《数据仓库与数据挖掘》

实验报告册

20 18 - 20 19 学年 第 一 学期

班 级：

学 号：

姓 名：

授课教师： 王晓茹 实验教师： 王晓茹

实验学时： 2 实验组号： 1

计算机系

**实验二 设计并构造AdventureWorks数据仓库实例（2学时）**

**【实验要求】**

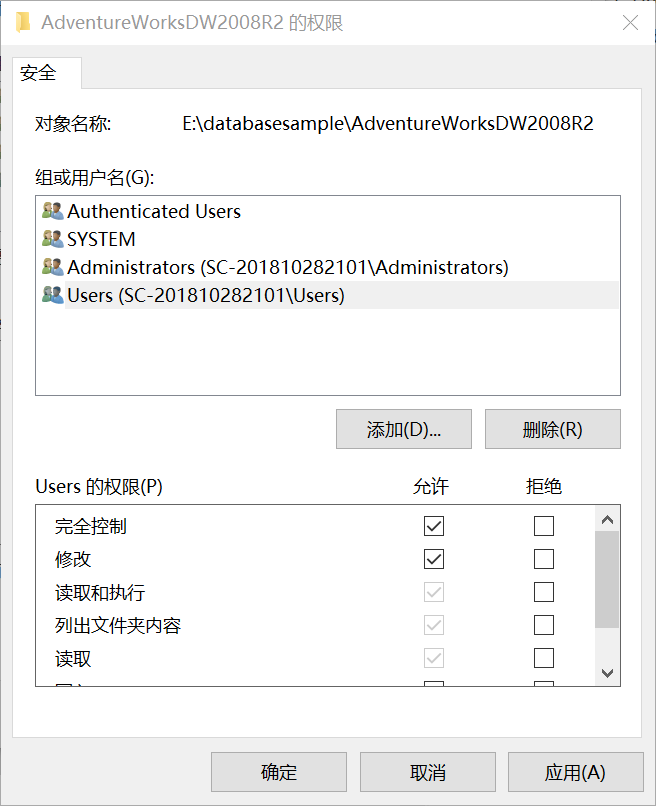
在SQL Server 平台上，利用AdventureWorks数据库作为商业智能解决方案

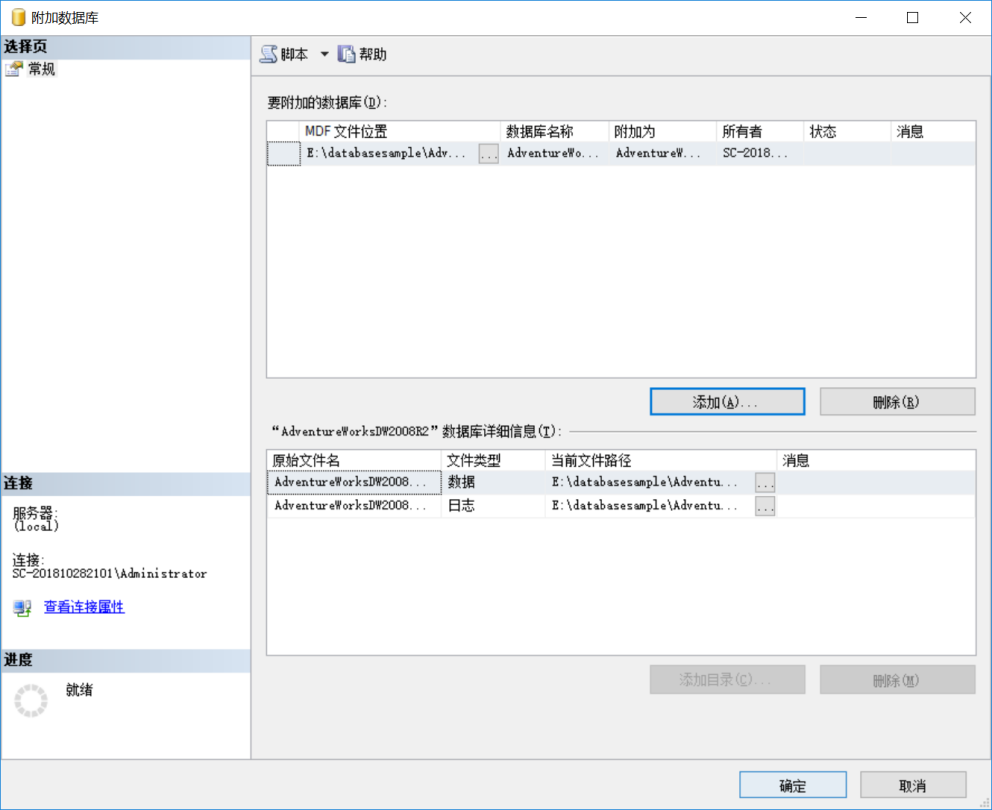
的数据源，设计并构造数据仓库，建立OLAP和数据挖掘模型，并以输出报表的形式满足决策支持的查询需求。

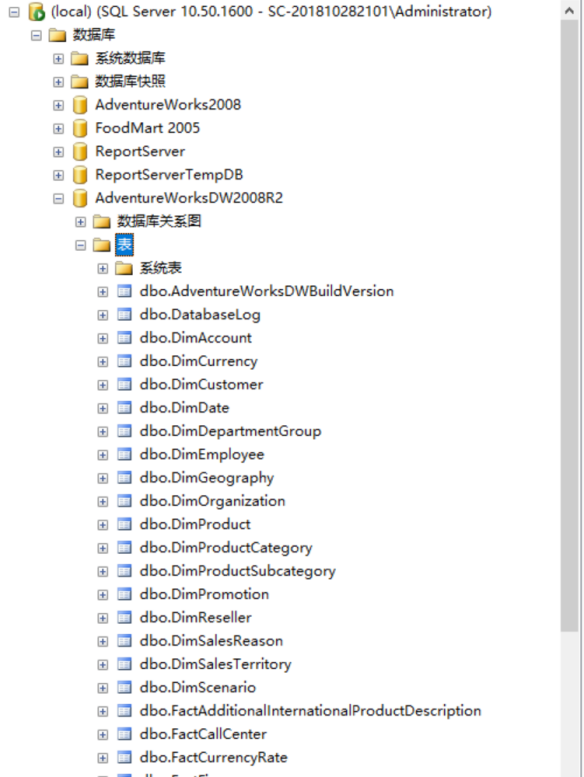
**【实验内容】**

**前期工作准备**

1. **安装SQL SERVER**，这里选择SQL SERVER的版本是2008 R2企业版。
2. **附加AdventureWorks数据库**
3. 老师给的AdventureWorks sql server 2005示例数据库.msi在我的电脑上安装失败，这里选择从网上下载sql server 2008版AdventureWorksDW数据库。
4. 打开Management Studio附加数据库，在“对象资源管理器”的“数据库”上右击，选“附加(A)…”，在弹出的“附加数据库”对话框内，添加要附加的mdf文件，点击“确定”，发现出错了，经寻找出错原因发现是权限问题。
5. 在mdf文件所在文件夹右击，属性，安全，编辑，选中Users，在下面的权限列表中队“完全控制”打钩，确定，在附加一次数据库，问题解决。







步骤 1： 需求分析：以决策者的视角分析和设计数据仓库的需求；

**选定AdventureWorks数据库作为商业智能解决方案的数据源，设计帮助企业进行销售管理的需求。**

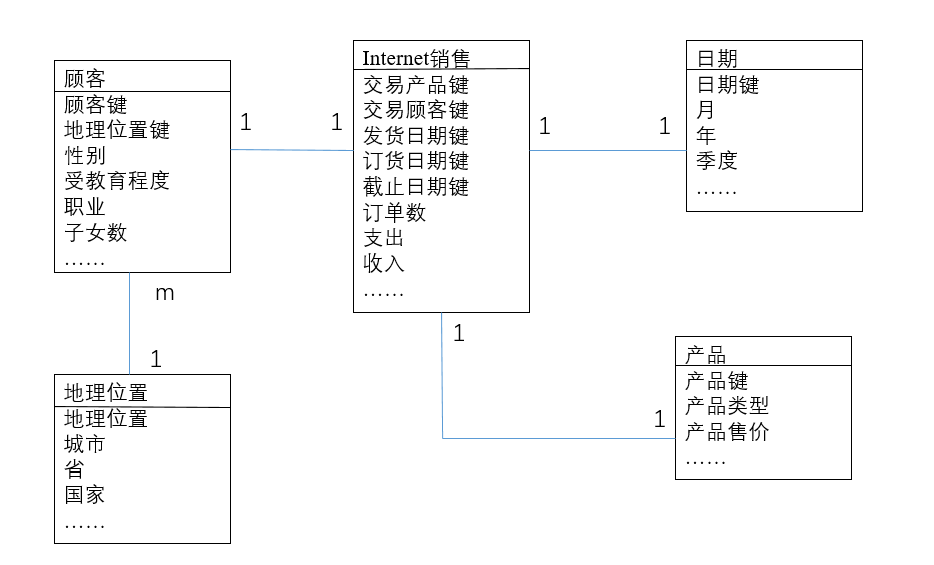
步骤 2： 根据所设计的需求， 确定本数据仓库的主题和主题与边界；

**选定网络销售为主题域，选定网络销售主题域的商业维度为产品、客户、日期和地理位置。**

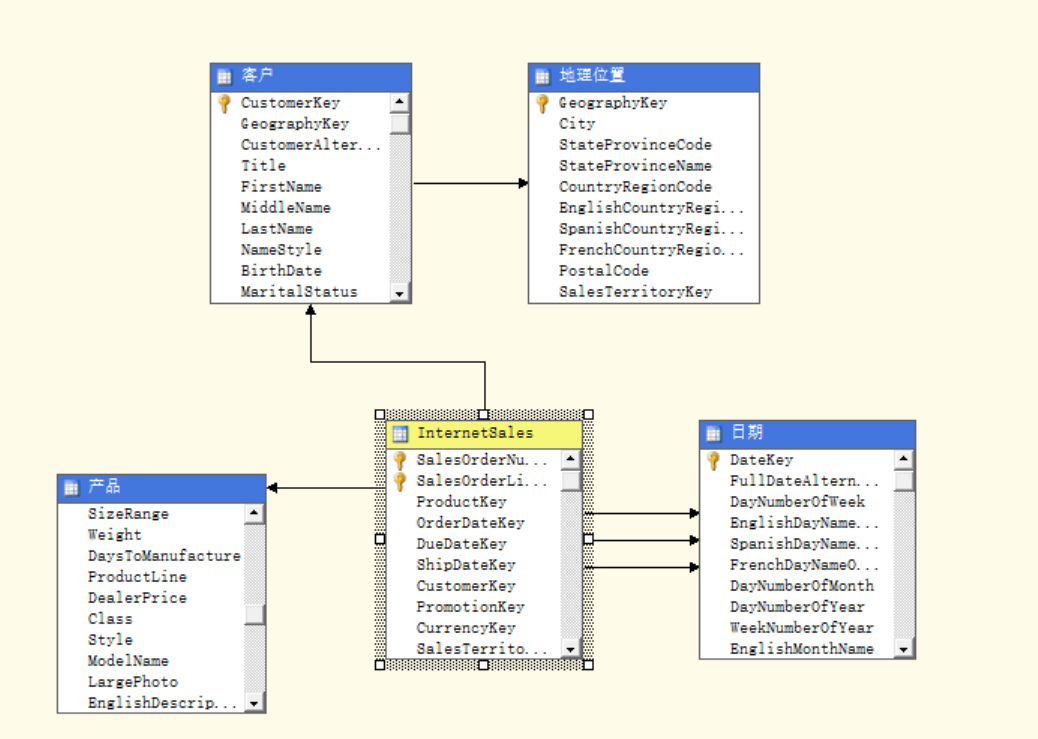
步骤 3： 设计并构造逻辑模型；

**确定FactInternetSales表为事实表、确定DimCustomer、DimDate、DimGeography、和DimProduct为使用的维度表。**

**ER图：**



**ER图转化为雪花模型：**

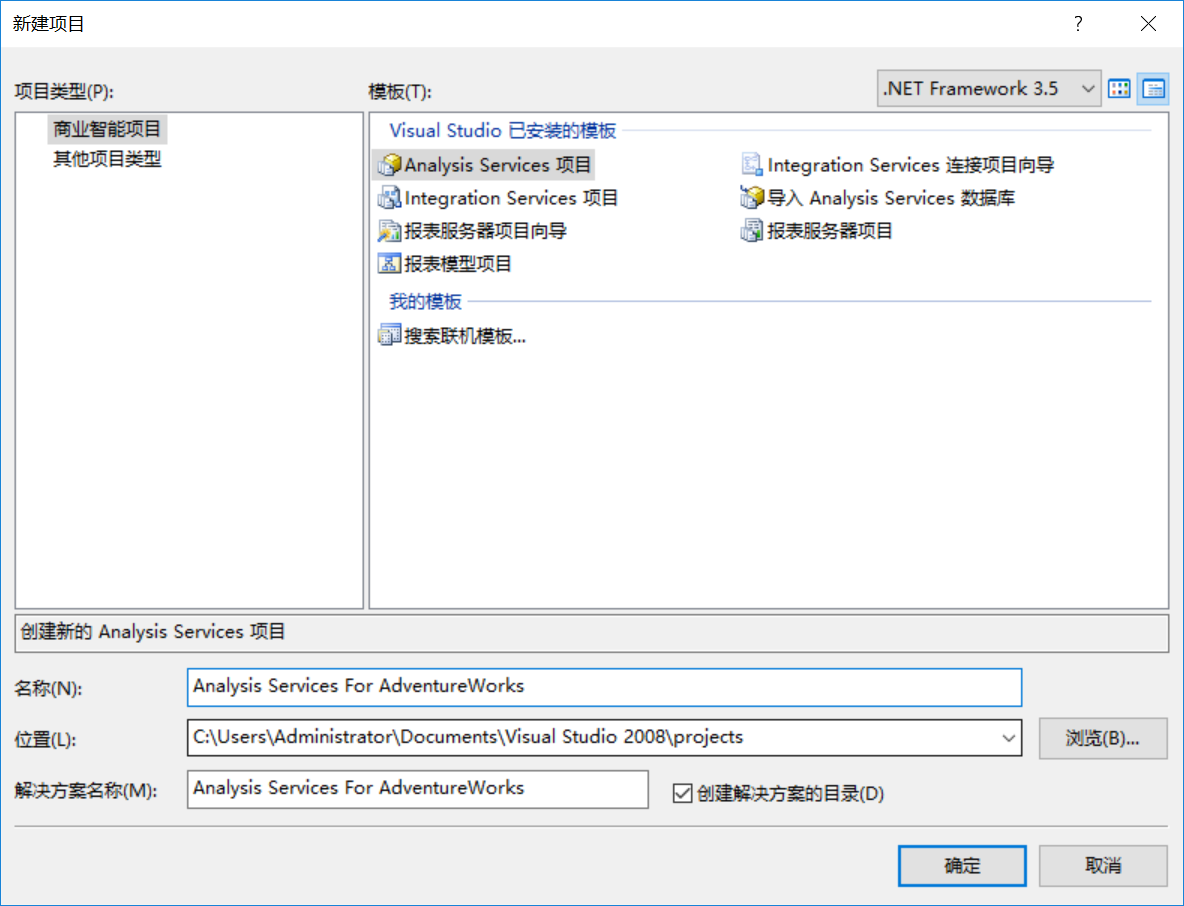


步骤 4： 进行数据转换和抽取，建立数据仓库：创建数据源，建立 OLAP 和

挖掘模型，使用多维数据集进行分析，建立数据挖掘结构和数据挖掘模型，创建报表。

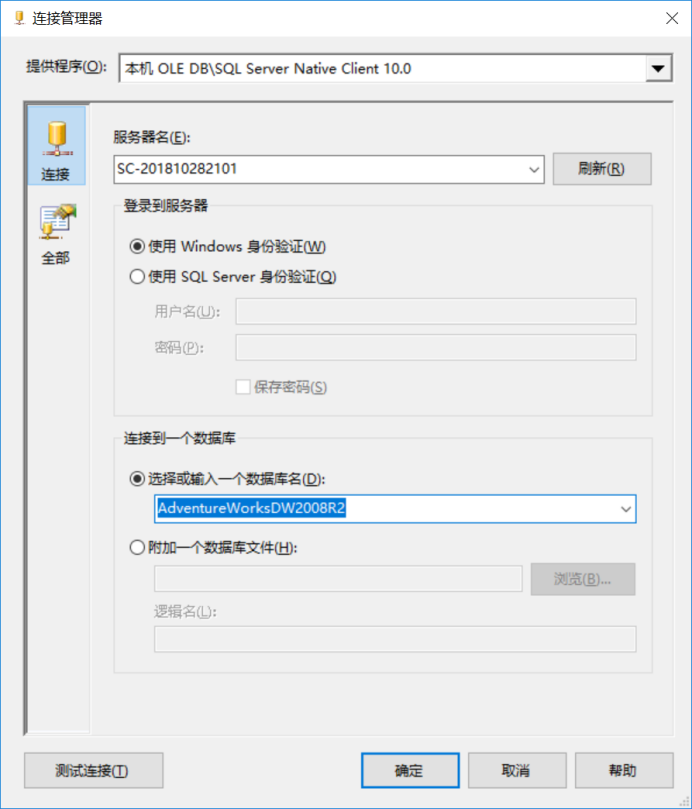
1. **创建Analysis Services项目**

打开SQL Server Business Intelligence Development Studio，点击文件—>新建—>项目，项目类型选择商业智能项目，Visual Studio已安装的模板选择Analysis Service项目，名称起名为Analysis Services For AdventureWorks，点击确定。

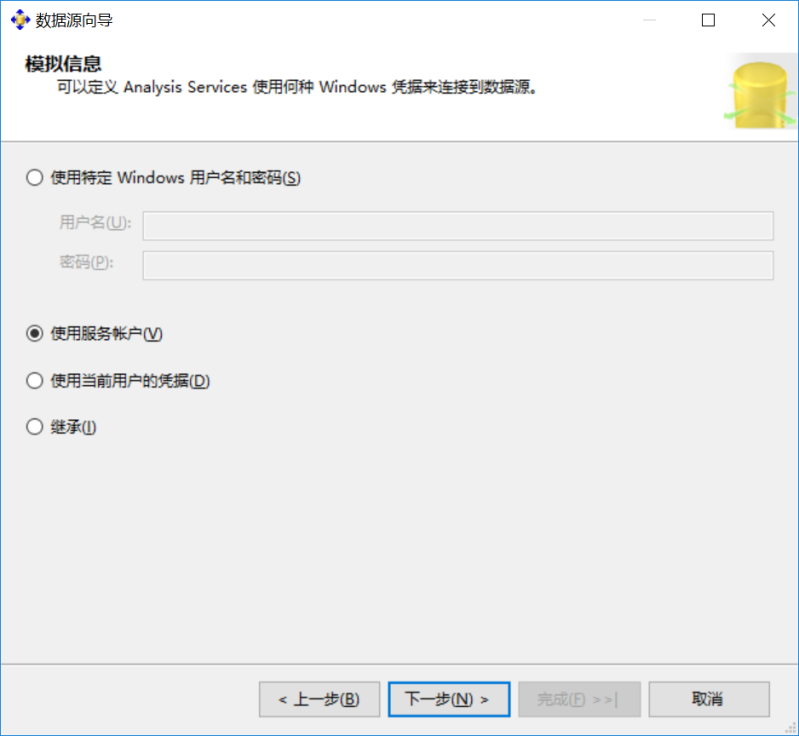


**2、定义数据源**

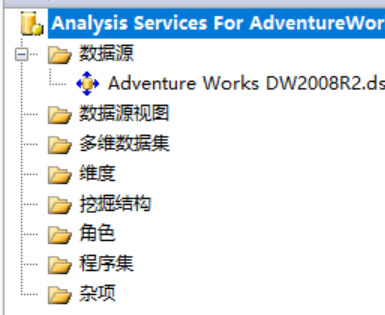
1）在解决方案资源管理器中，右键单击“数据源”，然后单击“新建数据源”。在“欢迎使用数据源向导”页上，单击“下一步”可打开“选择如何定义连接”页。在“选择如何定义连接”页上，可以基于新连接、现有连接或以前定义的数据源对象来定义数据源。本实验我们基于新连接定义数据源。选中“基于现有连接或新连接创建数据源”，再单击“新建”。在“连接管理器”对话框中，为数据源定义连接属性。在“提供程序”列表中，确保已选中“本机 OLE DB\SQL Server Native Client 10.0”。在“服务器名称”文本框中，选择本机名。确保已选中“使用 Windows 身份验证”。在“选择或输入数据库名称”列表中，选择AdventureWorksDW2008R2。点击“测试连接”以测试与数据库的连接。单击“确定”然后单击“下一步”。



2）在该向导的“模拟信息”页上，可以定义 Analysis Services 用于连接数据源的安全凭据。在选中“Windows 身份验证”时，模拟会影响用于连接数据源的 Windows 帐户。Analysis Services 不支持使用模拟功能来处理 OLAP 对象。选择“使用服务帐户”，然后单击“下一步”。

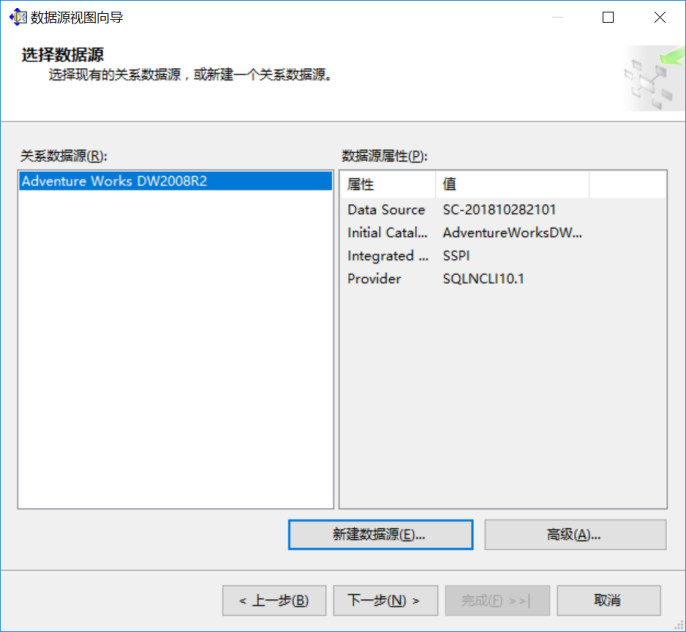


3）在“完成向导”页上，键入名称AdventureWorksDW2008R2，然后单击“完成”以创建新数据源。

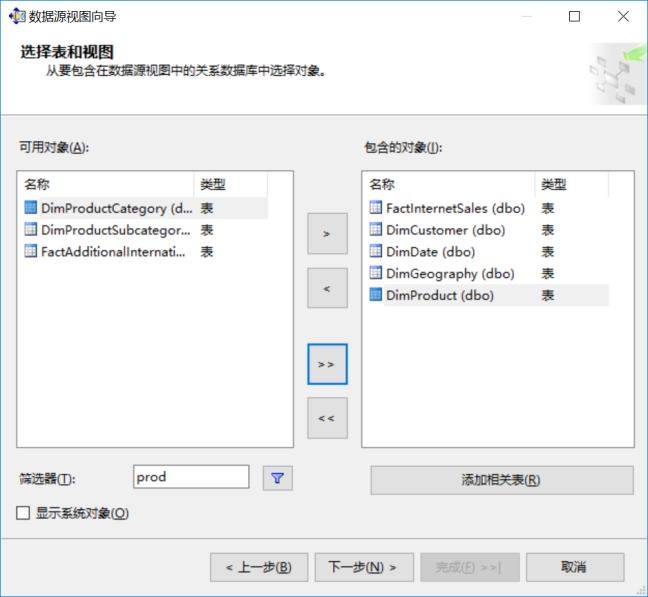


**3、定义数据源视图**

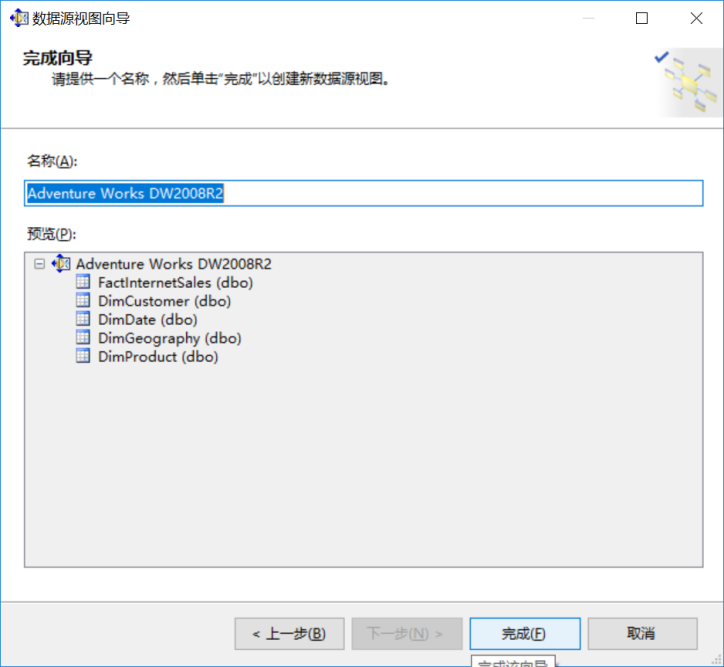
1）在解决方案资源管理器中，右键单击“数据源视图”，再单击“新建数据源视图”。在“欢迎使用数据源视图向导”页中，单击“下一步”。此时将显示“选择数据源”页。“关系数据源”下的AdventureWorksDW2008R2数据源已被选中。单击“下一步”。

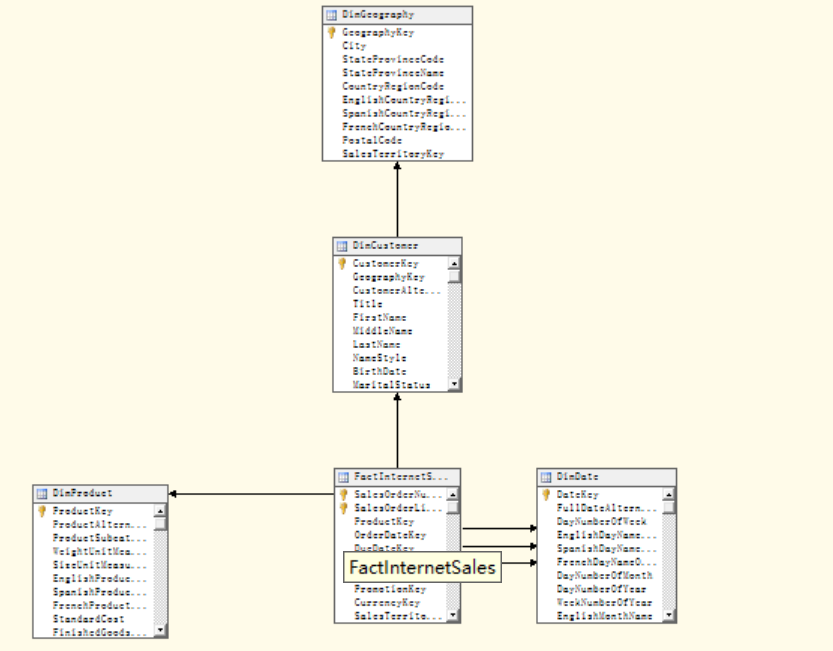


2）在“选择表和视图”页上，可以从选定的数据源提供的对象列表中选择表和视图。可以筛选此列表，为选择表和视图提供帮助。这里我们选择FactInternetSales、DimCustomer、DimDate、DimGeography、和DimProduct五个表。



3）单击“完成”，此时，AdventureWorksDW2008R2数据源视图将显示在解决方案资源管理器的“数据源视图”文件夹中。数据源视图的内容还将显示在 Business Intelligence Development Studio 的数据源视图设计器中。此设计器包含以下元素：“关系图”窗格:其中将以图形方式显示各个表及其相互关系;“表”窗格:其中将以树的形式显示各个表及其架构元素;“关系图组织程序”窗格:在其中创建子关系图，用于查看数据源视图的子集;一个特定于数据源视图设计器的工具栏。点击关系图窗格中的箭头即可查看某种关系的详细信息。注意在 FactInternetSales 表和 DimDate 表之间存在三种关系。每个销售都具有三个与其关联的日期：订单日期、到期日期和发货日期。

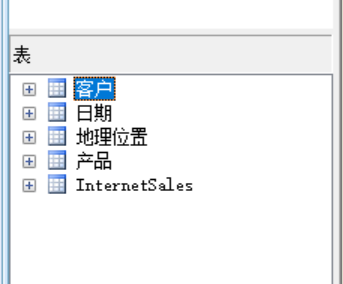




**3、修改表的默认名字，提供更友好的名字**

在数据源视图设计器的“表”窗格中，右键单击 FactInternetSales 表，然后单击“属性”以便在AdventureWorksDW2008R2数据源视图中显示FactInternetSales对象的属性，打开属性工具边栏。

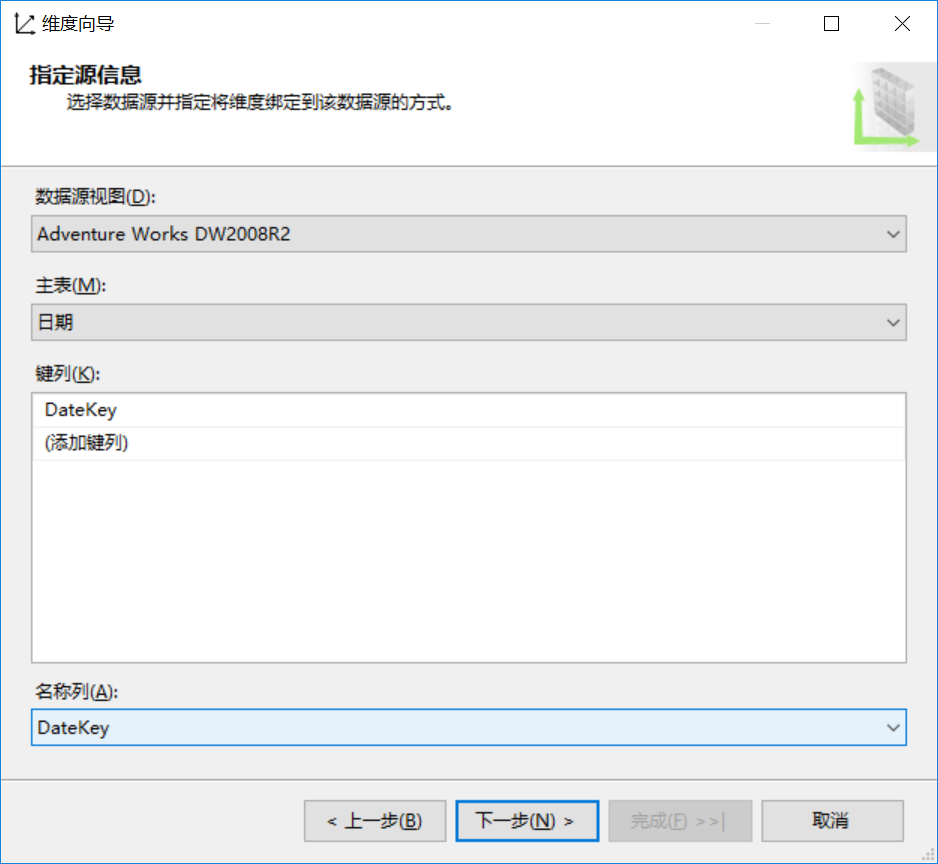
将FactInternetSales对象的FriendlyName属性更改为 InternetSales。同理更改其他表的FriendlyName属性以提供更友好的名字。



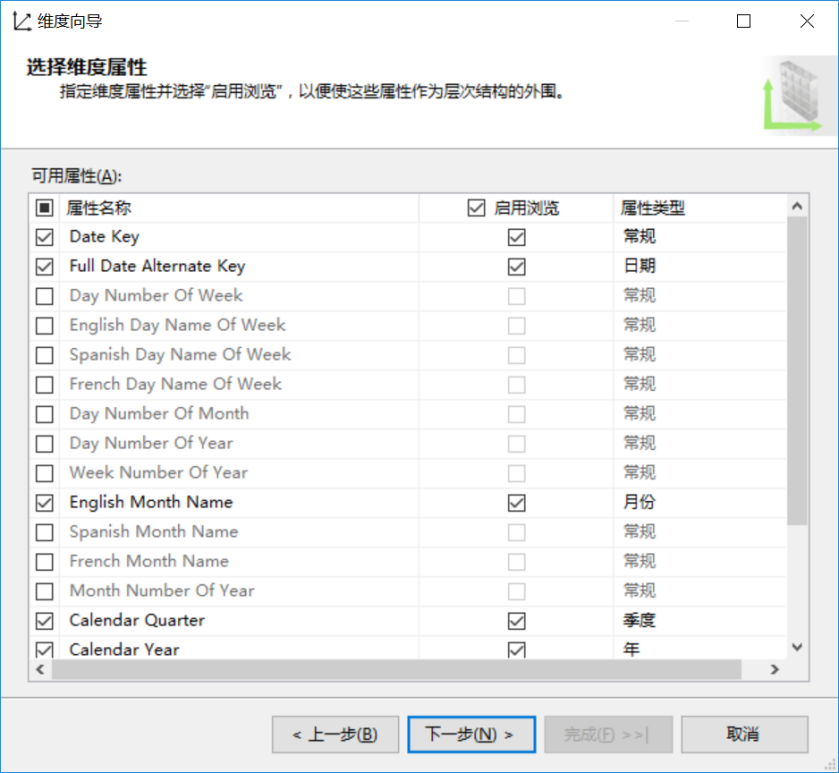
**4、定义维度**

在解决方案资源管理器中，右键单击“维度”，然后单击“新建维度”。在“欢迎使用维度向导”页上，单击“下一步”。在“选择创建方法”页上，验证是否选择了“使用现有表”选项，然后单击“下一步”。在“指定源信息”页上，验证是否选择了AdventureWorksDW2008R2数据源视图。在“主表”列表中，选择“日期”，

键列设置为“DateKey”,名称列设置为“DateKey”，点击“下一步”。



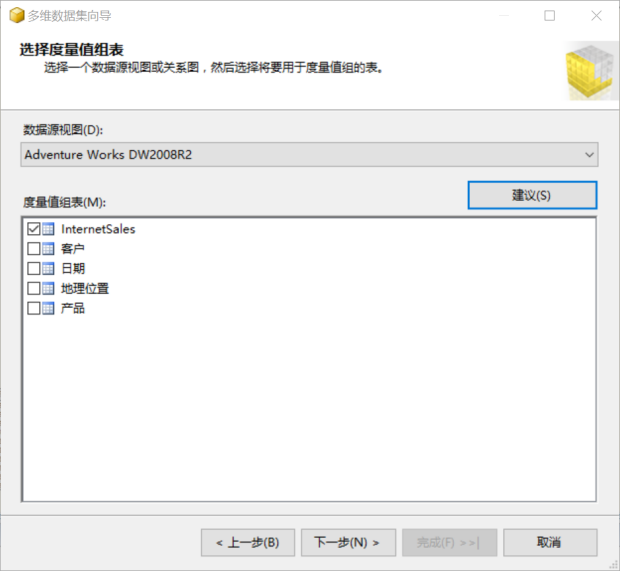
在“选择维度属性”页上，选中下列属性旁的复选框：English Month Name，Calendar Quarter、Calendar Year、Calendar Semester。将“完整日期备用键”属性的“属性类型”列的设置从“定期”更改为“日期”。为此，请单击“属性类型”列中的“定期”。然后单击箭头展开选项。接下来，单击“日期”>“日历” > “日期”。单击“确定”。 重复这些步骤来如下更改下列属性的属性类型：“英文月份名称”更改为“月份”， “日历季度”更改为“季度”， “日历年”更改为“年”，“日历半年”更改为“半年”。

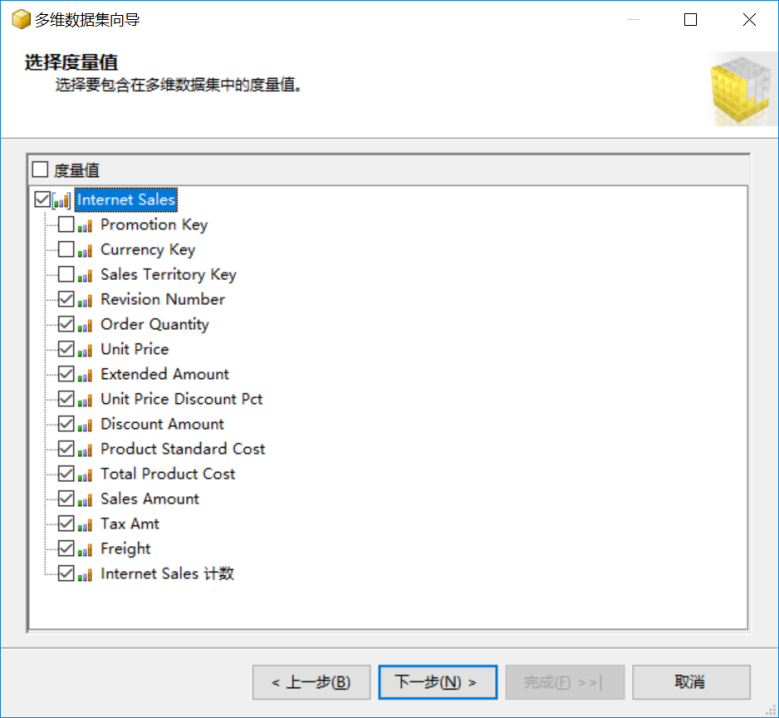


单击“下一步”，单击“完成”以完成日期维度的创建。同理完成“客户”维度、“地域”维度和“产品”维度的创建。

**5、定义多维数据集及其属性**

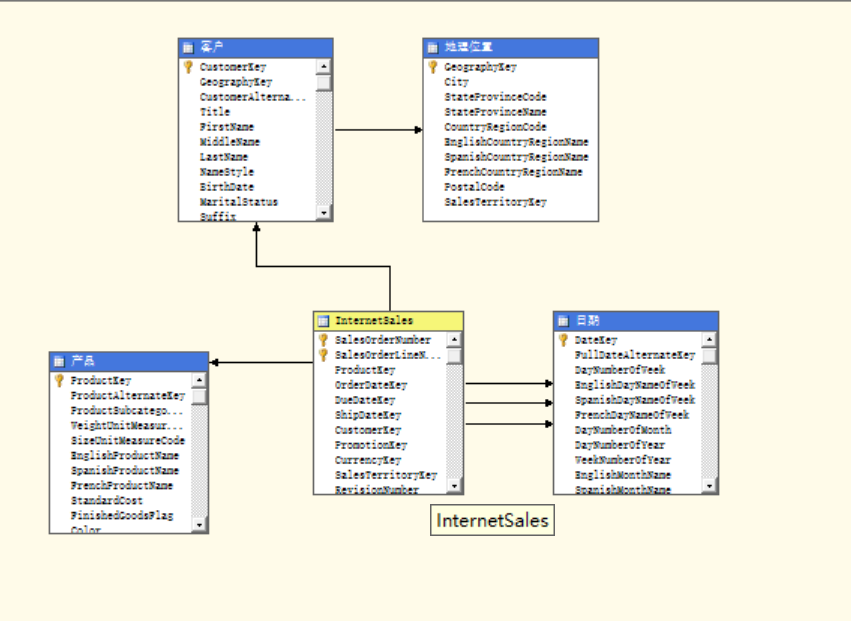
在解决方案资源管理器中，右键单击“多维数据集”，然后单击“新建多维数据集”。在“欢迎使用多维数据集向导”页上，单击“下一步”。在“选择创建方法”页上，确认已选中“使用现有表”选项，然后单击“下一步”。在“选择度量值组表”页上，确认已选中AdventureWorksDW2008R2数据源视图。度量值组表选择InternetSales，单击“下一步”，在选择度量值窗口中清除“Promotion Key”、“Currency key”、“Sales Territory Key”三个度量值的复选框（这里选择你感兴趣的度量值）。单击“下一步”。





选择现有维度窗口中选择已定义的四个维度，在选择新维度窗口中所有可以建立的新维度的勾选取消。在完成向导中点击“完成”。

在解决方案资源管理器的 Analysis Services For AdventureWorks项目中，AdventureWorksDW2008R2多维数据集显示在“多维数据集”文件夹中，而数据库维度则显示在“维度”文件夹中。此外，多维数据集设计器在开发环境的中央显示 AdventureWorksDW2008R2多维数据集。事实表示黄色的，维度表是蓝色的。



**6、部署Analysis Service项目**

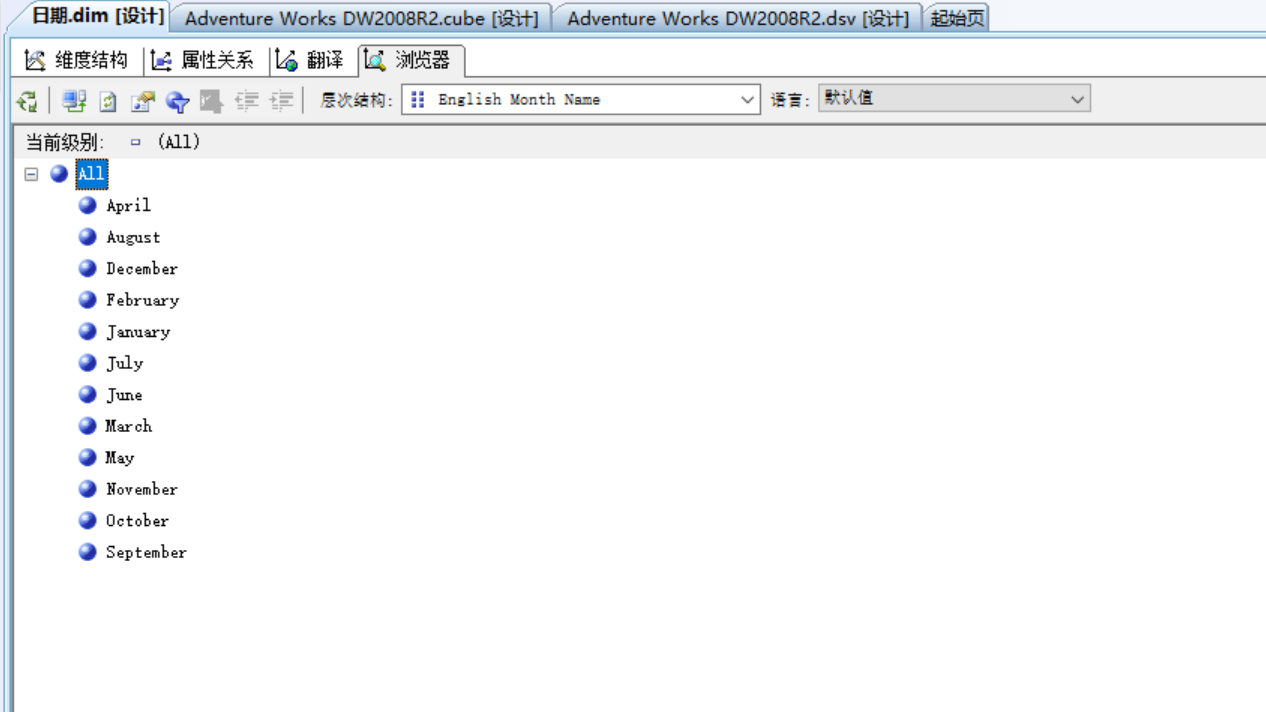
点击“浏览器”选项卡，提示无法访问Analysis Services For AdventureWorks数据库。由于浏览多维数据集尚未部署到 Analysis Services 实例中，因此无法对其进行浏览。此时，Analysis Services For AdventureWorks项目中的多维数据集只是一个可以部署到任何 Analysis Services 实例的多维数据集定义。部署和处理多维数据集时，将在 Analysis Services 实例中创建定义的对象，然后用基础数据源的数据填充这些对象。

在解决方案资源管理器中，右键单击“Analysis Services For AdventureWorks”项目，单击“部署”。

Business Intelligence Development Studio 将生成Analysis Services For AdventureWorks项目，然后使用部署脚本将其部署到指定的 Analysis Services 实例中。部署进度将在下列两个窗口中显示：“输出”窗口和“部署进度 – Analysis Services For AdventureWorks”窗口。打开“输出”窗口，如果需要，可通过单击“视图”菜单上的“输出”实现。“输出”窗口显示部署的整体进度。“部署进度 – Analysis Services Tutorial”窗口显示部署过程中每个步骤的详细信息。

**8、浏览已部署的多维数据集**

切换到 Business Intelligence Development Studio 中的“日期”维度的“维度设计器”。单击“浏览器”选项卡，层次结构下拉框中选择“English Month Names”可显示，日期维度月份属性层次结构的“所有”成员。



切换到 Business Intelligence Development Studio 中的“多维数据集设计器”。选择“浏览器”选项卡，然后在设计器的工具栏上单击重新连接图标。该设计器的左窗格会显示AdventureWorksDW2008R2 多维数据集中的对象。在“浏览器”选项卡的右侧有两个窗格：上部窗格是“筛选器”窗格，下部是“数据”窗格。



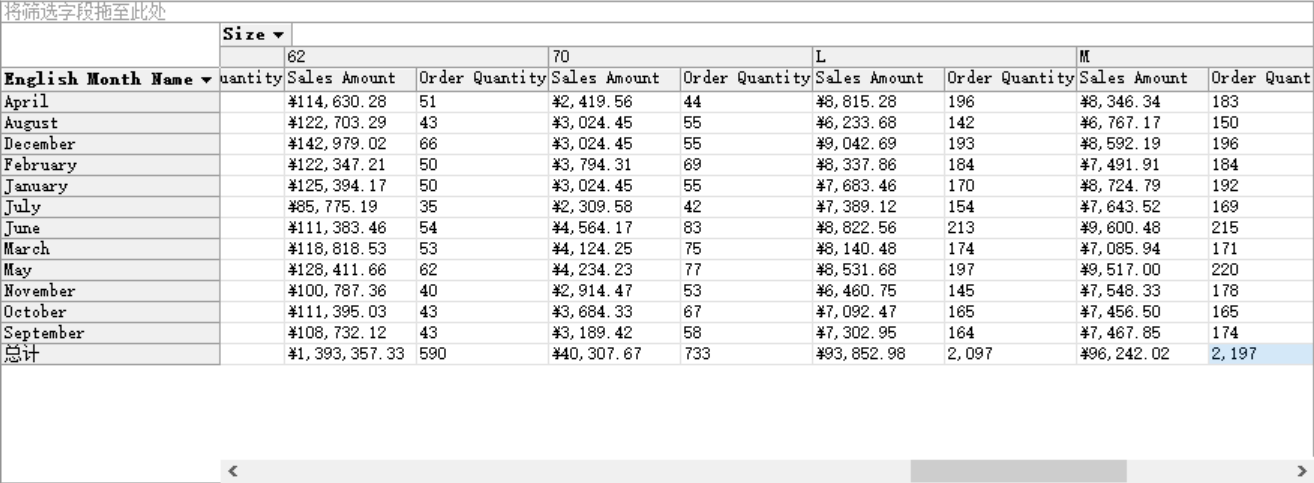
**9、创建报表**

将Internet Sales的“Sales Amount”和“Order Quantity”属性拖动到“将汇总或明细字段拖至此处”区域，将“Ship Date”的“English Month Name”属性拖动到“将行字段拖至此处”区域。



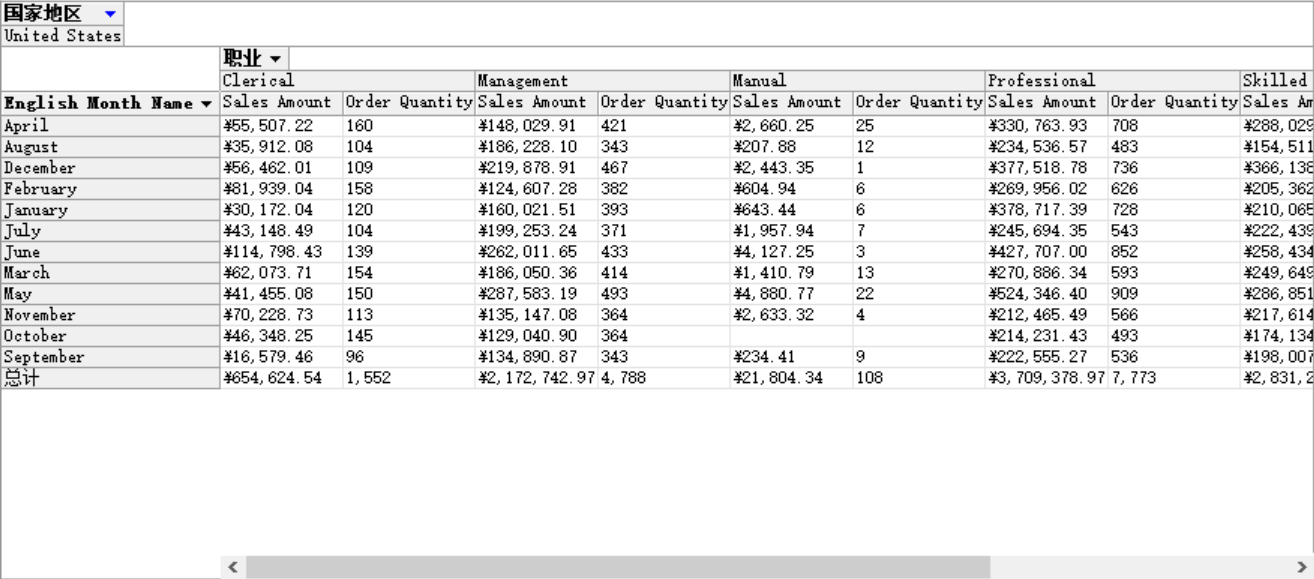
**结论：五月销售额和订单数最多，九月销售额最低，八月订单数最少，企业可以据此调整产品的生产数量。**

将“产品”的“Size”属性字段拖动到“将列字段拖至此处”区域。



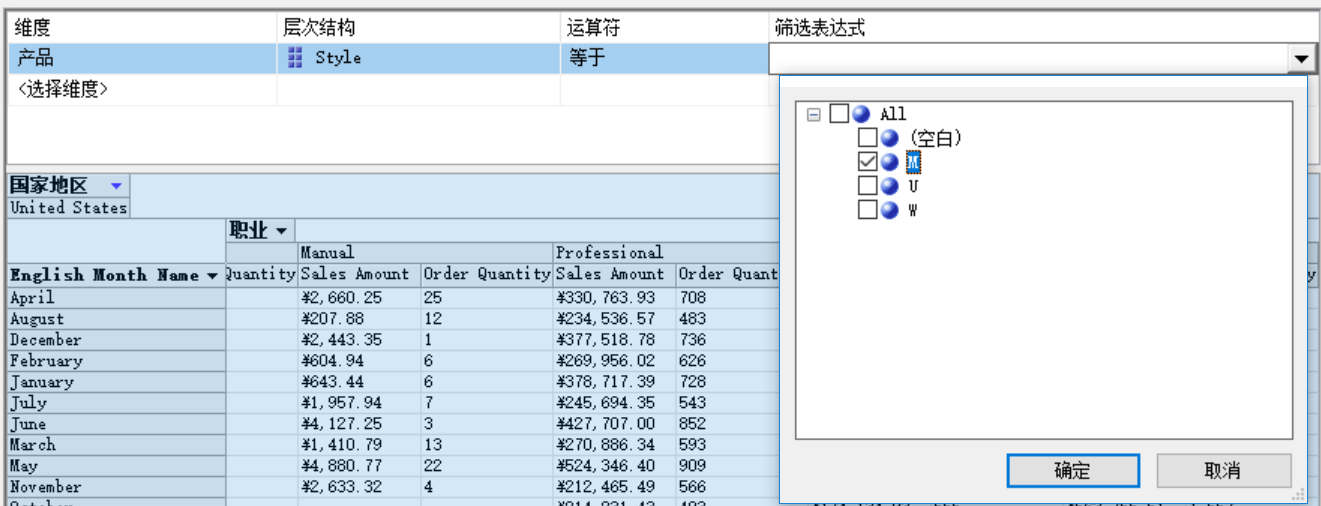
**结论：M号车订单数最多，56号车订单最少，企业可以根据订单数量的比例调整不同型号车的生产数量比例。**

将“客户”的“职业”属性拖动到“将行字段拖至此处”区域，将“客户”的“国家”属性拖动到“将筛选字段拖至此处”区域，并把筛选条件设置为“United States”。



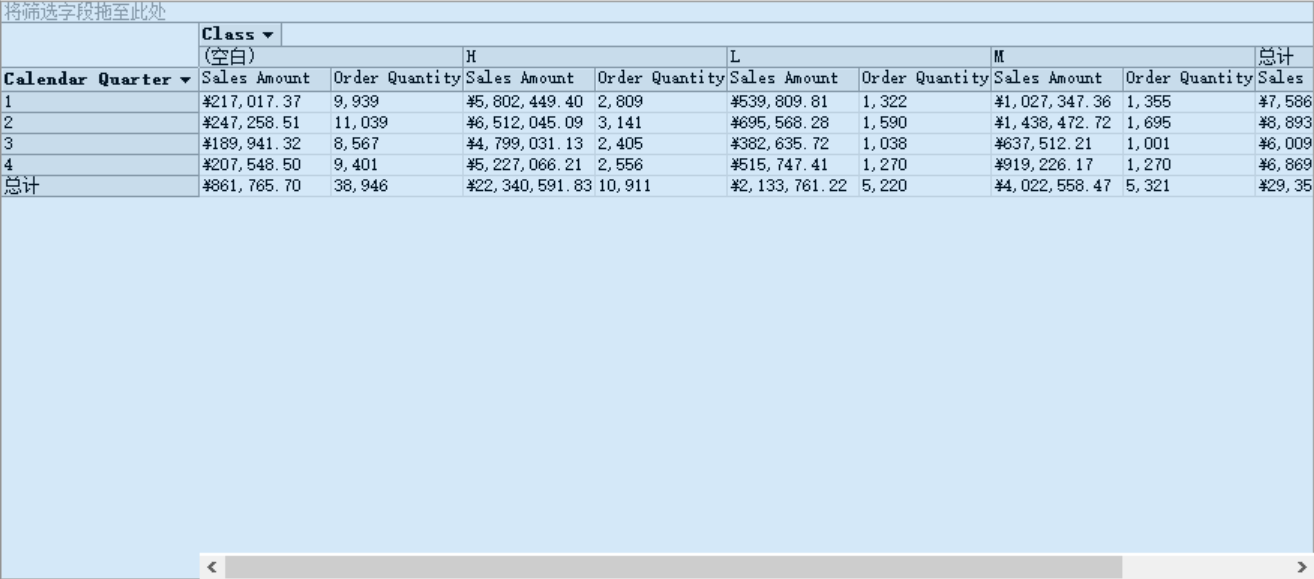
**结论：在美国，职业为Professional和Skilled Manual对公司产品的需求比较大，可以据此制定目标客户。**

还可以通过以下操作来对维度的属性进行筛选操作。

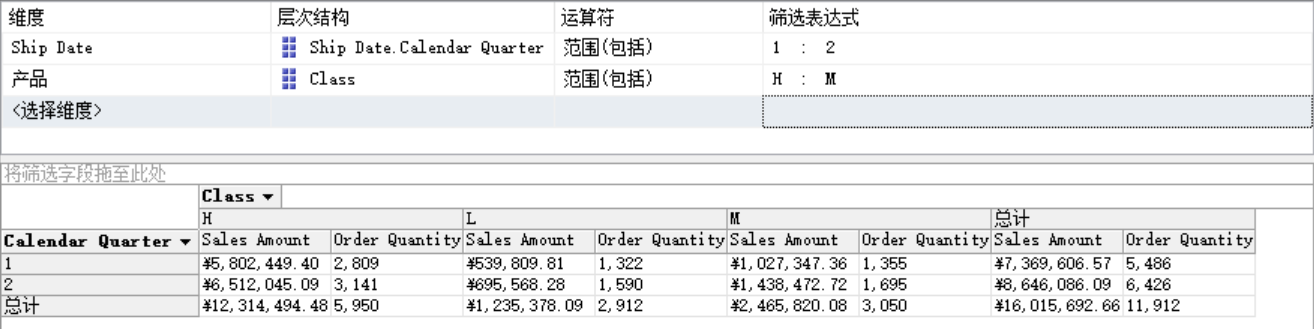


**其它多维数据分析的基本操作：**

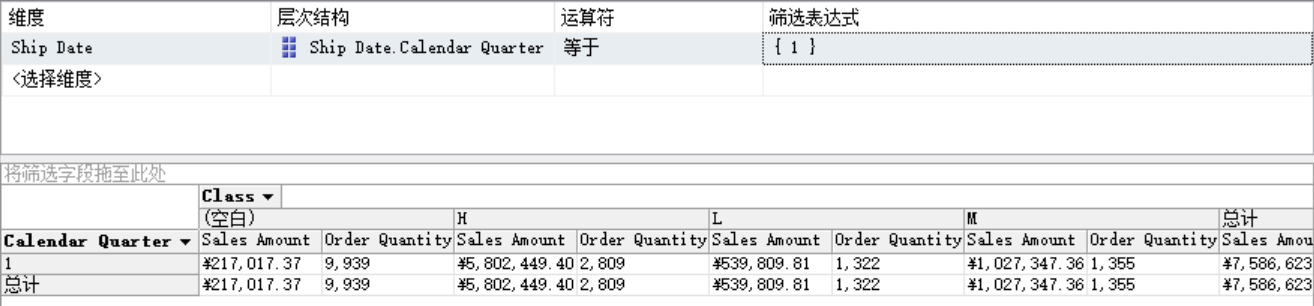
**原报表：**



**切块：**

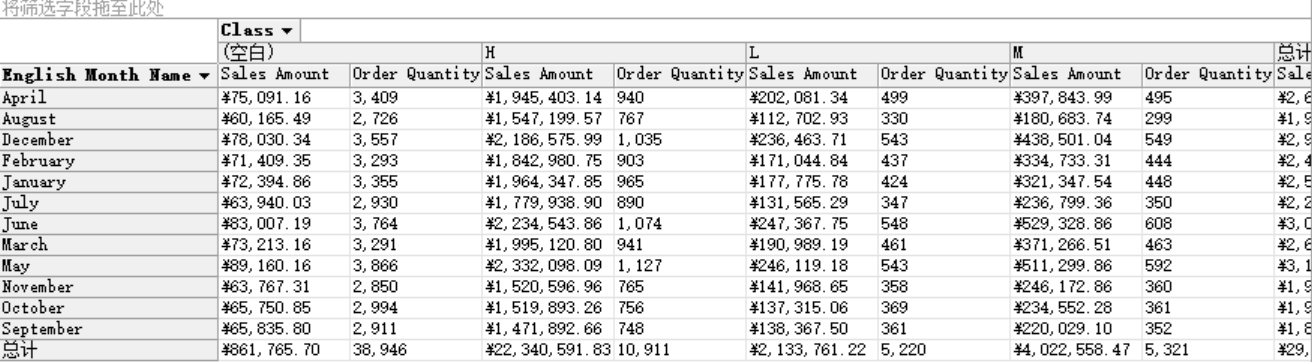


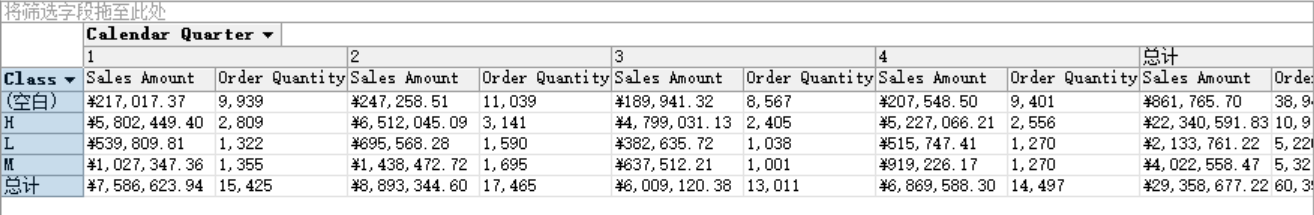
**切片：**



**钻取：**

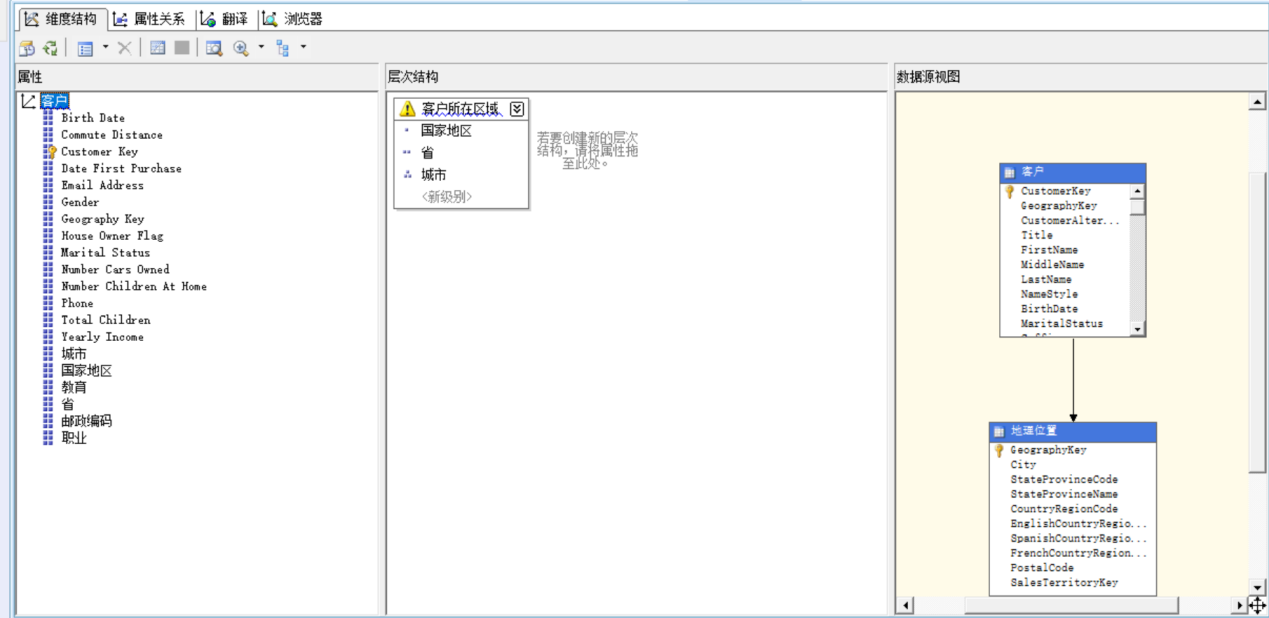
（对时间维度向下钻取）



**旋转：**

**【实验小节】**

1. 当新建完多维数据集后，点击浏览器卡发现提示“无权访问Analysis Services For AdventureWorks数据库,或该数据库不存在”，经查阅资料以后得知，浏览的多维数据集尚未部署到 Analysis Services 实例中，因此无法对其进行浏览。此时，Analysis Services For AdventureWorks 项目中的多维数据集只是一个可以部署到任何 Analysis Services 实例的多维数据集定义。部署和处理多维数据集时，将在 Analysis Services 实例中创建定义的对象，然后用基础数据源的数据填充这些对象。
2. 多维数据集向导可以使用不带聚合的多维联机分析处理 (MOLAP) 存储模式，为多维数据集定义单个分区。通过 MOLAP，所有叶级别数据和所有聚合均存储在多维数据集中，以便最大限度地提高性能。聚合是预先计算好的数据汇总，聚合可以在问题提出之前准备好答案，从而可以缩短查询响应时间。可在“分区”选项卡上定义其他分区、存储设置和写回设置。
3. 以上实验在对Internet Sales事实表的OLAP分析中无法选择二级维度表“地理位置”的元素，解决此问题可以通过在“客户”维度设计器中将“地理位置”维度的属性拖动到“客户”维度的属性列表中解决。



1. 掌握了SQL SERVER自带的Analysis Manager的使用方法；学会了对某个数据源的联机分析；学会了如何建立和编辑多维数据集以及进行OLAP操作。
2. 对比比较了数据仓库不同的概念模型：星型模型、雪花模型、事实模型。
3. 对比比较了数据仓库不同的存储模型：ROLAP、MOLAP、HOLAP。
4. 实验了诸如切片、切块、旋转和钻取等OLAP的基本操作。

写在最后的话：这门课学长戏称它为研究生几个“巨麻烦”的课之一，但是一个学期下来我真的是收获满满，在这里谢谢王晓茹教授的精心备课，感谢助教们判我们的作业（同样身为助教的我知道其中的心酸），给老师和助教提前拜个早年O(∩\_∩)O！