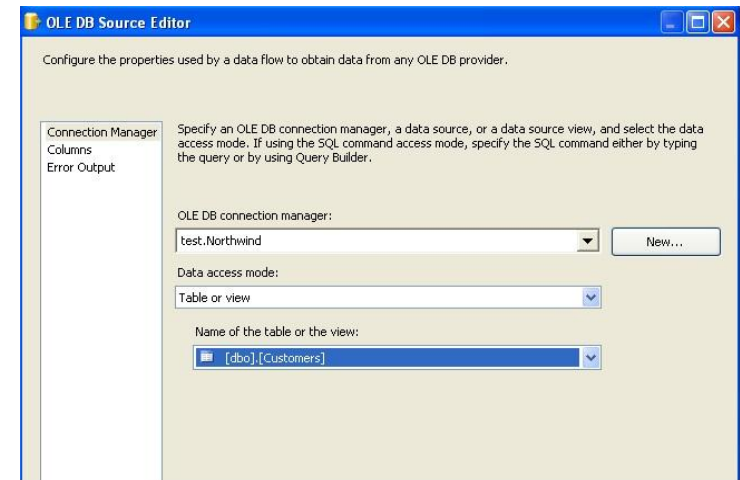
```
SELECT Products.ProductID, Products.ProductName,
Suppliers.companyName, Categories.CategoryName, Products.UnitPrice

FROM Products INNER JOIN Suppliers
ON Products.SupplierID = Suppliers.SupplierID
INNER JOIN Categories
ON Products.CategoryID = Categories.CategoryID
```

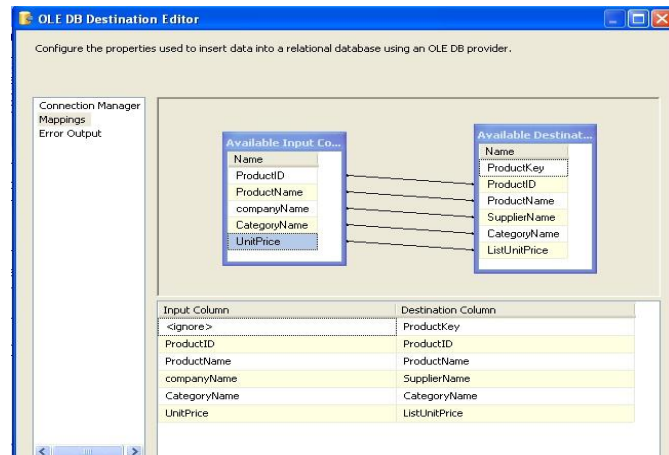
20. สร้าง OLTP Destination ชื่อ DW_Product พร้อมทั้งสร้าง Link ระหว่าง OLTP_Product ไปยัง DW_Product โดยทำเหมือนขั้นตอนในข้อ 11-14 โดยกำหนดค่า Connection Manager ดังรูปที่ 20 และ Mappings ของ DW_Product ดังรูปที่ 21 แล้วกดปุ่ม OK



รูปที่ 20 การกำหนดค่า Connection Manager ของ DW_Product



รูปที่ 22 การกำหนดค่า Connection Manager ของ OLTP_Customer

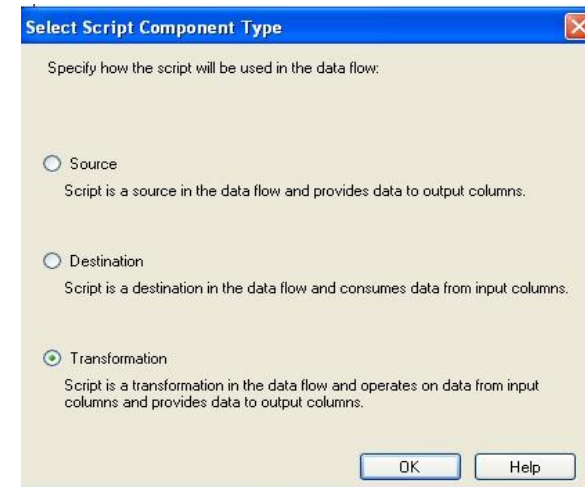


รูปที่ 21 ค่า Mappings จาก Source (OLTP_Product) ไปยัง Destination (DW_Product)

2.1.4 สร้างการดึงข้อมูลเข้าสู่คลังข้อมูลในตาราง Customer_Dim

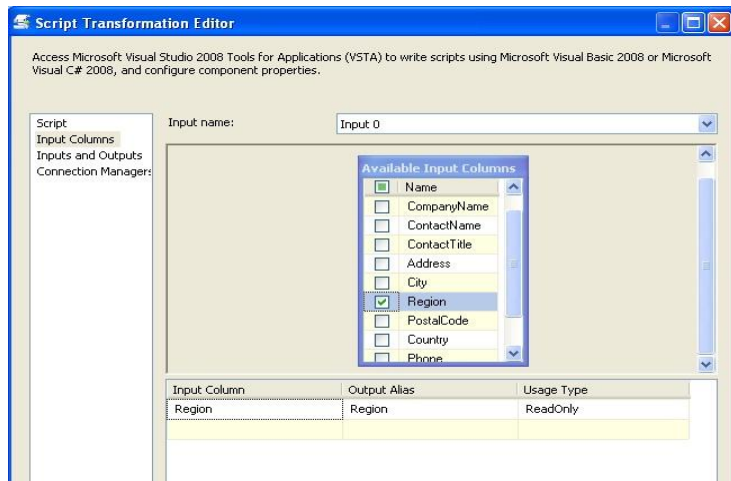
21. สร้าง OLTP Datasource ชื่อ OLTP_Customer โดยทำเหมือนขั้นตอน 11 พร้อมกำหนดค่า Connection Manager ของ OLTP_Customer ดังรูปที่ 22 แล้วกดปุ่ม OK

22. สร้าง Script Component แล้วเลือกตามรูปที่ 25 พร้อมทั้งสร้าง Link ระหว่าง OLTP_Customer ไปยัง Script Component



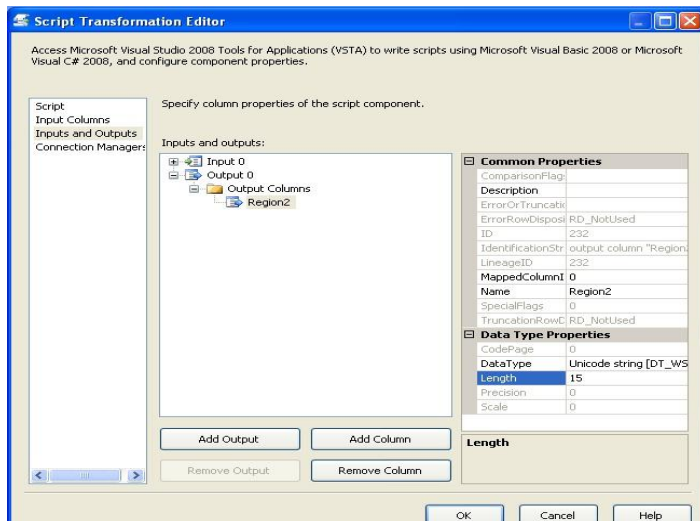
รูปที่ 23 Script Component Type

23. กด Double-clicks ที่ Script Component แล้วเลือกตามรูปที่ 24



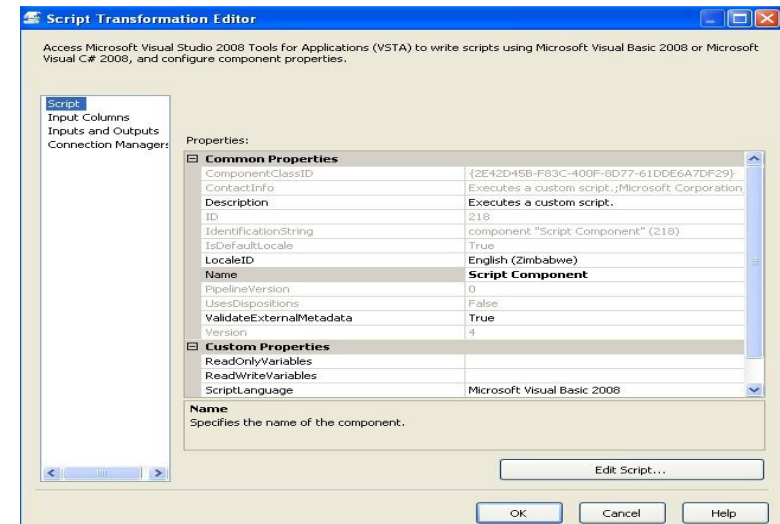
รูปที่ 24 Input Column ใน Script Component

24. เปลี่ยนไปที่ Input and Output แล้วเลือกไปที่ Output กดปุ่ม Add Column แล้วตั้งชื่อว่า Region2 แล้วเลือก Properties ตามรูปที่ 25



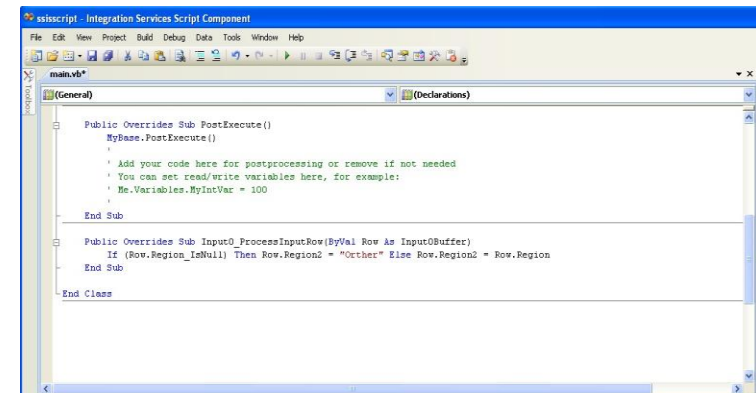
รูปที่ 25 Output Column ใน Script Component

25. เปลี่ยนไปที่ Script แล้วกดปุ่ม Design Script ดังรูป 26



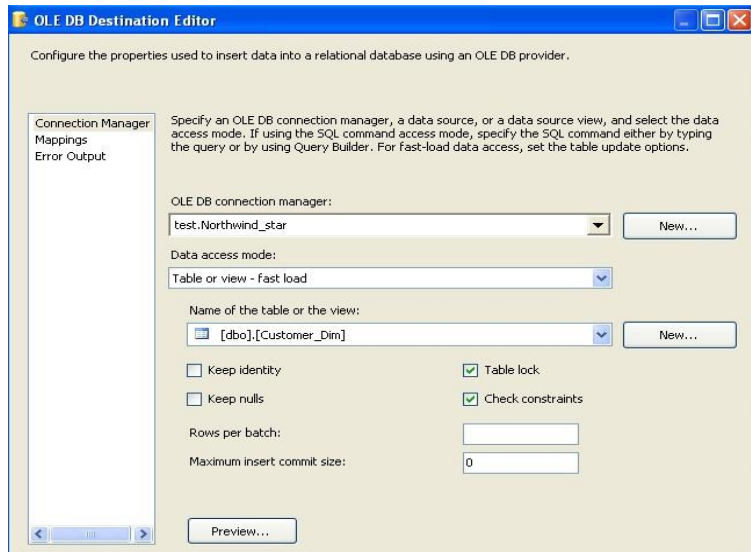
รูปที่ 26 Design Script ใน Script Component

26. เขียน Script ตามรูป 29



รูปที่ 27 Script

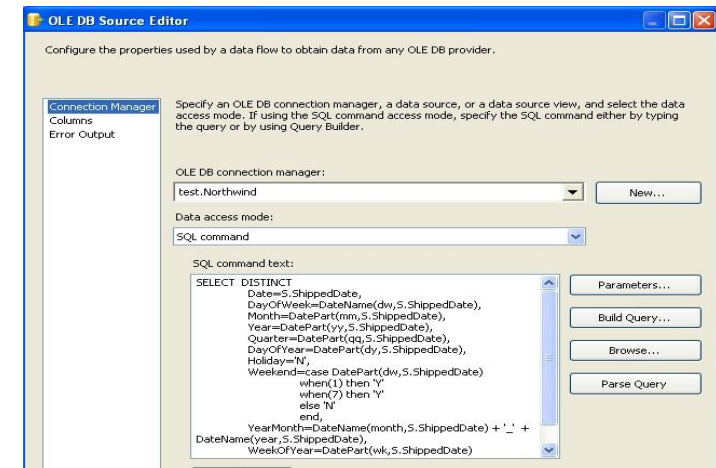
27. สร้าง Link ระหว่าง ระหว่าง Script Component ไปยัง DW_Customer โดยกำหนดค่า Connection Manager ดังรูปที่ 28 และ Mappings ของ DW_Customer ดังรูปที่ 29 แล้วกดปุ่ม OK



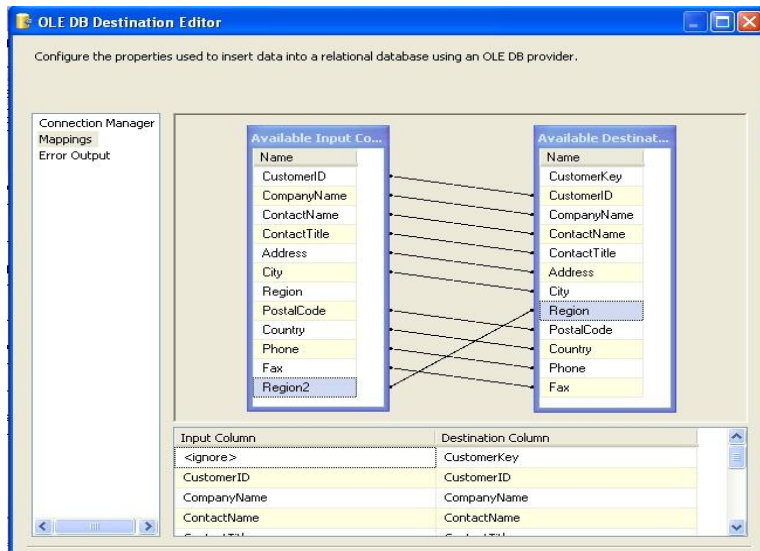
รูปที่ 28 การกำหนดค่า Connection Manager ของ DW_Customer

2.1.5 สร้างการดึงข้อมูลเข้าสู่คลังข้อมูลในตาราง Time_Dim

28. สร้าง OLTP Datasource ชื่อ OLTP_Time โดยทำเหมือนขั้นตอน 12 พร้อมกำหนดค่า Connection Manager ของ OLTP_Time ดังรูปที่ 30 แล้วกดปุ่ม OK



รูปที่ 30 การกำหนดค่า Connection Manager ของ OLTP_Time



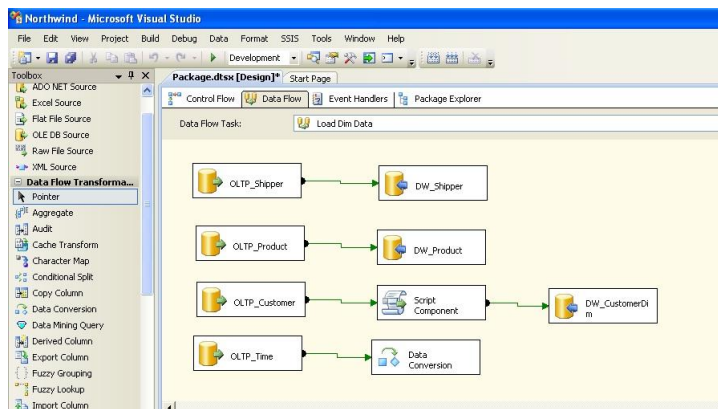
รูปที่ 29 ค่า Mappings จาก Source (OLTP_Customer) ไปยัง Destination (DW_Customer)

```

SELECT DISTINCT
Date=S.ShippedDate,
DayOfWeek=DateName(dw,S.ShippedDate),
Month=DatePart(mm,S.ShippedDate),
Year=DatePart(yy,S.ShippedDate),
Quarter=DatePart(qq,S.ShippedDate),
DayOfYear=DatePart(dy,S.ShippedDate),
Holiday='N',
Weekend=case DatePart(dw,S.ShippedDate)
when(1) then 'Y'
when(7) then 'Y'
else 'N'
end,
YearMonth = DateName(month,S.ShippedDate)+'_' + DateName(year,S.ShippedDate),
WeekOfYear=DatePart(wk,S.ShippedDate)
FROM Orders S
WHERE S.ShippedDate IS NOT NULL

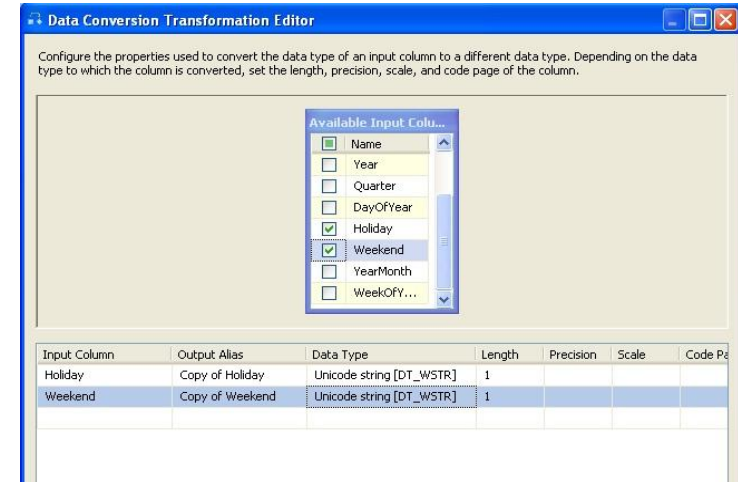
```

29. กด Double-clicks ที่ Data Conversion แล้วสร้าง Link จาก OLTP_Time ไปยัง Data Conversion
 ดังรูปที่ 31



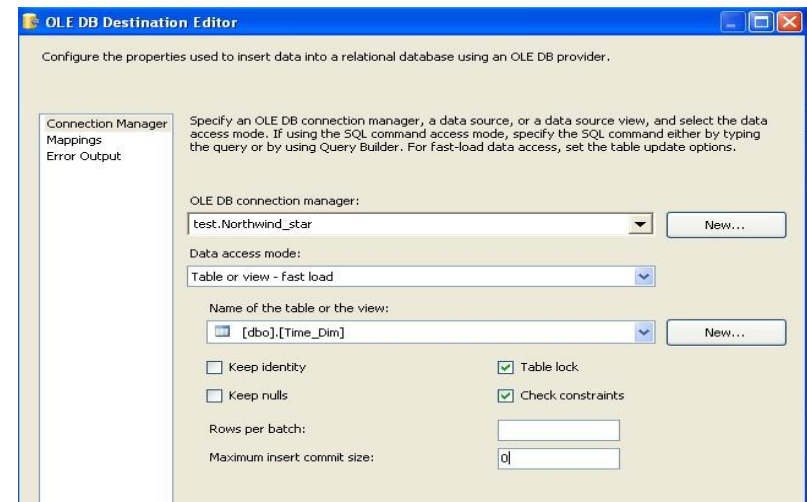
รูปที่ 31 Data Conversion

30. กด Double-clicks ที่ Data Conversion แล้วเลือกตามรูปที่ 32 (อย่าลืมเปลี่ยน Data Type)

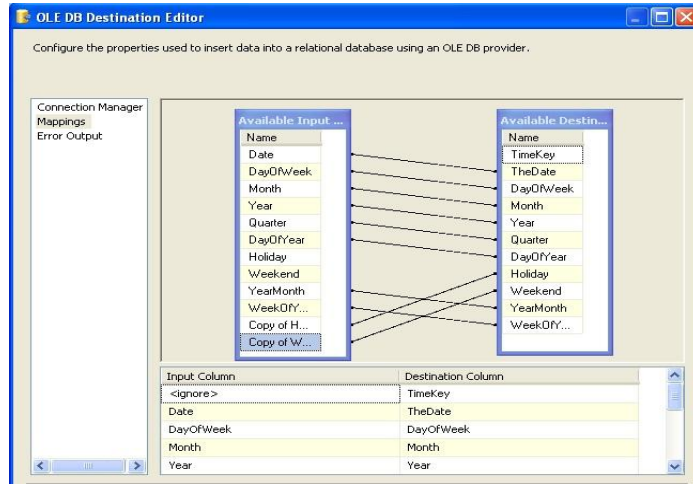


รูปที่ 32 Data Conversion

31. สร้าง OLTP Destination ชื่อ DW_Time พร้อมทั้งสร้าง Link ระหว่าง Data Conversion ไปยัง DW_Time โดยกำหนดค่า Connection Manager ดังรูปที่ 33 และ Mappings ของ DW_Time ดังรูปที่ 34 แล้วกดปุ่ม OK



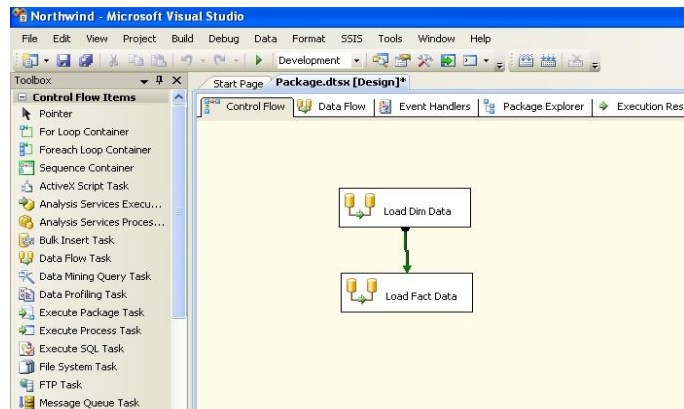
รูปที่ 33 การกำหนดค่า Connection Manager ของ DW_Time



รูปที่ 34 ค่า Mappings จาก Source (OLTP_Time) ไปยัง Destination (DW_Time)

2.2 สร้าง Data Flow สำหรับดึงข้อมูลเข้าสู่ Fact Table

32. สร้าง Data Flow ชื่อ Load Fact Data โดยทำเหมือนกับขั้นตอนที่ 10 หลังจากนั้นให้คลิกที่ Data Flow "Load Dim Data" แล้วลากลูกศรสีเขียวมายัง Data Flow "Load Fact Data" ซึ่งหมายความว่าต้องทำงานใน Data Flow "Load Dim Data" ให้เสร็จก่อนจึงจะทำงานในส่วนของ Data Flow "Load Dim Data" ดังรูปที่ 35

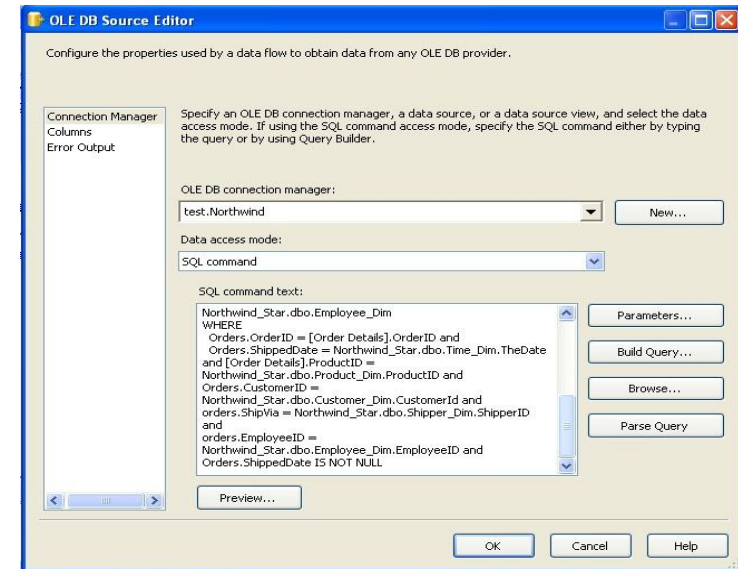


รูปที่ 35 การสร้าง Data Flow "Load Fact Data"

33. Double-clicks ที่ Data Flow "Load Fact Data" จะมาสู่ Tab Data Flow

34. สร้าง OLTP Destination ชื่อ DW_Order โดยทำเหมือนในขั้นตอน 11 พร้อมกับกำหนดค่า

Connection Manager ของ DW_Order ดังรูปที่ 36 แล้วกดปุ่ม OK



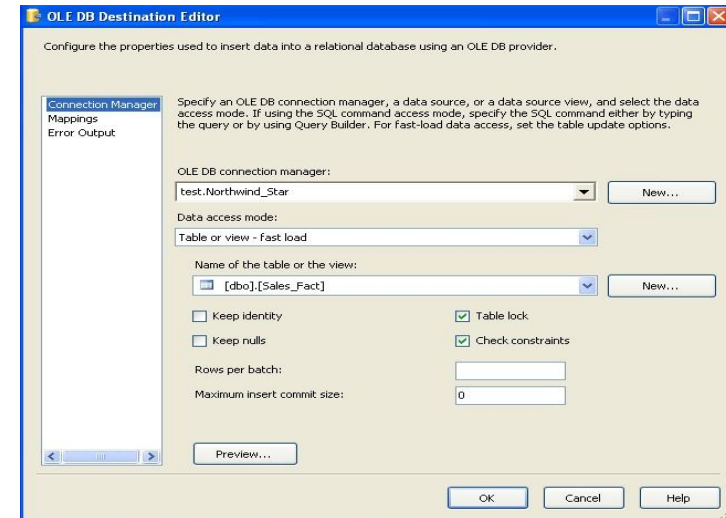
รูปที่ 36 การกำหนดค่า Connection Manager ของ OLTP_Purchase_Data

```

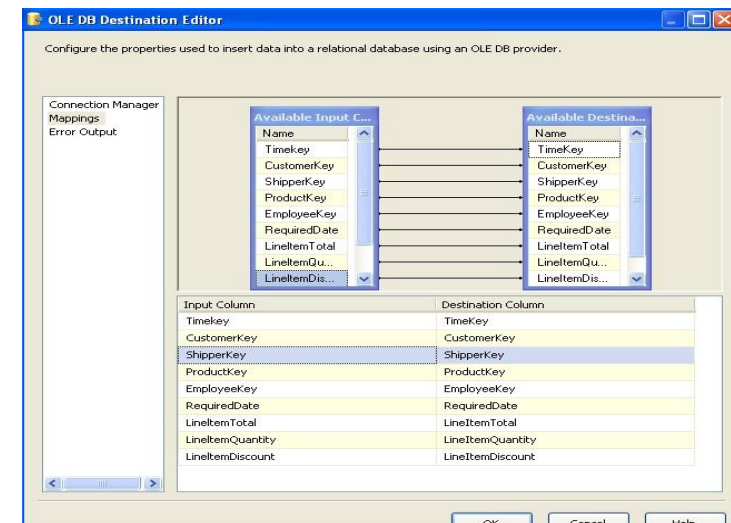
SELECT Northwind_Star.dbo.Time_Dim.TimeKey,
Northwind_Star.dbo.Customer_Dim.CustomerKey,
Northwind_Star.dbo.Shipper_Dim.ShipperKey,
Northwind_Star.dbo.Product_dim.ProductKey,
Northwind_Star.dbo.Employee_Dim.EmployeeKey,
Orders.RequiredDate,
LineItemTotal = [Order details].unitprice*[Order details].quantity,
LineItemQuantity = [Order details].quantity,
LineItemDiscount = [Order details].discount*[Order details].unitprice
FROM Orders,[Order Details],Northwind_Star.dbo.Time_Dim,
Northwind_Star.dbo.Product_Dim, Northwind_Star.dbo.Customer_Dim,
Northwind_Star.dbo.Shipper_Dim,
Northwind_Star.dbo.Employee_Dim
WHERE
Orders.OrderId = [Order Details].OrderID and
Orders.ShippedDate = Northwind_Star.dbo.Time_Dim.TheDate and
[Order Details].ProductId = Northwind_Star.dbo.Product_Dim.ProductID and

Orders.CustomerID =Northwind_Star.dbo.Customer_Dim.CustoMerID and
Orders.ShipVia = Northwind_Star.dbo.Shipper_Dim.ShipperId and
Orders.EmployeeId = Northwind_Star.dbo.Employee_Dim.EmployeeId and
Orders.ShippedDate IS NOT NULL

```



รูปที่ 37 การกำหนดค่า Connection Manager ของ DW_Fact

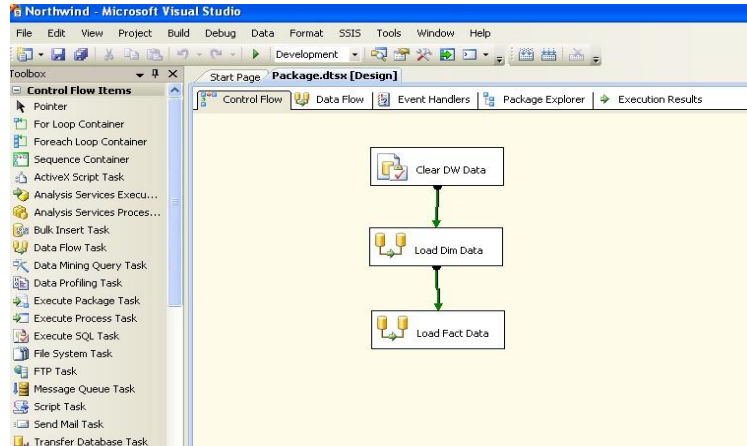


รูปที่ 38 ค่า Mappings จาก Source (OLTP_Order) ไปยัง Destination (DW_Fact)

35. สร้าง OLTP Destination ชื่อ DW_Fact พร้อมทั้งสร้าง Link ระหว่าง OLTP_Order ไปยัง DW_Fact โดยทำเหมือนขั้นตอนในข้อ 11-14 โดยกำหนดค่า Connection Manager ดังรูปที่ 37 และ Mappings ของ DW_Fact ดังรูปที่ 38 แล้วกดปุ่ม OK

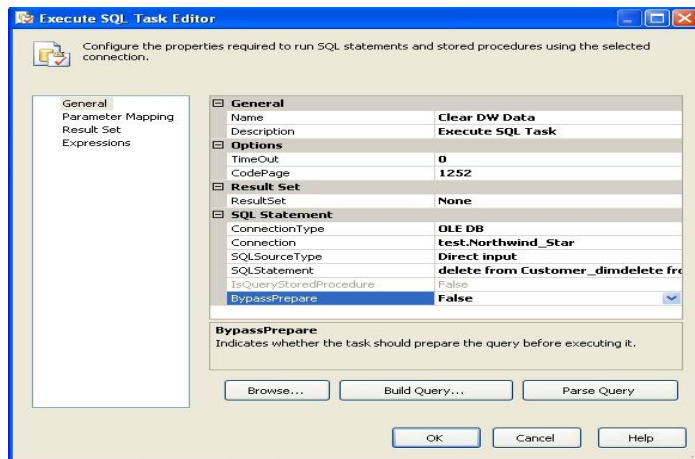
2.3 สร้าง Excute SQL Task สำหรับลบข้อมูลใน Dimension table และ Fact table เมื่อมีการ Load ข้อมูลใหม่

36. คลิกที่ Tab Control Flow หลังจากนั้นทำการคลิกลาก Excute SQL Task จากเมนูย่อย Control Flow Items ใน Tool Box ด้านซ้ายมือ แล้วสร้าง Link ไปยัง Data Flow "Load Dim Data" ดังรูปที่ 39 พร้อมกับเปลี่ยนชื่อเป็น "Clear DW Data"



รูปที่ 39 การสร้าง Excute SQL Task

37. Double-clicks ที่ "Clear DW Data" แล้วทำการกำหนดค่า Properties ในรูปที่ 40 ดังตารางข้างล่าง

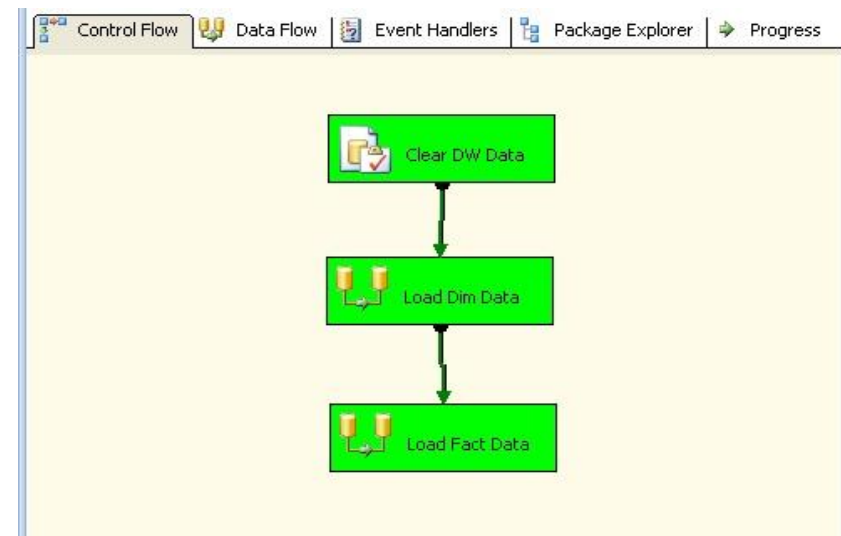


รูปที่ 40 กำหนดค่าต่างๆ ของ Excute SQL Task

General Page:	Connection	--	Servername.Northwind_Star
	SQL Statement	--	delete from Customer_dim
			delete from Time_dim
			delete from Product_dim
			delete from Employee_dim
			delete from Shipper_dim
			delete from Sale_Fact;

3. Execute SSIS Package เพื่อดึงข้อมูลจาก Northwind เข้าสู่ Northwind_Star (คลังข้อมูล)

38. คลิกที่ Tab Control Flow แล้วคลิกที่ปุ่ม Execute บนเมนูบาร์ เพื่อทำการรัน Project ก็เป็นอันเสร็จสิ้นการดึงข้อมูลดังรูป 41



รูปที่ 41 ดึงข้อมูลเสร็จสิ้น