

---

## 第8章 定义数据变换和移动

在本课程中，将定义“数据仓库中心”如何为数据仓库移动数据及将数据变换为其格式。首先，您将定义一个进程，该进程包含变换和移动进程中的一系列步骤。然后，将指定要为仓库进行变换的源表。最后，将定义使用以下两种不同变换方法的数据变换步骤：

- 将用程序来把数据装入到仓库数据库中。
- 将选择源数据并将表与 SQL 语句相连。

特别地，您将定义“教程市场”进程，该进程执行下列处理：

1. 将“人口统计”文件装入到仓库数据库中。
2. 从 GEOGRAPHIES 表中选择数据，并创建目标表。
3. 连接“人口统计”表和 GEOGRAPHIES 目标表中的数据。

最终结果是 LOOKUP\_MARKET 目标表。

---

### 定义进程

在此练习中，将为“教程市场”进程定义进程对象。

要定义进程对象：

1. 从“数据仓库中心”窗口，展开**主题区**树。
2. 展开 **TBC 教程**主题区，该主题区是在第23页的『第5章 定义主题区』中定义的。
3. 用鼠标右键单击**进程**文件夹，并单击**定义**。  
“定义进程”笔记本打开。
4. 在**名称**字段中，输入进程的名称：  
教程市场

该名称最长可为 80 个字符，且是区分大小写的。该名称的第一个字符必须是字母数字。在英文中，不能将 & 用作首个字符。此字段是必需的。

5. 在**管理员**字段中，输入您的姓名作为进程定义的联系人。
6. 在**说明**字段中，输入进程的说明：  
创建 LOOKUP\_MARKET 表的进程
7. 单击**安全性**标签。

8. 在**可用安全性组**列表中，单击**教程仓库组**，它是在第19页的『定义仓库组』中定义的。

把进程添加至安全性组将授权组中的用户（此情况中为您）打开对象，并将对象添加至该进程。

9. 单击 **>**。

“教程仓库组”就移至**已选择的安全性组**列表中。

10. 单击**确认**。

“定义进程”笔记本关闭。

---

## 打开进程

在此练习中，将打开进程，以便您可以图形方式来定义下一个练习的进程中的数据流。

要打开“教程市场”进程：

1. 用鼠标右键单击“教程市场”进程。
2. 单击**打开**。

---

## 将表添加至进程

要定义数据流，需要添加步骤变换的每个源及从变换中得到的目标表。

在“教程市场”进程中，您要将“人口统计”文件装入目标数据库，因此，需要为进程的步骤添加源文件和 **DEMOGRAPHICS\_TARGET** 目标表。“人口统计”源文件是“教程文件源”仓库源的一部分，它是在第25页的『第6章 定义仓库源』中定义的。**DEMOGRAPHICS\_TARGET** 目标表是“教程目标”仓库目标的一部分，它是在第35页的『第7章 定义仓库目标』中定义的。

要添加“人口统计”文件：

1. 单击**添加数据**图符：



在想要放置表的地方单击画布。“添加数据”窗口打开。

2. 在**可用的源表和目标表**列表中，展开**仓库源树**。

将显示在仓库中定义的仓库源的列表。

3. 展开“教程文件源”仓库源树。
4. 展开**文件树**。

在该树形结构中，您应可看到  
X:\sql11b\samples\db2samp1\dw\demographics.txt，其中，X 是安装了样本的驱动器。

5. 选择“人口统计”文件。
6. 单击 ➤ 以将“人口统计”文件添加至已选择的源表和目标表列表。

要添加 DEMOGRAPHICS\_TARGET 目标表：

1. 在可用的源表和目标表列表中，展开仓库目标树。  
将显示在仓库中定义的仓库目标的列表。
2. 展开“教程目标”仓库目标树。
3. 展开表树。  
在该列表中应可看到 DEMOGRAPHICS\_TARGET 目标表。
4. 选择 DEMOGRAPHICS\_TARGET 目标表。
5. 单击 ➤ 以将 DEMOGRAPHICS\_TARGET 目标表添加至已选择的源表和目标表列表。

在此练习的下一部分，需要添加 GEOGRAPHIES 源表。当您定义从 GEOGRAPHIES 表选择数据的步骤时，可指定“数据仓库中心”自动生成目标表，因此不需要添加目标表。

要添加 GEOGRAPHIES 源表：

1. 从“添加数据”窗口，选择 GEOGRAPHIES 表。
2. 单击 ➤ 以将 GEOGRAPHIES 表添加至已选择的源表和目标表列表。
3. 单击**确认**以将所选择的文件和表添加至进程。  
所选择的文件和表显示在“进程模型”窗口中。

最后一个步骤将把“人口统计”表和“地理”表用作源，因此您不需要为该步骤指定源。可以指定“数据仓库中心”自动生成结果目标表 LOOKUP\_MARKET，因此不需要为该步骤指定目标表。

---

## 将步骤添加至进程

现在，需要添加用来定义如何将源数据变换为目标数据的步骤。必须定义三个步骤：

### 装入人口统计数据

这是一个 DB2 程序步骤，它将数据从“人口统计”文件装入到“TBC 仓库”数据库的表中，该数据库是在第5页的『第2章 创建仓库数据库』中创建的。

## 选择地理数据

这是一个 SQL 步骤，它从“地理”源表中选择列。

## 连接市场数据

这是一个 SQL 步骤，它连接“地理”表和“人口统计”表，并将所连接的表写入“教程仓库”数据库中。

## 定义“装入人口统计”步骤

在此练习中，将定义“装入人口统计”步骤：

1. 从调色板（窗口左边的工具栏）中，单击 **DB2 程序**图符，该图符是自底部数上来的第二个图符：



每个程序图符都表示一个程序组，它是相类似程序的分组。

2. 单击 **DB2 UDB → 装入**。

程序组中的程序是从程序图符级联的。

3. 单击您想将步骤放在画布（窗口右边的空白区域）上的那个位置。

表示该步骤的图符即被添加至窗口。

4. 用鼠标右键单击该步骤。

5. 单击**特性**。

“步骤”笔记本打开。

6. 在**名称**字段中，输入步骤的名称：

装入人口统计数据

7. 在**管理员**字段中，输入您的姓名作为此步骤的联系人。

8. 在**说明**字段中，输入步骤的说明：

将人口统计数据装入仓库中。

9. 单击**确认**。

“步骤特性”笔记本关闭。

10. 单击**任务流**图符：



11. 单击**数据链接**图符：



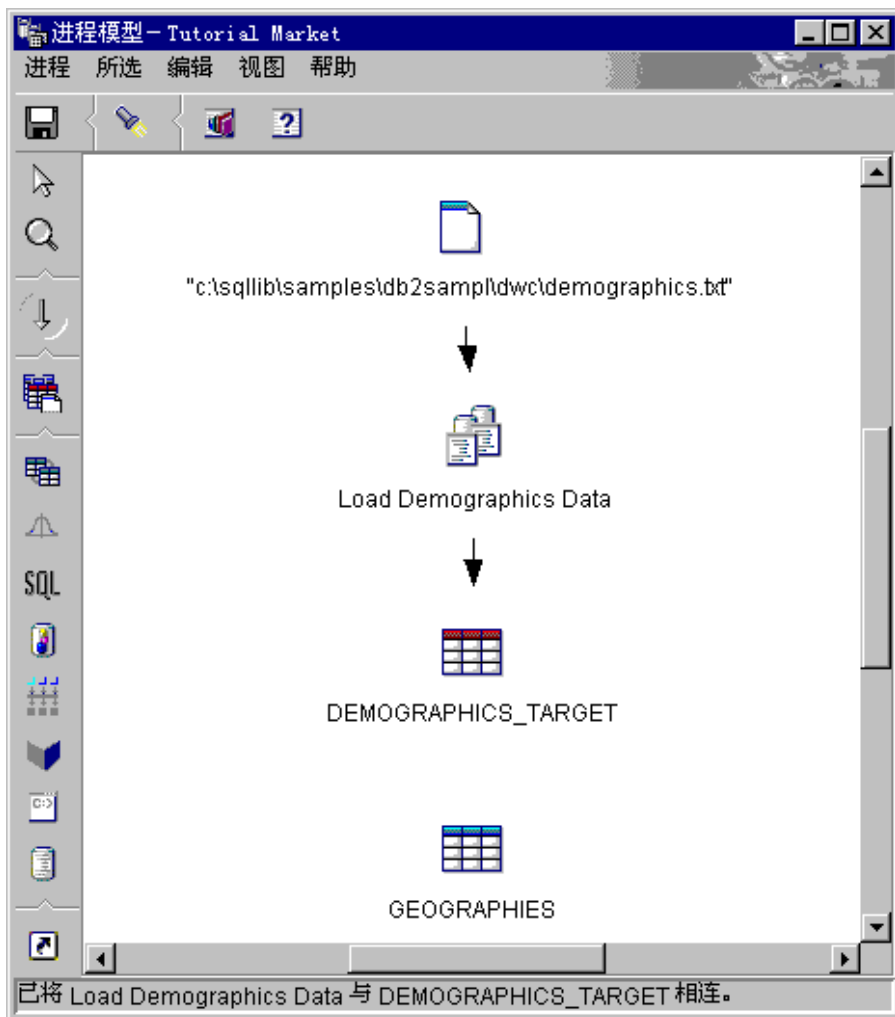
将使用“数据链接”图符来定义从源文件至目标表（通过在某一步骤中进行变换）的数据流。

12. 单击“人口统计”源文件的中部，并将鼠标拖至“装入人口统计数据”步骤。

“数据仓库中心”就会在源文件和步骤之间划一条线。

这指示“人口统计”源文件包含该步骤的源数据。

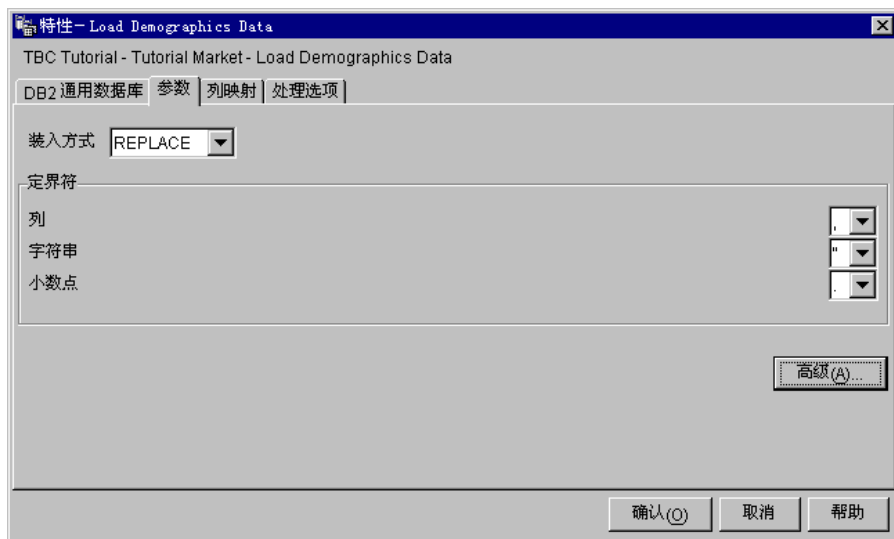
13. 单击“装入人口统计数据”步骤的中部，并将鼠标拖至“人口统计表”目标表。



这指示“人口统计表”目标表包含该步骤的目标数据。

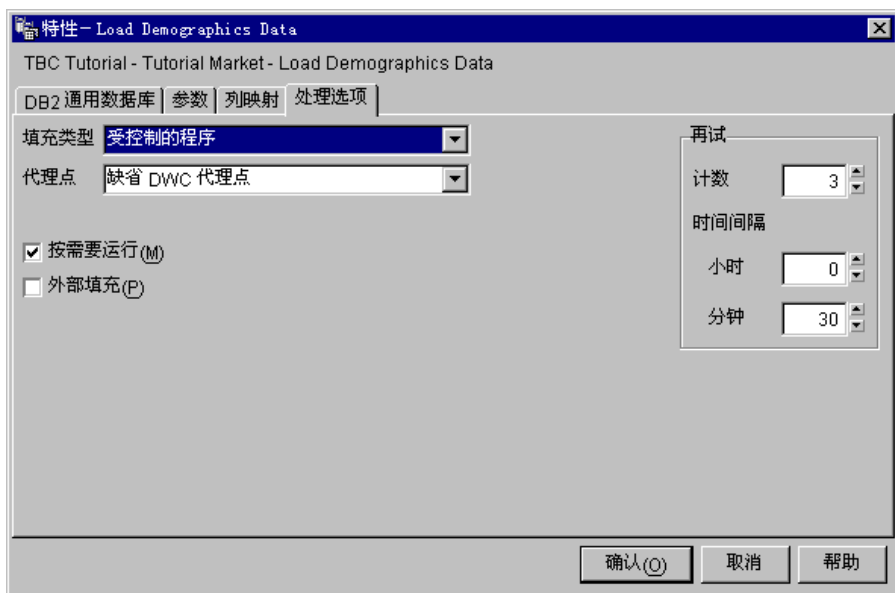
14. 用鼠标右键单击该步骤。

15. 单击**特性**。  
“步骤”笔记本打开。
16. 单击**参数**标签。



17. 从**装入方式**列表中，单击**替换**。  
接受该页上的其余缺省值。  
因为 Load 程序不使用列映射来生成目标表，所以您可以跳过“列映射”页。  
Load 程序根据源文件生成目标表。

18. 接受“处理选项”页上的缺省值。



有关此页上的值的详情，参见联机帮助中的『将数据装入表中』。

19. 单击**确认**。

“步骤”笔记本关闭。

## 定义“选择地理”步骤

在此练习中，将定义“选择地理”步骤：

1. 从调色板，单击 **SQL** 图符：

**SQL**

2. 单击您想将步骤放在画布上的那个位置。  
表示该步骤的图符即被添加至窗口。
3. 用鼠标右键单击该步骤。
4. 单击**特性**。  
“步骤”笔记本打开。
5. 在**名称**字段中，输入步骤的名称：  
选择地理数据
6. 在**管理员**字段中，输入您的姓名作为该步骤的联系人姓名。
7. 在**说明**字段中，输入步骤的说明：

从仓库源中选择“地理”数据

8. 单击**确认**。

“步骤”笔记本关闭。

9. 单击**任务流**图符:



10. 单击**数据链接**图符:



11. 单击“地理”源表的中部，并将鼠标拖至“选择地理数据”步骤的中部。

“数据仓库中心”即划一条线，它指示“地理”源表包含该步骤的源数据。

因为您将指定“数据仓库中心”要创建目标表，所以不需要将目标表与该步骤链接。

12. 用鼠标右键单击“选择地理数据”步骤。

13. 单击**特性**。

“步骤”笔记本打开。

14. 单击 **SQL 语句** 标签。

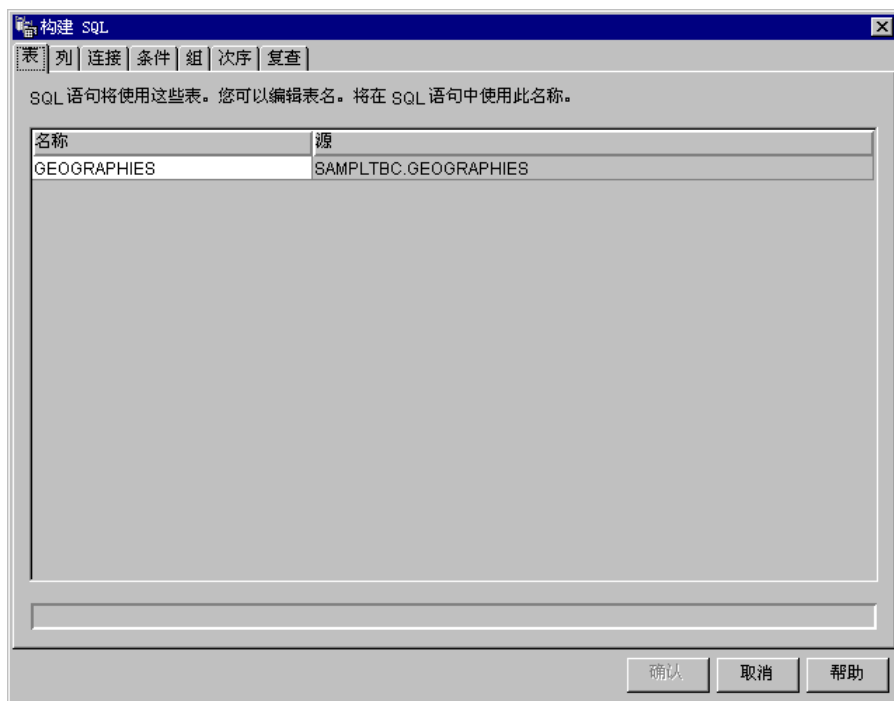
15. 单击**构建 SQL** 以通过使用“SQL 辅助程序”向导来构建 SQL，将会根据您在该向导中所作的选择来生成 SQL。

“SQL 辅助程序”打开。

16. 单击**表**标签。

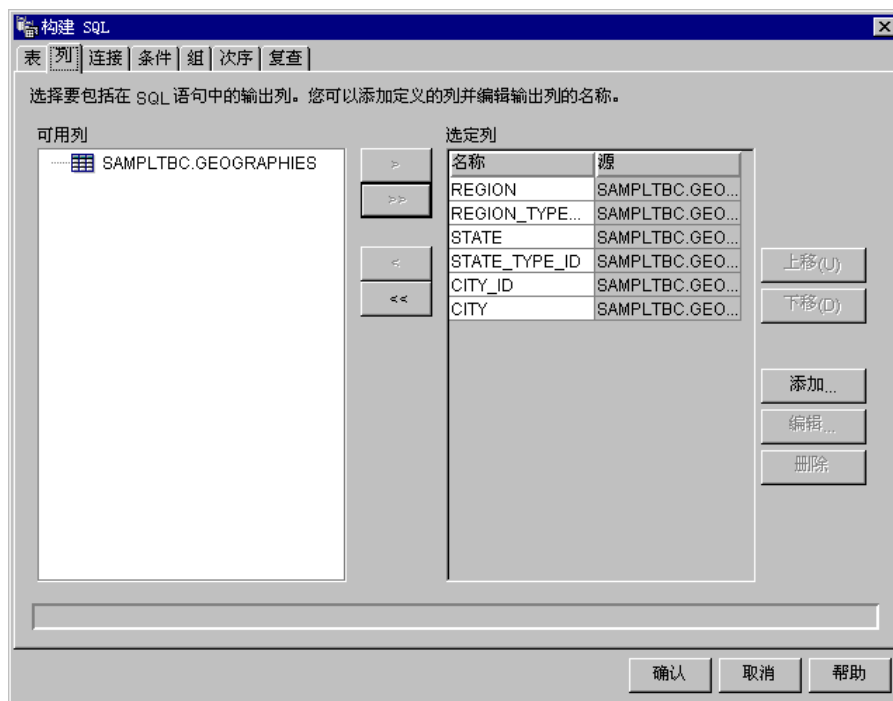


17. 验证是否列示了 GEOGRAPHIES。



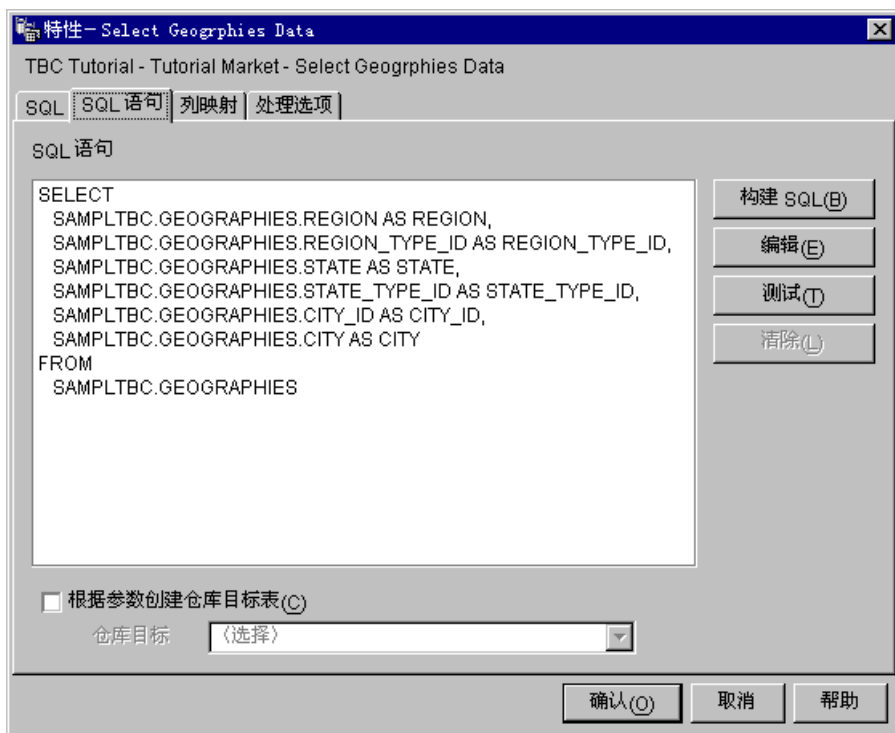
18. 单击列标签。

19. 单击 **>>**，以将“地理”表中的所有列添加至输出列表。



20. 单击**复查**标签以查看您刚才构建的 SQL 语句。
21. 单击**确认**。

“SQL 辅助程序” 关闭。您刚才构建的 SQL 语句显示在 “SQL 语句” 页上。



22. 单击**测试**以测试刚才生成的 SQL。

“数据仓库中心” 返回 SELECT 语句的样本结果。这些结果应该与您浏览 “地理” 源表的样本数据时在第11页的『第3章 浏览源数据』中获得的结果相同。

23. 单击**关闭**来关闭该窗口。

24. 选择**根据参数创建仓库目标表**校验框。

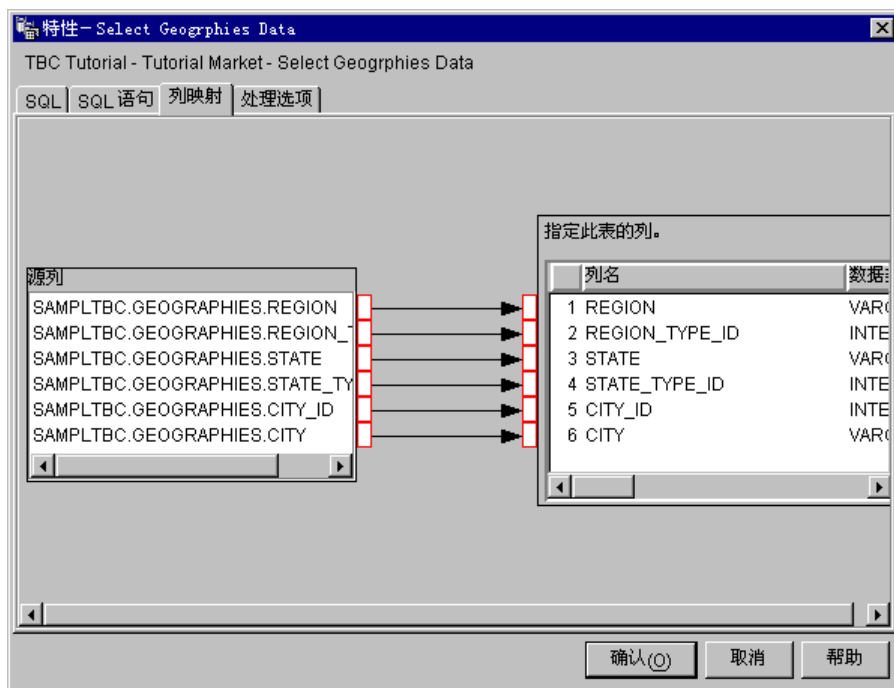
选择此校验框将指定 “数据仓库中心” 会根据在 “列映射” 页上指定的值来创建目标表。

25. 从**仓库目标**列表，单击**教程目标**。

仓库目标就是将在其中创建目标表的数据库或文件系统。

26. 单击**列映射**标签。

27. 验证源列是否映射至正确的目标列。



28. 单击**处理选项**标签。

29. 在**填充类型**列表中，单击**替换**。

30. 验证是否选择了**按需运行**校验框。

接受此页上的其余缺省值。有关此页上的值的详情，参见联机帮助。

31. 单击**确认**。

“步骤”笔记本关闭。“数据仓库中心”生成一个目标表“Select\_Geog\_T”。因为该目标表名是混合大小写的，所以括在引号中。

32. 更改目标表的名称：

a. 用鼠标右键单击“Select\_Geog\_T”目标表。

b. 单击**特性**。

“表”笔记本打开。

c. 在**表模式**字段中，输入 IWH。

d. 在**表名字**段中，为该表输入新名称：

GEOGRAPHIES\_TARGET

e. 在**说明**字段中，输入表的说明：

为“选择地理数据”步骤自动生成的目标表。

- f. 在“商业名”字段中，输入表的描述名：  
地理目标
- g. 清除 **OLAP 模式** 的一部分校验框。
- h. 单击列标签以查看列定义。
- i. 单击**确认**。“表”笔记本关闭。

## 定义“连接市场数据”步骤

要定义“连接市场数据”步骤：

- 1. 从调色板，单击 **SQL** 图符。
- 2. 单击您想将步骤放在画布上的那个位置。  
表示该步骤的图符即被添加至窗口。
- 3. 用鼠标右键单击该步骤。
- 4. 单击**特性**。  
“步骤”笔记本打开。
- 5. 在**名称**字段中，输入步骤的名称：  
连接市场数据
- 6. 在**管理员**字段中，输入您的姓名作为该步骤的联系人。
- 7. 在**说明**字段中，输入步骤的说明：  
将“地理”表与“人口统计”表相连
- 8. 单击**确认**。  
“步骤”笔记本关闭。
- 9. 单击**任务流**图符：

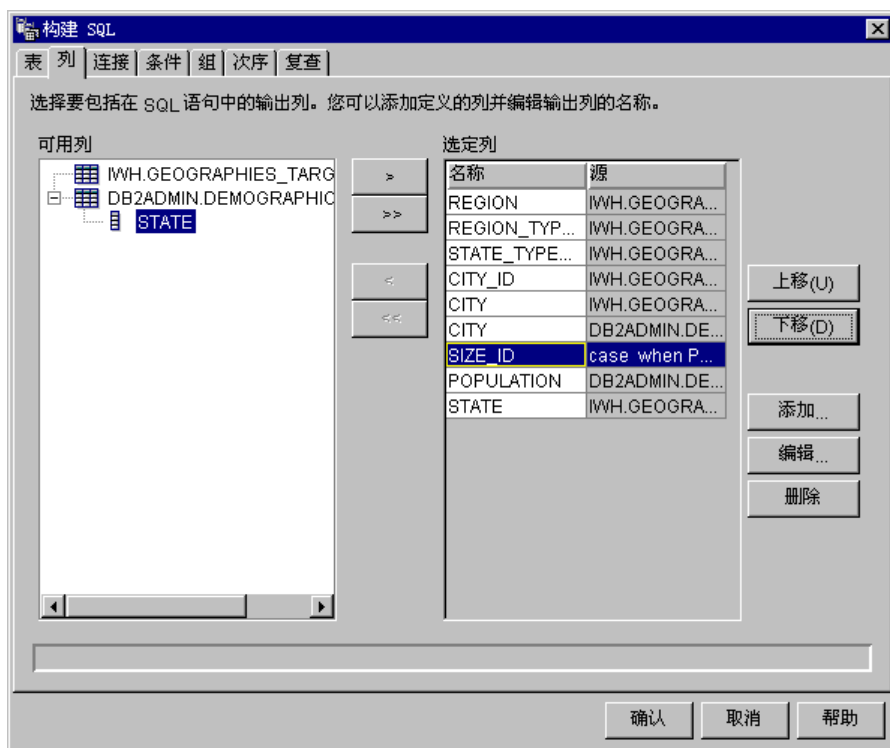


- 10. 单击**数据链接**图符：



- 11. 单击“地理目标”表的中部，并将鼠标拖至“连接市场数据”步骤。对“人口统计目标”表和“连接市场数据”步骤重复此步骤。  
“数据仓库中心”将划一些线，它们指示“地理目标”表和“人口统计目标”表包含该步骤的源数据。  
因为您将指定“数据仓库中心”要创建目标表，所以不需要将目标表与该步骤链接。

12. 用鼠标右键单击“连接市场数据”步骤。
13. 单击**特性**。  
“步骤”笔记本打开。
14. 单击 **SQL 语句** 标签。
15. 单击**构建 SQL** 以让“数据仓库中心”生成 SQL。（否则，您可以输入您自己的 SQL。）  
“SQL 辅助程序”打开。
16. 单击**表** 标签。
17. 验证是否选择了 DEMOGRAPHICS\_TARGET 和 GEOGRAPHIES\_TARGET 表。
18. 单击**列** 标签。

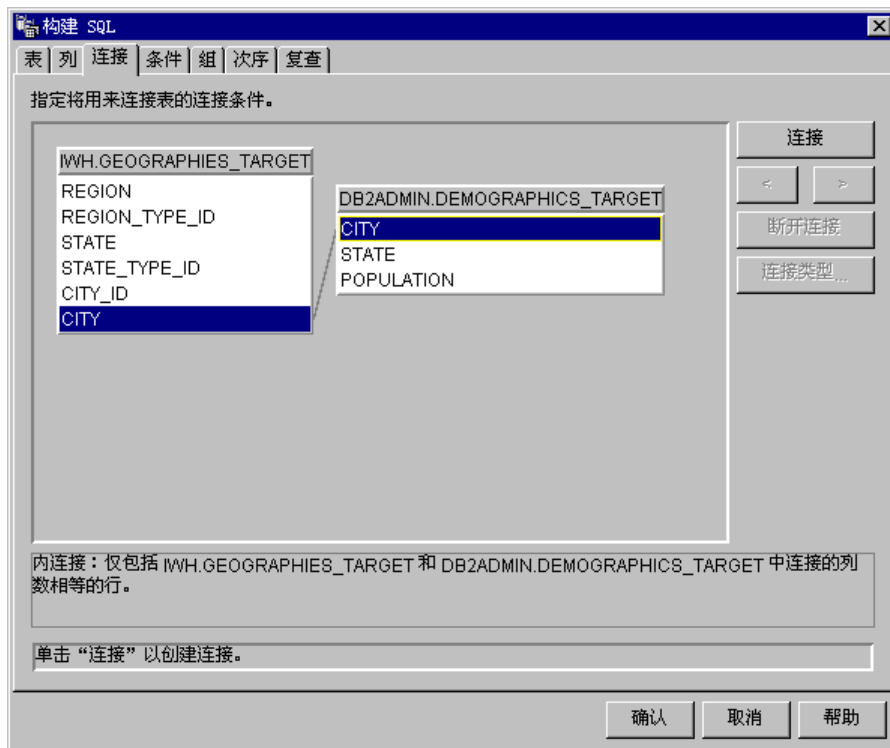


19. 单击 **>>**，以将所有列从“地理”表和“人口统计”表添加至输出列列表。
20. 从输出列列表，单击 **DEMOGRAPHICS\_TARGET.STATE**。
21. 单击 **<**。  
DEMOGRAPHICS\_TARGET.STATE 列即移至可用列列表。

22. 单击 **DEMOGRAPHICS\_TARGET.CITY**。
23. 单击 **<**。
- DEMOGRAPHICS\_TARGET.CITY 列即移至可用列列表。
- 24.
- 不需要 DEMOGRAPHICS\_TARGET.STATE 和 DEMOGRAPHICS\_TARGET.CITY 列，原因是已经在 GEOGRAPHIES\_TARGET 表中定义了它。将通过使用两个表中的 CITY 列（它是唯一的）来连接这两个表。
25. 单击**添加**。
- “表达式构建程序”窗口打开。
26. 在表达式字段中，输入以下 CASE 语句：
- ```
case
when POPULATION < 9000000 then 1
when POPULATION < 15000000 then 2
Else 3
end
```



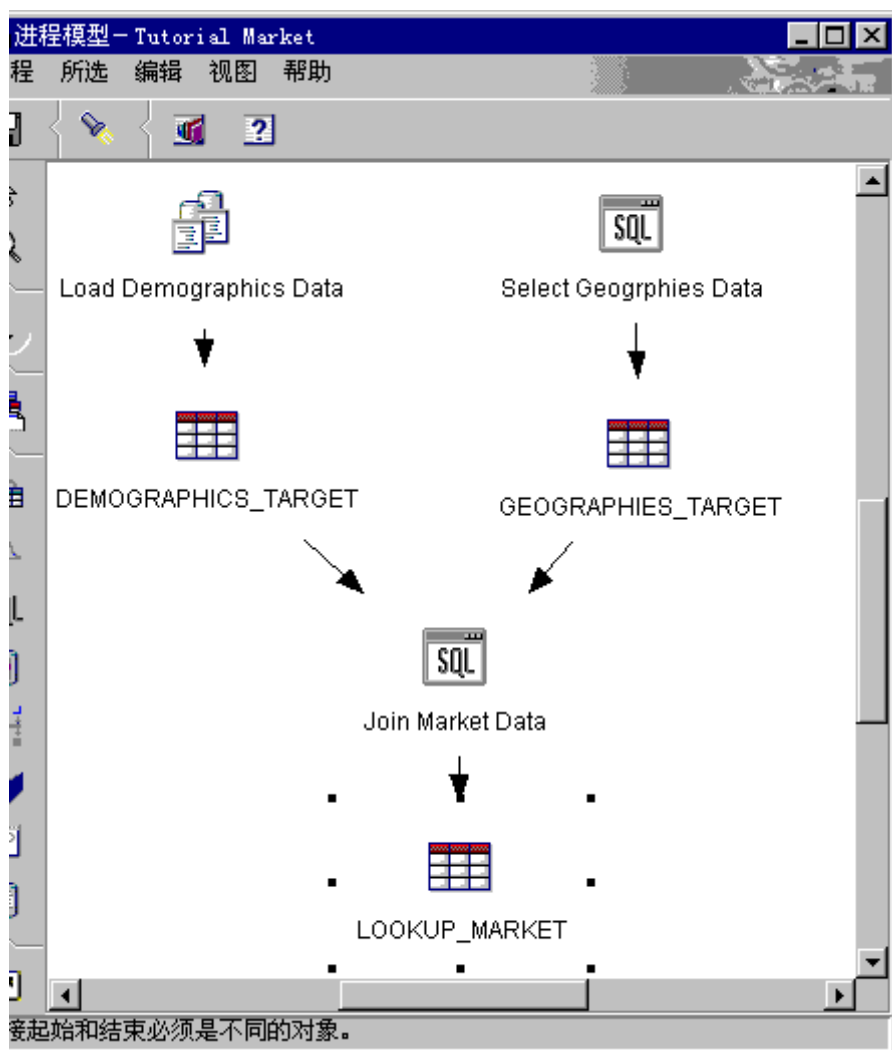
27. 单击**确认**。  
新列即被添加至**已选择的列**列表。
28. 单击新列的**名称**字段，并输入该列的名称：  
SIZE\_ID
29. 按 Enter 键。
30. 单击**上移**以将 SIZE\_ID 列移到 POPULATION 列的上面。
31. 单击**连接**标签。
32. 在 GEOGRAPHIES 表中，选择 CITY 列。
33. 在 GEOGRAPHIES 表中，选择 CITY 列。
34. 单击**连接**。  
“SQL 辅助程序”将在两个 CITY 列之间划一条线，指示这两个表是在该列上进行连接的。



35. 单击**复查**标签以查看您刚才构建的 SQL 语句。
36. 单击**确认**。  
“SQL 辅助程序”关闭。



37. 选择**根据参数创建仓库目标表**校验框。  
选择此校验框即指定“数据仓库中心”将根据在“SQL 语句”页和“列映射”页上指定的值来创建目标表。
38. 从**仓库目标列表**，单击**教程目标**。
39. 单击**列映射**标签。
40. 验证源列是否映射至正确的目标列。
41. 单击**处理选项**标签。
42. 在**填充类型**列表中，单击**替换**。
43. 选择**按需运行**校验框来按需要运行步骤。  
接受此页上的其余缺省值。有关此页上的值的详情，参见联机帮助。
44. 单击**确认**。  
“步骤”笔记本关闭。
45. 更改目标表的名称：
  - a. 用鼠标右键单击“Join\_Market\_T”目标表。
  - b. 单击**特性**。  
“表”笔记本打开。
  - c. 在**表模式**字段中，输入 IWH。
  - d. 在**表名字段**中，为该表输入新名称：  
LOOKUP\_MARKET
  - e. 在**说明**字段中，输入表的说明：  
市场维数据
  - f. 验证是否选择了**OLAP 模式的一部分**校验框和**维度表**单选按钮。  
此表是其中一个将在第99页的『第16章 从“数据仓库中心”内创建星型模式』中包括到星型模式中的维度表。
  - g. 单击**列**标签以查看列定义。
  - h. 清除 CITY ID 列的**可为空**校验框，因为在第71页的『第11章 定义目标表的关键字』中，将把此列定义成表的主关键字
  - i. 单击**确认**。 “表”笔记本关闭。



## 您刚才做了什么

在“教程市场”进程中定义了 LOOKUP\_MARKET 维度表，它包含三个步骤：

- 装入人口统计数据
- 选择地理数据
- 连接市场数据

每个步骤的源和目标分别是：

| 步骤       | 源                    | 目标         |
|----------|----------------------|------------|
| 装入人口统计数据 | “人口统计” 文件            | “人口统计目标” 表 |
| 选择地理数据   | “地理” 表               | “地理目标” 表   |
| 连接市场数据   | “人口统计目标” 表和 “地理目标” 表 | “查找市场” 表   |

对于此教程，当您定义每个步骤的特性时，就添加了每个步骤的数据链接。可以完成此任务的另一种方法是：同时添加进程中的所有步骤，将这些步骤与它们的源和目标链接，然后定义每个步骤的特性。“数据仓库中心” 对您在“步骤” 笔记本中可更改的步骤指定缺省名。

### 定义星型模式的其余步骤（可选）

您在前一节中创建的表 LOOKUP\_MARKET 是第v页的『教程中的商业问题』的示例中的其中一个维度表。本节叙述为星型模式构建其余维度表和事实表的步骤。

本节是可选的，但是，若不完成本节中的步骤，则将不能进行下列课程：

- 第71页的『第11章 定义目标表的关键字』
- 第83页的『第14章 编目仓库中的数据以供最终用户使用』
- 第87页的『第15章 使用商业元数据』
- 第99页的『第16章 从“数据仓库中心” 内创建星型模式』

若想跳过本节，则转至第63页的『第9章 测试仓库步骤』。

当定义每个表时，必须为该表定义新进程。然而，您将复制样本中定义的步骤，而不是为进程定义自己的步骤。步骤的定义在您正在使用的“数据仓库中心” 中。当复制步骤时，“数据仓库中心” 将复制该步骤所使用的源，并生成目标表。

要定义“产品” 维度表：

1. 遵循第41页的『定义进程』中的步骤来定义新进程。将进程命名为教程产品。
2. 从“数据仓库中心” 主窗口展开树，直到看到“样本产品” 进程为止。
3. 用鼠标右键单击“样本产品” 进程。
4. 单击**打开**。  
“进程模型” 窗口打开。
5. 用鼠标右键单击“选择产品” 步骤。

6. 单击**复制**。  
“复制步骤”窗口打开。
7. 在**名称**字段中，输入步骤副本的名称：  
教程选择产品
8. 在**新进程**字段中，选择或输入想要将该步骤复制至的进程的名称：  
教程产品
9. 验证是否选择了**复制目标表**校验框。
10. 在**仓库目标**字段中，从列表选择**教程目标**仓库目标。



11. 单击**确认**。  
步骤及其源被复制到“教程产品”进程中。“数据仓库中心”生成相应的目标表。
12. 遵循第42页的『打开进程』中的步骤来打开“教程产品”进程。
13. 验证进程是否包括下列对象：
  - PRODUCT 源表
  - “教程选择产品”步骤
  - TARGET\_PRODUCT 目标表
14. 将 TARGET\_PRODUCT 目标表重命名为 LOOKUP\_PRODUCT。

对其余维度表和事实表重复此过程。

| 维   | 教程过程  | 样本进程  | 样本步骤  | 教程步骤    | 仓库目标 | 源表                                | 目标表               | 新目标表名           |
|-----|-------|-------|-------|---------|------|-----------------------------------|-------------------|-----------------|
| 时间  | 教程时间  | 采样时间  | 选择时间  | 教程选择时间  | 教程目标 | TIME                              | TARGET_TIME       | LOOKUP_TIME     |
| 方案  | 教程方案  | 样本方案  | 选择方案  | 教程选择方案  | 教程目标 | SCENARIO                          | TARGET_SCENARIO   | LOOKUP_SCENARIO |
| 事实表 | 教程事实表 | 样本事实表 | 事实表连接 | 教程事实表连接 | 教程目标 | SALES, INVENTORY, 和 PRODUCT_COSTS | TARGET_FACT_TABLE | FACT_TABLE      |

您刚才做了什么

现在，您已经在“数据仓库中心”中定义了五个进程以及与它们相关联的源、目标和步骤：

- 教程市场
- 教程产品
- 教程方案
- 教程时间
- 教程事实表

