

## **Курс з інтелектуального аналізу даних (Data Mining) з використанням платформи MS SQL**

Data Mining - сучасна технологія аналізу інформації з метою знаходження у накопичених даних раніше невідомих, нетривіальних та практично корисних знань, необхідних для прийняття оптимальних рішень у різних галузях людської діяльності.

Microsoft SQL Server 2019 надає інтегроване середовище для створення моделей Data Mining та роботи з ними. В курсі описано чотири сценарії (адресне розсилання, прогнозування, аналіз купівельного кошика та аналіз послідовності покупок) для демонстрації методів використання алгоритмів Data Mining, а також супутнього інструментарію, що входить до складу SQL Server Analysis Services 2019.

Засоби аналізу даних OLAP та Data Mining об'єднані у два середовища розробки: Business Intelligence Development Studio та SQL Server Management Studio. У програмі Business Intelligence Development Studio можна створювати від'єднані від сервера проекти. У цьому випадку готовий проект можна завантажити на сервер, але ніщо не заважає працювати з проектом, який працює із сервером безпосередньо.

Усі засоби керування моделями Data Mining доступні у редакторі моделей. З його допомогою можна створювати, переглядати, порівнювати різні моделі, а також створювати на їх основі прогнози.

Після створення моделі можна провести її аналіз на предмет виявлення цікавих для нас шаблонів (паттернів) та правил. Залежно від застосовуваного алгоритму серед розробки використовується різні уявлення просматриваємої моделі.

Оскільки найчастіше проект містить кілька моделей Data Mining, попередньо створених на основі наших припущень, виникає завдання визначення найбільш адекватної з них. Для вирішення цього завдання в редакторі є засіб порівняння моделей Mining Accuracy Chart. З використанням цього інструменту можна передбачити точність моделі та вибрати найкращу з них.

Для створення прогнозів використовується мова Data Mining Extensions (DMX), яка є розширенням SQL і містить команди для створення, зміни та здійснення передбачень на основі різних моделей. Створення таких прогнозів може бути складним завданням, тому в редакторі є інструмент під назвою Prediction Query Builder, який є візуальним засобом створення DMX-запитів.

Крім інструментарію для роботи з моделями, не менш значущими є способи створення моделей. Ключовим моментом створення моделі є вибір алгоритму виявлення даних. SQL Server 2019 Analysis Services включає наступні дев'ять алгоритмів:

- Дерево рішень (Microsoft Decision Trees)
- Кластеризація (Microsoft Clustering)
- "Наївний" Байєс (Microsoft Na?ve Bayes)
- Кластеризація послідовностей (Microsoft Sequence Clustering)
- Тимчасові ряди (Microsoft Time Series)
- Асоціативні правила (Microsoft Association)
- Нейронна мережа (Microsoft Neural Network)
- Лінійна регресія (Microsoft Linear Regression)
- Логістична регресія (Microsoft Logistic Regression)

Використовуючи комбінацію цих алгоритмів можна створювати рішення для більшої частини завдань з виявлення прихованих закономірностей у великих обсягах даних.

Найбільш важливі кроки створення моделі Data Mining полягають у отриманні, об'єднанні, очищенні та передобробці даних для подальшого використання алгоритмом. До складу SQL Server 2019 входять засоби перетворення даних - SQL Server Integration Services (SSIS), які дозволяють об'єднувати, очищати, перевіряти і проводити попередню обробку даних.

Для демонстрації можливостей SQL Server ми використовуватимемо демо-проект AdventureWorksDW. Ця база включена в поставку SQL Server 2019 та спроектована належним чином для роботи з OLAP та моделями Data Mining. Щоб ця база була доступна, слід вибрати опцію установки цього проєкту під час інсталяції SQL Server.

### **Проект Adventure Works**

Проект AdventureWorksDW описує роботу виробника велосипедів - компанії "Adventure Works Cycles". Компанія займається виробництвом та реалізацією велосипедів з металевих та композиційних матеріалів на території Північної Америки, Європи та Азії. Головне виробництво, яке має 500 співробітників, знаходиться в місті Bothell, штат Вашингтон. Декілька регіональних офісів знаходяться безпосередньо на території ринків збуту.

Компанія реалізує продукцію оптом для спеціалізованих магазинів та в роздріб через інтернет. Для вирішення демонстраційних завдань ми будемо використовувати в базі AdventureWorksDW дані про інтернет продаж, оскільки вони містять дані, що добре підходять для аналізу.

### **База даних**

В базі інтернет продажів міститься інформація про 9 242 клієнтів, які живуть у 6 країнах, що належать до трьох регіонів:

- Серверна Америка (83%)
- Європа (12%)
- Австралія (7%)

Дані відносяться до періоду з 2002 до 2004 року. Вся продукція класифікована за категоріями, моделями та товарами.

### **Середовище розробки - Business Intelligence Development Studio**

Середовище розробки "Business Intelligence Development Studio" є інтегрованим середовищем розробки для створення проектів аналізу даних (business intelligence) і включає безліч інструментів. Вона дозволяє працювати з моделями у від'єднаному від сервера режимі, а потім розгортати оновлений проект на сервері.

Можна відзначити такі переваги роботи в єдиному середовищі:

- Середовище є потужним засобом управління проектами
- Ви можете об'єднувати кілька проектів в одне рішення (solution) та використати для них загальне уявлення.
- Повна інтеграція із системою зберігання вихідного коду дозволяє зручно організувати спільну роботу над проектом.

Проект Analysis Services є центральним у вирішенні business intelligence. Проект включає моделі Data Mining і куби OLAP, а також допоміжні об'єкти наповнення аналітичної бази даних. Зі студії можна створювати проекти аналізу даних і розгортати їх на безлічі серверів.

При роботі з раніше створеним проектом, який вже знаходиться на сервері, можна працювати з ним безпосередньо.

## **Среда управления - SQL Server Management Studio**

Среда SQL Server Management Studio является оболочкой для множества административных функций для управления компонентами SQL Server. Эта среда отличается от Business Intelligence Development Studio в первую очередь тем, что изменения сохраняются на сервер в реальном времени.

После предварительной очистки и предобработки данных большая часть всех дальнейших действий по созданию модели выполняется в среде Business Intelligence Development Studio. В ней выполняются действия по созданию и проверке моделей, а также итеративное тестирование для выявления наиболее адекватной модели. После получения нужной модели ее можно развернуть на сервере. После этого фокус смещается с разработки на поддержку и использование модели для чего подходит SQL Server Management Studio. С ее помощью можно администрировать базу данных и выполнять просмотр и создание предсказаний аналогично тому, как это выполняется в среде Business Intelligence Development Studio

## **Сервисы интеграции - Integration Services**

Integration Services (сервисы интеграции, SSIS) включают в себя средства получения (Extract), преобразования (Transform) и загрузки (Load) (ETL) данных из различных источников. Они используются для некоторых очень важных задач в проектах извлечения данных, таких как извлечение, объединение, очистка и предобработка данных, предшествующие созданию модели. В проектах извлечения данных и в процессе настройки моделей обычно необходимо проводить неоднократную предобработку и очистку данных. С помощью SSIS можно объединить задачи предобработки данных в единый пакет.

Среда SSIS содержит визуальный редактор, с помощью которого удобно создавать и отлаживать пакеты преобразования данных. Из него можно

загрузить пакеты на сервер и настроить планировщик для их выполнения. Это удобно использовать, к примеру, для автоматической предобработки новых поступающих данных.

## **Алгоритмы Data Mining**

Алгоритмы Data Mining являются основой для создания моделей. Набор алгоритмов Data Mining, входящих в SQL Server 2019, позволяют производить различные виды анализа данных.

### **□ Дерево решений - Microsoft Decision Trees**

Алгоритм "Дерево решений" предназначен для решения задач классификации и регрессии и хорошо подходит для прогнозирования. В алгоритме Microsoft Decision Trees используются как дискретные, так и непрерывные атрибуты.

В процессе построения модели алгоритм итеративно вычисляет степень влияния каждого входного атрибута модели на значения выходного атрибута и использует атрибут, влияющий на выходную переменную в наибольшей степени для разбиения узла дерева решений. Узел верхнего уровня описывает распределение значений выходного атрибута по всей совокупности данных. Каждый последующий узел описывается распределением выходного атрибута при соблюдении условий на входные атрибуты, соответствующие этому узлу. Модель продолжает расти до тех пор, пока разбиение узла на последующие узлы увеличивает вероятность того, что выходной атрибут будет принимать какое-то определенное значение по сравнению со всеми другими значениями, т.е. разбиение увеличивает качество прогноза. Алгоритм осуществляет поиск атрибутов и их значений, разбиение по которым позволяет с большей вероятностью правильно предсказать значение выходного атрибута.

## Кластеризация - Microsoft Clustering

Алгоритм кластеризации использует итеративный метод группировки записей набора данных в кластеры, обладающие сходными характеристиками. Используя разбиение на кластеры можно выявить в исследуемом массиве данных такие связи, которые невозможно обнаружить простым просмотром этих данных. Кроме того, с помощью алгоритмов кластеризации можно осуществлять прогнозирование. К примеру, объединить в группу людей, которые живут в одном районе, водят одну марку машин, имеют сходные предпочтения в пище и покупают один тип продукции. Такое объединение и есть кластер. Другой кластер может включать в себя людей, посещающих один ресторан, имеющих один уровень дохода и едущих дважды в год в отпуск в другие страны. Оценивая распределение данных в этих кластерах, можно лучше понять взаимосвязи различных характеристик исследуемых объектов, а также как эти взаимосвязи влияют на значение прогнозируемого атрибута.

В Microsoft Analysis Services 2019 для кластеризации используется модификации алгоритмов максимизации ожидания (Expectation Maximization) и К-ближайших соседей (K-Means).

В первом случае каждый кластер характеризуется своей функцией распределения атрибутов входных данных, которые вычисляются в ходе итеративного процесса. В процессе выполнения алгоритма итеративно максимизируется функция правдоподобия на пространстве параметров функций распределения для каждого кластера. При этом предполагается, что непрерывные атрибуты имеют совместное многомерное нормальное распределение, а дискретные - совместное дискретное распределение.

Во втором случае итеративно минимизируется сумма квадратов расстояний (в различных метриках) от каждого элемента данных до центра соответствующего кластера.

## **Наивный алгоритм Байеса - Microsoft Na?ve Bayes**

Наивный алгоритм Байеса предназначен для решения задач классификации и прогнозирования. В процессе его реализации вычисляются вероятности состояний входных атрибутов для каждого состояния выходного атрибута. Эти значения используются для вычисления вероятности того, что выходной атрибут принимает то или иное состояние при заданных значениях входных атрибутов. Алгоритм принимает только дискретные или дискретизированные атрибуты модели, а также исходит из предположения о стохастической независимости входных атрибутов ("наивность" алгоритма как раз обусловлена этими предположениями). Алгоритм Microsoft Na?ve Bayes представляет собой простую модель для анализа данных, которую можно рассматривать как средство первоначального разведочного анализа. Вследствие того, что большинство требуемой для построения модели информации вычисляется в процессе обработки соответствующего куба, результаты алгоритма Байеса возвращаются очень быстро. Это служит еще одним преимуществом использования алгоритма в качестве оптимального разведочного механизма для классификации и прогнозирования.

## **Временные ряды - Microsoft Time Series**

Алгоритм Microsoft Time Series создает модели, предназначенные для прогнозирования значений непрерывных переменных по времени и использованием как OLAP, так и реляционных источников данных. Например, можно использовать этот алгоритм для прогнозирования объема продаж и прибыли по историческим данным в разрезе регионов продаж.

Алгоритм позволяет осуществлять прогнозирование по нескольким непрерывным переменным. Временной ряд должен характеризоваться временной характеристикой по которой осуществляется прогноз, а также может содержать произвольный набор других характеристик, разделяющий ряд на несколько (например, регион, продавца и т.д.).

Алгоритм Microsoft Time Series может анализировать и использовать корреляции между различными прогнозируемыми характеристиками. В этом случае результат прогнозирования будет зависеть не только от исторических значений той же переменной, но и значений других прогнозируемых переменных. Например, продажи в определенном магазине могут зависеть от прошлых продаж в другом магазине.

Алгоритм Microsoft Time Series является разновидностью алгоритмов ART (Autoregressive Trees - деревья авторегрессии). В процессе реализации алгоритма строится дерево решений, листьям которого соответствуют функции линейной регрессии.

### **Ассоциативные правила - Microsoft Association**

Алгоритм Microsoft Association предназначен в первую очередь для анализа покупательских корзин. Анализируется каждая пара атрибут=значение (например продукт=велосипед) как значение некой логической переменной. Алгоритм сканирует транзакции в базе данных для определения "частых" наборов таких пар. Набор считается "частым" если его support (поддержка, т.е. число транзакций, в которые входит данный набор) превышает определенный порог. Например, частым набором может быть {Пол="мужской", Семейный статус = "женат", Возраст="30-35"}. Каждый набор имеет размер, например, в этом случае, он составляет 3.

Часто транзакции описываются вложенными таблицами, например, строками заказа продажи в заказе. В этом случае ключ вложенной таблицы служит названием атрибута, а факт его наличия - значением. Например {Товар("Велосипед")=existing, Товар("Велосипедная кепка")=existing}.

Алгоритм Microsoft Association помимо выявления частых наборов предназначен для выявления правил вида  $A, B \Rightarrow C$ , характеризуемых вероятно-

стью выполнения (confidence), где  $\{A, B\}$ ,  $\{C\}$  частые наборы. Символ ' $\Rightarrow$ ' означает, что появление набора  $C$  в транзакции можно предсказать из факта появления наборов  $A$  и  $B$ . Например,  $\{\text{Фотокамера}=\text{existing}, \text{Батарейки}=\text{existing}\} \Rightarrow \{\text{Фотопленка}=\text{existing}\}$ . Порог значения вероятности соблюдения правила является параметром модели, определяющим факт рассмотрения каждого правила.

Ассоциативные правила используются для оптимизации кросс-продаж, целевых рекламных кампаний, определения политики скидок, в оптимизации размещения товаров на товарных полках в супермаркетах и т.д.

### **Кластеризация последовательностей действий - Microsoft Sequence Clustering**

Алгоритм Microsoft Sequence Clustering анализирует последовательности каких-либо фактов, представляющих собой временные последовательности дискретных переменных. Обычно такие последовательности атрибутов рассматриваются как происхождение событий в определенном порядке (например, порядок просмотра страниц сайта). Алгоритм предназначен для прогнозирования наступления последующих событий на основании уже осуществленного перехода между состояниями.

Алгоритм Microsoft Sequence Clustering является гибридом алгоритма последовательностей действий и алгоритмом кластеризации. В процессе реализации алгоритма выполняется группировка последовательности переходов в типичные кластера, характеризующиеся паттернами переходов. Каждый из этих кластеров анализируется с точки зрения распределения вероятности переходов отдельно. Типичным сценарием использования для этого алгоритма является задача анализа клиентов web-портала. У web-портала имеется набор связанных доменов, таких как новости, погода, финансы, почта, спорт и т.д. Каждый web-посетитель характеризуется последовательностью "кликов" - переходов в рамках каждого из этих доменов. Алгоритм Microsoft Sequence

Clustering может сгруппировать этих web-посетителей в более-менее однородные группы на основании их паттернов навигации по portalу. Эти группы могут быть визуализированы, обеспечивая представление шаблонов использования portalа каждой группой посетителей.

В алгоритме используется алгоритм Expectation Maximization для выявления кластеров, а также Марковские процессы первого порядка для моделирования переходов между состояниями (событиями).

### **Нейронные сети - Microsoft Neural Network**

Алгоритм Microsoft Neural Network предназначен для создания моделей классификации и регрессии путем конструирования многослойной нейронной сети перцептронов. Как и в случае алгоритма дерева решений, для каждого состояния выходного атрибута алгоритм вычисляет распределение вероятности входных атрибутов. Обработывается полный набор записей, при этом итеративно сравниваются предсказанные значения классификатора с известным значением. Ошибки классификации на первой итерации подаются на вход сети для изменения параметров на следующей итерации и т.д. Впоследствии, полученные вероятности используются для прогнозирования значения выходного атрибута на основании значений входных атрибутов. Одно из самых значительных отличий между этим алгоритмом и алгоритмом дерева решений состоит в том, что процесс обучения заключается в оптимизации параметров сети для минимизации ошибки классификации, в то время как алгоритм дерева решений осуществляет разбиение узлов для максимизации информации (минимизации энтропии). Алгоритм поддерживает как непрерывные так и дискретные типы атрибутов.

### **Линейная регрессия - Microsoft Linear Regression**

Алгоритм Microsoft Linear Regression представляет собой алгоритм регрессии являющийся частным случаем алгоритма Microsoft Decision Trees,

получаемый в случае запрета на разбиение узлов в дереве решений. Формула регрессии определена на всем объеме данных, т.е. на корневом узле дерева. Алгоритм предназначен для прогнозирования непрерывных атрибутов.

### **Логистическая регрессия - Microsoft Logistic Regression**

Алгоритм Microsoft Logistic Regression представляет собой алгоритм регрессии являющийся частным случаем алгоритма Microsoft Neural Network, получаемый в случае удаления скрытого слоя нейросети. Алгоритм поддерживает прогнозирование значений как непрерывных, так и дискретных атрибутов.

### **Описание работы с курсом**

В ходе данного курса вы будете работать в среде разработки Business Intelligence Development Studio (рис.1).

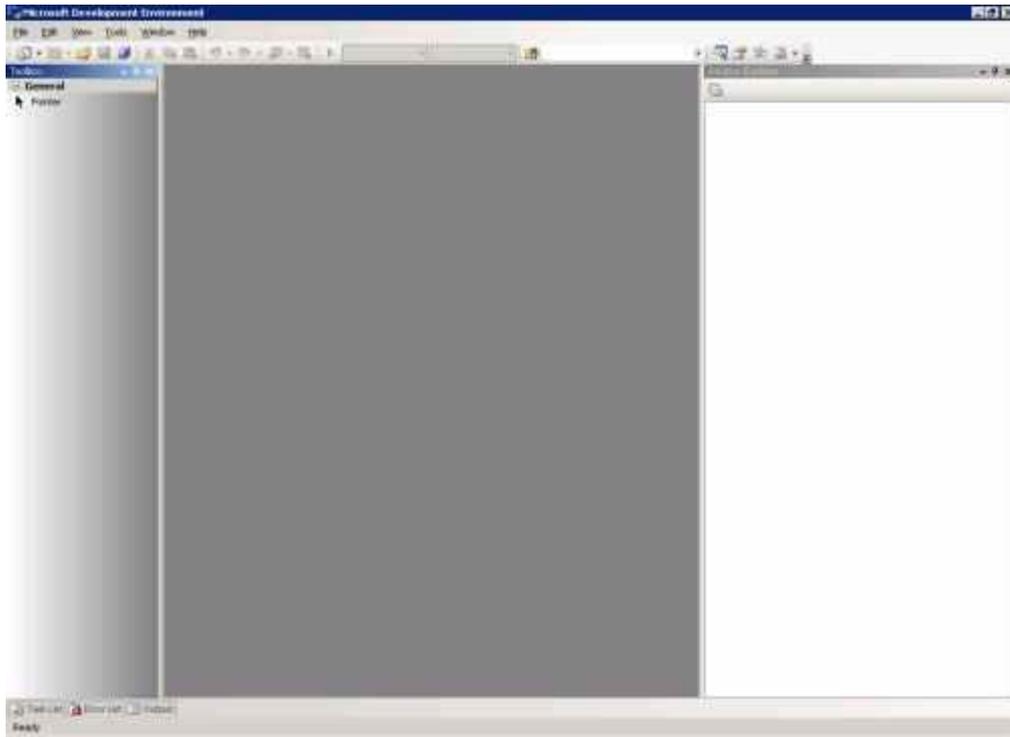


рис. 1 Business Intelligence Studio

Курс разделено на три части: подготовка базы данных SQL Server, подготовка базы данных Analysis Services и построение и работа с моделями Data Mining.

### **Подготовка базы данных SQL Server**

База данных AdventureWorksDW, которая используется в этом обзорном курсе, устанавливается вместе с SQL Server (по умолчанию эта опция отключена) и уже содержит представления (views), которые будут использоваться для создания моделей.

В дополнения к ним необходимо будет импортировать в базу данных таблицу prospective customers (в виде плоского файла). Один из сценариев, который мы рассмотрим, анализирует целевые рассылки с использованием построителя запросов для прогнозов.

### **Импорт таблицы Prospective Customers**

Вы будете использовать таблицу `prospective customer` в сценарии целевой рассылки для проверки качества прогнозирования модели. Вы можете скачать эту таблицу с сайта `Betaplace (www.Betaplace.com)`. Сначала необходимо импортировать плоский файл в базу данных `AdventureWorksDW` и назвать колонки этой таблицы так же как и в представлении, служащем источником данных для целевой рассылки.

В приложении A содержится сценарий импорта данных и переименования колонок. Загрузите плоский файл на ваш локальный диск. Запустите `SQL Management Studio` и откройте окно нового запроса к серверу БД `AdventureWorksDW`. Скопируйте SQL-запросы из приложения A в окно запросов. Измените значение переменной `@data_path` так, чтобы она содержала путь к файлу `"Prospect.csv"`. Выполните запрос.

Будет создана таблица **Prospects**, которая будет содержать ту же структуру столбцов, как и представление `vTargetedmail`, за исключением следующих столбцов:

- Age
- Bike Buyer
- Region

Так же, ключи клиентов (`CustomerKey`, `CustomerAlternateKey`) заменены на ключи перспективных клиентов (`ProspectKey`, `ProspectAlternateKey`).

### **Подготовка базы данных Analysis Services**

Прежде чем вы начнете создавать и работать с моделями `Data Mining`, вы должны выполнить следующие действия:

1. Создать новый проект `Analysis Services`.
  1. Создать новый источник данных (`data source`).

2. Создать новое представление данных (data source view).

## Создание проекта Analysis Services

Каждый проект Analysis Services определяет схему объектов для единственной базы данных Analysis Services, которая описывается моделями Data Mining, OLAP-кубами и дополнительными объектами.

1. Откройте Business Intelligence Development Studio.
2. Выберите **New** и **Project** из меню **File**.
3. Выберите проект Analysis Services в качестве типа нового проекта и назовите его **AdventureWorks**.
4. Нажмите **Ok**.

Новый проект откроется в Business Intelligence Development Studio.

## Создание источника данных

Источник данных - это данные о соединении с внешним источником данных, которые хранятся в проекте и в базе данных Analysis Services. В источнике данных хранится информация об имени сервера, базы данных, где хранится информация, используемая в проекте, а так же другие свойства соединения.

1. Правой кнопкой мыши щелкните по узлу **Data Source** вашего решения и выберите **New Data Source**.
2. На первой странице мастера выберите **Next**.
3. Щелкните **New** чтобы добавить соединение к базе данных **AdventureWorksDW**.
4. В диалоговом окне **Connection Manager** выберите провайдер данных **Microsoft OLE DB Provider for SQL Server**, имя сервера -

AdventureWorksDW, введите данные о пользователе, под учетной записью которого будет производиться соединение.

5. Щелкните **ОК**.
6. Щелкните **Next**.

По умолчанию источник данных будет назван **Adventure Works DW**.  
Щелкните **Finish**

Новый источник данных **Adventure Works DW** появится в папке источников данных в дереве решения.

### **Создание представления данных**

Представление данных (Data Source View) обеспечивает уровень абстракции для источника данных, позволяя вам менять структуру исходных данных для удобства вашего проекта. Используя представление данных, вы можете выбрать только относящиеся к вашему проекту таблицы и представления, определить связи между таблицами, добавить вычисляемые поля, а также именованные запросы без необходимости вносить модификации в исходные данные.

1. В дереве решения щелкните правой кнопкой мыши на **Data Source View** и выберите **New Data Source View**.
2. На первой странице щелкните **Next**.
3. Выберите источник данных **Adventure Works DW**, который вы создали на прошлом шаге в окне **Relational data sources**. Нажмите **Next**.
4. Если вы хотите создать новый источник данных, нажмите **New Data Source**.
5. Выберите таблицы из списка нажмите на правую стрелку для включения их в представление данных:
  - Prospect

- vAssocSeqLineItems
- vAssocSeqOrders
- vTargetMail
- vTimeSeries

6. Нажмите Next.

7. По умолчанию представление данных будет названо Adventure Works DW. Нажмите **Finish**.

Откроется окно представления данных как показано на рис. 2. Далее можно производить изменения в представлении источника данных.

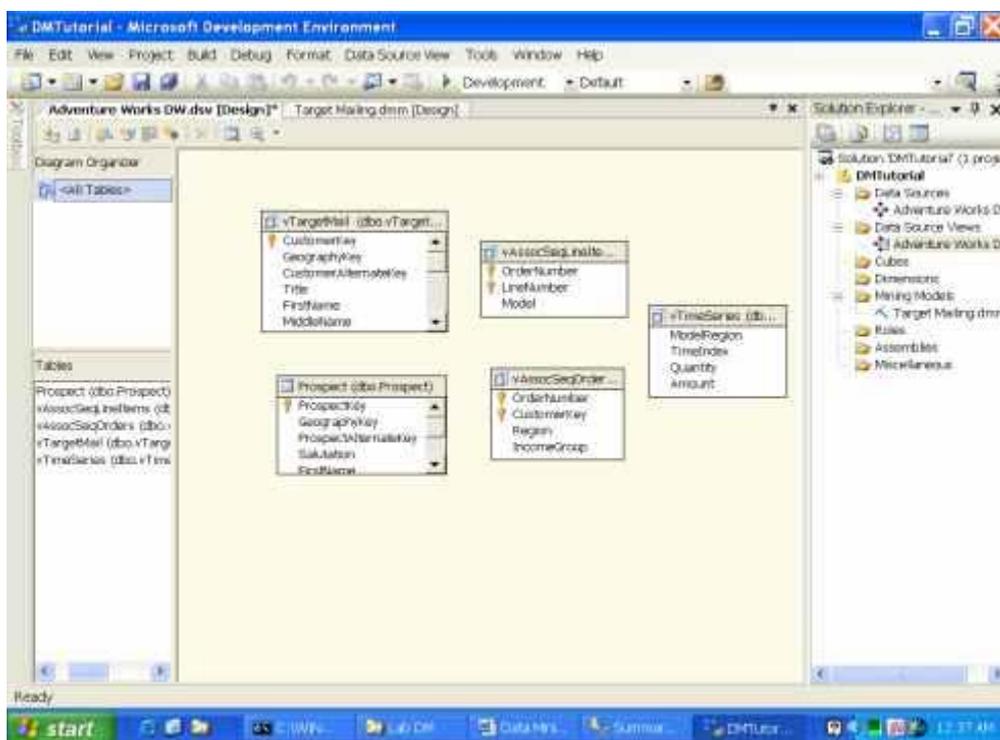


Рис. 2 Окно просмотра модели базы Adventure Works DW

### Редактирование представления источника данных

Используя редактор представления источника данных (Data Source View Editor), вы можете вносить изменения в представление данных. Например, вы можете переименовать объект, если его новое название будет больше

подходить к модели. При этом оригинальное название не изменится, но появится возможность обращаться к объекту через новое более удобное имя.

Для создания сценариев анализа рыночной корзины или кластеризации последовательностей необходимо создать новое соединение многие-к-одному между `vAssocSeqOrders` и `vAssocSeqLineItems`. Используя эту связь можно сделать `vAssocSeqLineItems` вложенной таблицей по отношению к `vAssocSeqOrders`, что необходимо для создания модели.

1. В окне просмотра выбрать поле `OrderNumber` из таблицы `vAssocSeqLineItems`
2. Перенести выбранную колонку в таблицу `vAssocSeqOrders` и поместить ее на колонку `OrderNumber`

Новая связь многие-к-одному между `vAssocSeqOrders` и `vAssocSeqLineItems` создана.

## **Построение и работа с моделями Data Mining**

Редактор Data Mining (показан на рисунке 4) содержит все инструменты и средства отображения для построения и работы с моделями Data Mining. Дополнительная информация по редактору доступна в разделе "Using the Data Mining Tools" в SQL Server Books Online.

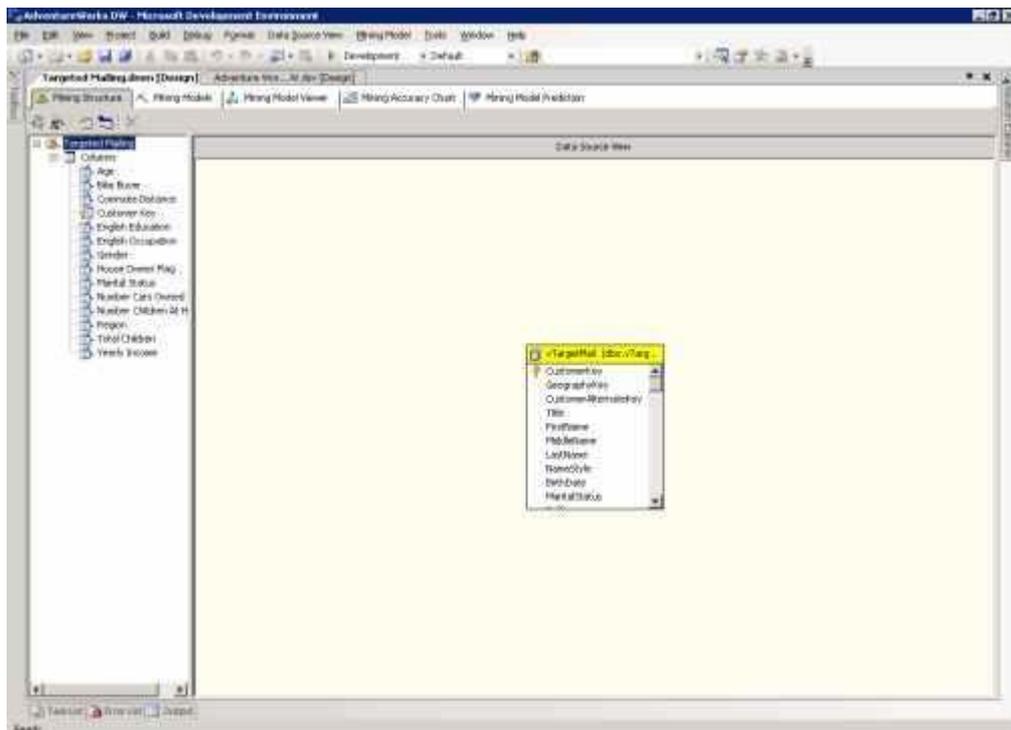


Рисунок 4 Редактор Data Mining

На протяжении этого руководства вы будете работать со следующими сценариями:

- Целевая отправка писем (Targeted mailing)
- Прогнозирование (Forecasting)
- Рыночная корзина (Market basket)
- Кластеризация последовательностей (Sequence clustering)

В сценарии Целевая отправка писем вы будете заниматься построением моделей, сравнением их ожидаемых возможностей (при помощи окна Mining Accuracy Chart), а также созданием прогнозов с использованием Prediction Query Builder.

## Адресная рассылка

Отдел маркетинга компании Adventure Works заинтересован в увеличении продаж при помощи проведения почтовой кампании, нацеленной на определённых покупателей. Исследуя их характеристики, они хотят обнаружить некоторый шаблон, применимый к потенциальным клиентам, который мог бы впоследствии использоваться для определения наиболее вероятных покупателей.

Кроме того, отдел маркетинга намеревается выявить логические группы среди уже существующих в их базе данных клиентов. Например, группа может содержать покупателей, объединённых по демографическому признаку, обладающих схожим набором покупок.

Adventure Works располагает списком как бывших, так и потенциальных клиентов.

В процессе решения этой задачи, отделу маркетинга потребуется выполнить следующие действия:

- Установить модели, способные выявить наиболее вероятных клиентов из списка потенциальных покупателей
- Кластеризовать своих существующих клиентов

Для выполнения этого сценария вам потребуется использовать Microsoft Naïve Bayes, Microsoft Decision Trees, и Microsoft Clustering алгоритмы. Сценарий содержит в себе 5 задач:

### **Описание структуры модели Data Mining**

Первый шаг заключается в создании новой структуры Data mining модели при помощи Mining Model Wizard. При этом также создаётся начальная модель, основанная на алгоритме Microsoft Decision Trees.

Для того чтобы создать структуру модели:

1. В Solution Explorer кликните правой кнопкой мыши на **Mining Models**, выберите **New Mining Model**. Откроется Mining Model Wizard.
2. Нажмите **Next** на странице приветствия.
3. Выберите **From existing relational database or data warehouse**, затем **Next**.
4. В разделе **Which data mining technique do you want to use?** выберите **Microsoft Decision Trees**. Вы создадите несколько моделей, опирающихся на эту начальную структуру, в основе которой лежит алгоритм Microsoft Decision Trees.
5. Нажмите **Next**. По умолчанию в качестве источника данных здесь выбран Adventure Works DW.
6. Нажмите **Next**.
7. Пометьте флагом **Case** таблицу **vTargetMail**, затем нажмите **Next**.
8. Пометьте флагом **Key** столбец **CustomerKey**. Если в исходной таблице присутствует ключевое поле, Mining Model Wizard автоматически выберет его в качестве ключевого для модели.
9. Пометьте флагом **Input** и **Predictable** столбец **BikeBuyer**. В результате столбец становится доступным для прогнозирования в новых наборах данных. После того как вы помечаете столбец подобным образом, становится доступной кнопка Suggest. Нажатие на эту кнопку вызывает диалоговое окно со списком наиболее связанных с данным полем столбцов, расположенных в порядке корреляции с прогнозируемым. Столбцы со значением больше 0.05 автоматически добавляются в модель. Если вы согласны с предлагаемым вариантом, нажмите кнопку **ОК** или определите нужные поля вручную.
10. Пометьте флагом **Input** столбцы, перечисленные в таблице.

**Age**

**YearlyIncome**

**Region**

|                   |                      |               |
|-------------------|----------------------|---------------|
| CommuteDistance   | HouseOwnerFlag       | TotalChildren |
| EnglishEducation  | LastName             |               |
| EnglishOccupation | MaritalStatus        |               |
| FirstName         | NumberCarsOwned      |               |
| Gender            | NumberChildrenAtHome |               |

Вы можете выбрать несколько столбцов, удерживая клавишу SHIFT.

11. Нажмите Next.

12. В поле Specify Columns' Content and Data Type нажмите Detect. При этом автоматически определяется, содержат ли столбцы с численными данными дискретные, либо непрерывные значения. Например, столбец может содержать информацию о заработной плате с непосредственными значениями сумм (непрерывные значения) или в нём могут содержаться целые числа, отвечающие неким диапазонам значений (1 = <\$25,000; 2 = от \$25,000 до \$50000 и так далее), то есть дискретные значения.

13. В поле Structure Name введите Targeted Mailing и нажмите Finish

14. Нажмите Finish. При этом откроется редактор Data mining, отображающий структуру Targeted Mailing, которую вы только что создали, см. Рисунок 5.

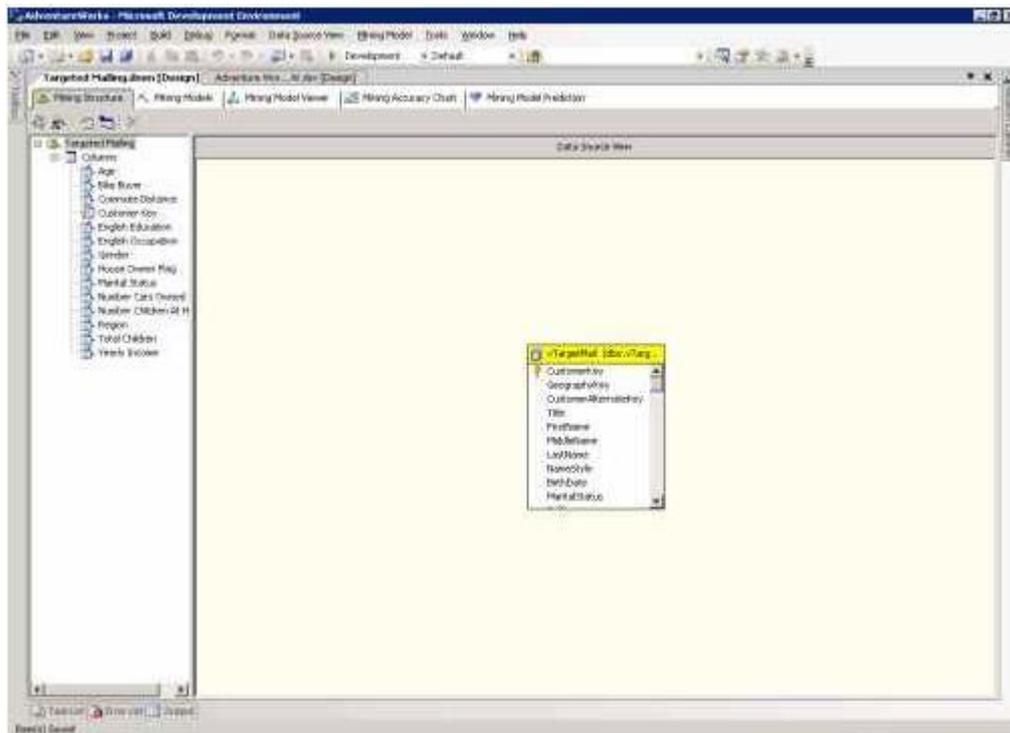


Рисунок 5 Страница структуры Targeted Mailing

## Редактирование Data Mining моделей

Начальная Data Mining структура содержит единственную модель, основанную на Microsoft Decision Trees. В этой части вы, при помощи страницы Mining Models редактора определите две дополнительные модели: модели Microsoft Na?ve Bayes и Microsoft Clustering.

Чтобы создать модель Microsoft Clustering

1. Нажмите на закладку **Mining Models**.
2. Кликните правой кнопкой по Targeted Mailing и выберите **New Mining Model**.
3. В поле **Model Name** введите *TM\_Clustering*.
4. В поле **Algorithm Name** выберите **Microsoft Clustering**.
5. Нажмите ОК.

Новая модель появится на странице Mining Models. Модель Microsoft Clustering позволяет группировать и прогнозировать непрерывные и дис-

кретные атрибуты. Вы можете изменить используемые столбцы и свойства для новой модели.

Установка столбца как Predict (прогнозируемого) не оказывает никакого эффекта на этапе обучения модели; но позволяет вам указать этот столбец в запросе на прогнозирование PREDICTION JOIN. При создании кластеров алгоритм игнорирует столбцы, помеченные как **PredictOnly**. Статистика для **PredictOnly** столбцов в кластерной модели определяется на финальном шаге, после того как операция кластеризации завершена. Это имеет смысл, если вы хотите видеть распределение атрибута по кластерам, созданным на основании других атрибутов и позволяет выявить более глубокие взаимосвязи.

Чтобы создать модель Microsoft Na?ve Bayes

1. Кликните правой кнопкой по Targeted Mailing и выберите New Mining Model.
2. В поле Model Name введите TM\_NaiveBayes.
3. В поле Algorithm Name выберите Microsoft Na?ve Bayes. После чего появится диалоговое окно с текстом, объясняющим, что алгоритм Microsoft Na?ve Bayes не поддерживает работу с столбцами Age, Geography Key и Yearly Income, содержащими непрерывные значения, и они будут проигнорированы.
4. Нажмите **Yes**.
5. Нажмите **ОК**.
6. После чего, новая модель появится на странице Mining Models. Хотя вы можете менять используемые столбцы и свойства для всех моделей на этой странице, в данном случае оставьте всё без изменений.

**Обработка Data Mining моделей**

Теперь, после того как структура и параметры моделей определены, вы можете осуществить развёртывание и обработку моделей.

Для того чтобы развернуть проект и обработать модели

- Нажмите F5.

Теперь база данных Analysis Services выложена на сервер и модели обработаны. Если база данных уже была на сервере, вы можете осуществить только обработку моделей, следуя следующему алгоритму.

Для обработки data mining моделей

- В меню Mining Model выберите пункт Process All Models. Откроется диалоговое окно Process Targeted Mailing.
- Нажмите Run.
- Откроется диалоговое окно Process Progress, отображающее информацию об обработке модели. В зависимости от производительности вашего компьютера, процесс может занять некоторое время.
- После того как завершится обработка, закройте оба диалоговых окна.

## Исследование моделей

После того как модели обработаны, вы можете просмотреть их, используя закладку Mining Model Viewer в редакторе Data Mining. При помощи выпадающего списка **Mining Model** в верхней части закладки можно исследовать модели, входящие в структуру.

### Модель Microsoft Decision Trees

Mining Model Viewer по умолчанию открывает модель Targeted Mailing первой в структуре. Tree viewer содержит две закладки, **Decision Tree** (дерево решений) и **Dependency Network** (сеть зависимостей).

## Decision Tree (Дерево решений)

На странице **Decision Tree** вы можете просмотреть все модели деревьев, составляющие модель Targeted\_Mailing. Для каждого прогнозируемого атрибута модели существует одно дерево, если только не задействовано feature selection. Так как ваша модель содержит всего один прогнозируемый атрибут, Bike Buyer, то и дерево тоже одно. Если бы их было несколько, можно было бы выбрать нужное при помощи Tree box.

Tree viewer по умолчанию показывает только первые три уровня дерева. Чтобы увидеть больше используйте Show Level slider или Default Expansion box. Для более подробной информации по настройке Tree viewer, обратитесь к разделу "Viewing with Tree Viewer" в SQL Server Books Online.

Чтобы создать дерево, показанное на рисунке 6

1. Установите Show Level 5

2. В списке Background нажмите 1.

Здесь задаётся состояние прогнозируемого атрибута. Изменяя это свойство, вы можете быстро узнать число случаев попадания покупателя велосипеда в каждый узел. Чем темнее узел, тем больше случаев попадания он содержит.

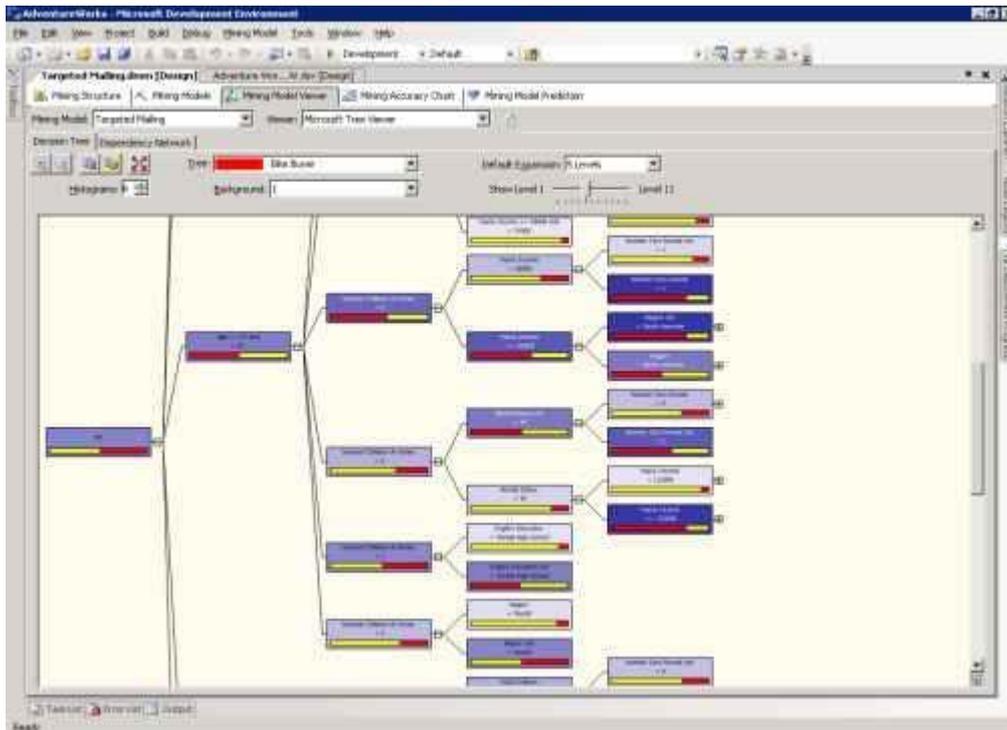


Рисунок 6 Страница Decision Tree для модели Targeted Mailing

Каждый узел в дереве решений несёт в себе следующую информацию:

- Условие, по которому попадаем в этот узел из предыдущего. Полный путь к узлу можно узнать из легенды или из подсказки.
- Гистограмма, описывающая распределение состояний для прогнозируемого столбца. Число состояний, отображаемых на гистограмме, регулируется при помощи Histogram control.
- Концентрацию случаев попадания, если состояние предсказываемого значения определено в контроле Background.

Если включен режим drillthrough ("проваливание" в исходные данные), вы можете просмотреть варианты для узла, кликнув по нему правой кнопкой мыши и выбрав Drillthrough.

### Сеть зависимостей (Dependency Network)

На странице Dependency Network показаны отношения между атрибутами, влияющими на прогнозирующую способность модели. Сеть зависимостей для модели Targeted Mailing представлена на рисунке 7.

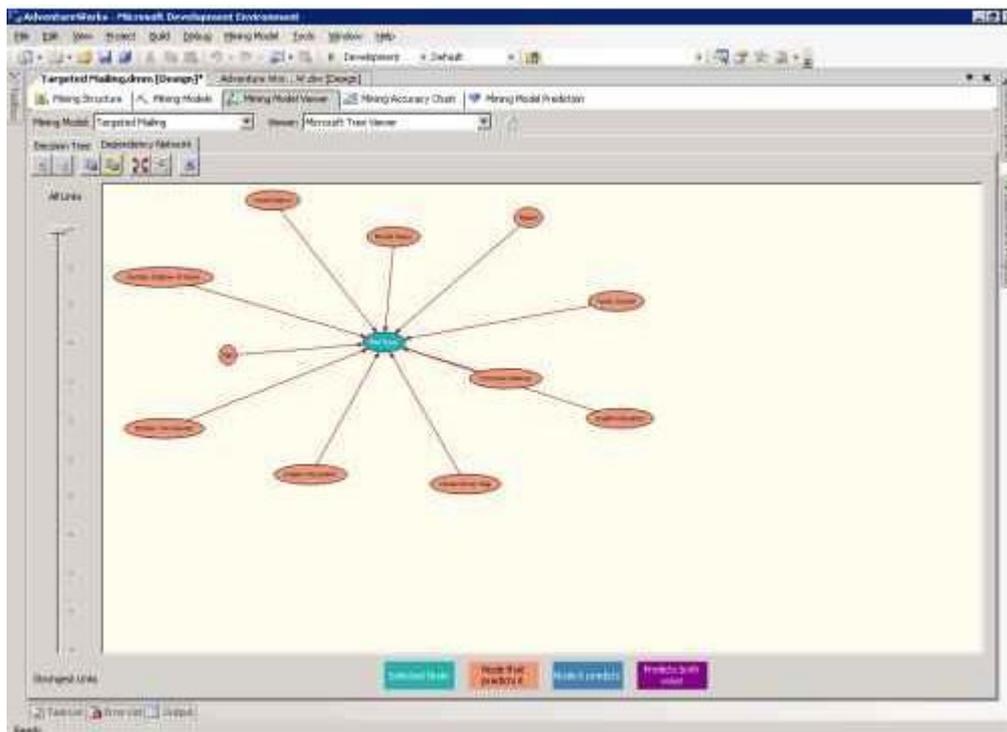


Рисунок 7 Страница Dependency Network модели Targeted Mailing.

Центральный узел на рисунке 7, Vike Buyer, представляет прогнозируемый атрибут модели. Каждый из узлов вокруг него отображает атрибут, влияющий на прогнозируемый. При помощи ползунка на левой стороне страницы можно указать силу показываемых связей. При перемещении ползунка вниз будут показываться только сильные связи.

По легенде цветов внизу окна можно узнать, какой узел является прогнозируемым и какие узлы его определяют.

## Модель Microsoft Clustering

Используйте combo box Mining Model для перехода к модели TM\_Clustering. Окно этой модели, Cluster viewer, содержит четыре закладки: Cluster Diagram, Cluster Profiles, Cluster Characteristics, и Cluster Discrimination. По умолчанию при первом открытии отображается страница Cluster Diagram.

Для более подробной информации по использованию Cluster viewer, обратитесь к разделу " Viewing with Cluster Viewer " в SQL Server Books Online.

### **Кластерная диаграмма (Cluster Diagram)**

На странице Cluster Diagram вы можете изучить отношения между кластерами, обнаруженными алгоритмом. Длина линий, соединяющих кластеры, отображает "близость", а их интенсивность показывает, насколько они схожи. Цвет кластера отвечает за частоту появления переменной и состояния (выбранные в полях Shading Variable и State вверху экрана) в каждом кластере. Переменная по умолчанию population, но вы можете поменять её на любую другую из модели и найти нужные вам кластеры. Используя полосу прокрутки в левой части экрана, вы можете отфильтровать слабые связи и найти кластеры, расположенные наиболее близко к друг другу.

Например, установите поле Bike Buyer в качестве Shading Variable и State в 1. Как показано на рисунке 8, кластер №5 содержит наибольшее число покупателей велосипедов. А самая сильная связь - между 4 и 7 кластерами.

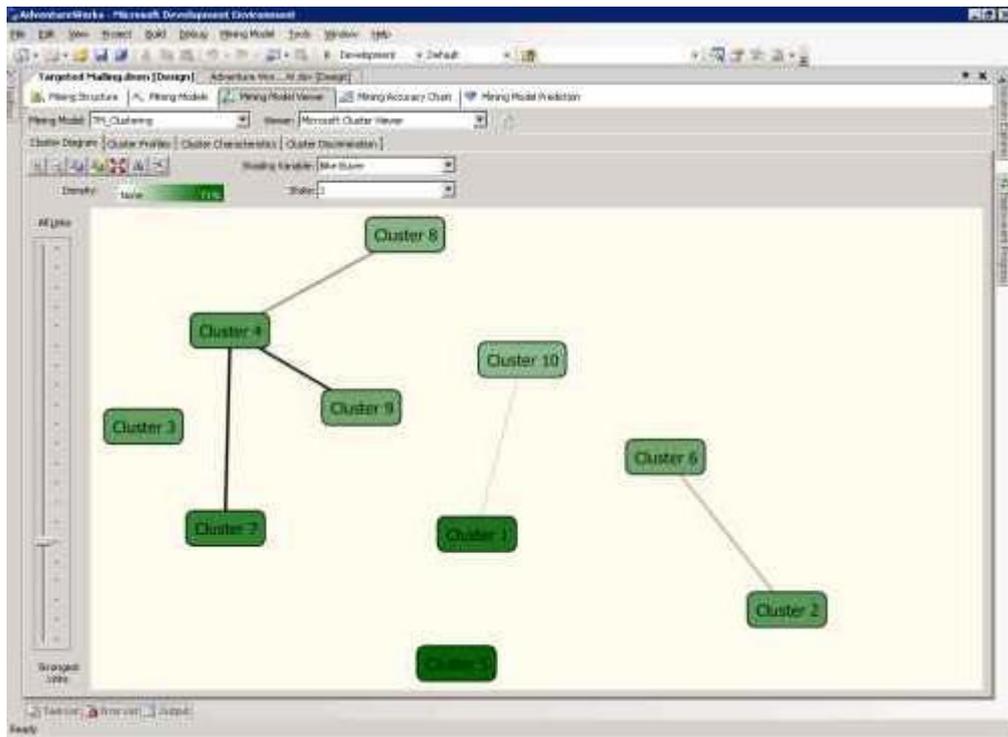


Рисунок 8 Кластерная диаграмма модели TM\_Clustering.

### Профили кластеров (Cluster Profiles)

Страница Cluster Profiles предоставляет общий вид модели TM\_Clustering. Как показано на рисунке 9, каждому кластеру модели на этой странице соответствует столбец с описывающими его данными. В первом столбце находится список атрибутов, с которыми связан по крайней мере один кластер. Все оставшиеся столбцы отображают распределение состояний этих атрибутов для каждого кластера. Распределение дискретной переменной показывается цветным столбцом, максимальное число отображаемых цветов-состояний определяется в поле Bars per histogram. Непрерывные значения отображаются diamond диаграммой, показывающей среднее и стандартное отклонение в каждом кластере.

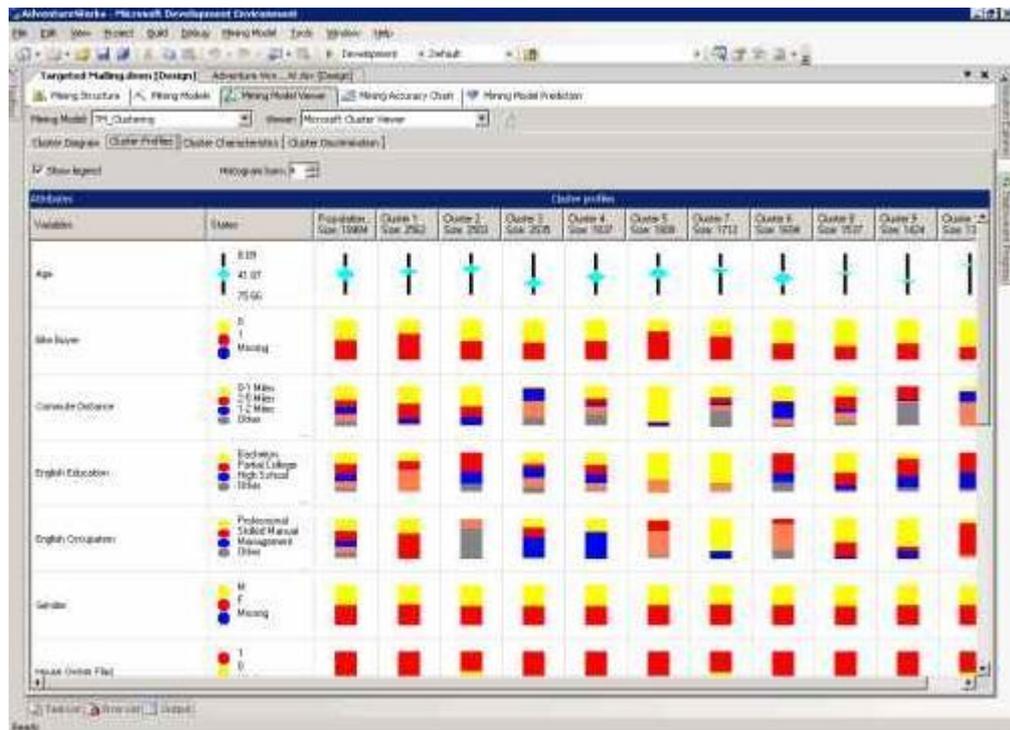


Рисунок 9 Страница Cluster Profiles модели TM\_Clustering.

### Характеристики кластеров (Cluster Characteristics)

Страница Cluster Characteristics позволяет более детально рассмотреть характеристики кластера. Для примера, на рисунке 10, вы можете видеть, что люди из Кластера №5 (покупатели велосипедов) обладают следующими характеристиками: они передвигаются на короткие дистанции (0-1 миль), у них нет машины и они женаты.

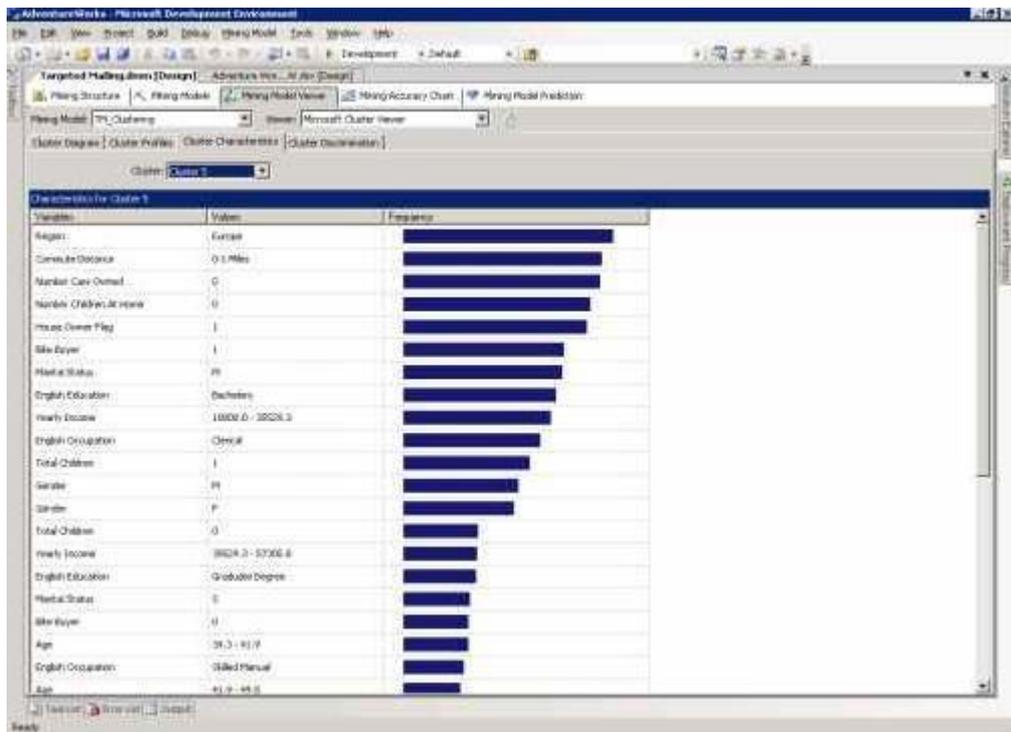


Рисунок 10 Окно характеристик кластера модели TM\_Clustering.

## Cluster Discrimination

На странице Cluster Discrimination вы можете посмотреть характеристики, отличающие один кластер от другого. После выбора кластеров (поля First cluster и Second cluster) система определяет различия и показывает их, сортируя по атрибутам, наиболее сильно отличающим кластеры.

На рисунке 11 приведено сравнение Кластера №5 и Кластера №7 модели TM\_Clustering. В кластере №5 наиболее высокая плотность покупателей велосипедов, а в №10, напротив, самая низкая. Видно, к примеру, что люди, попавшие в Кластер №10 скорее всего из Северной Америки и моложе (23-31), чем жители Европы из Кластера №5.

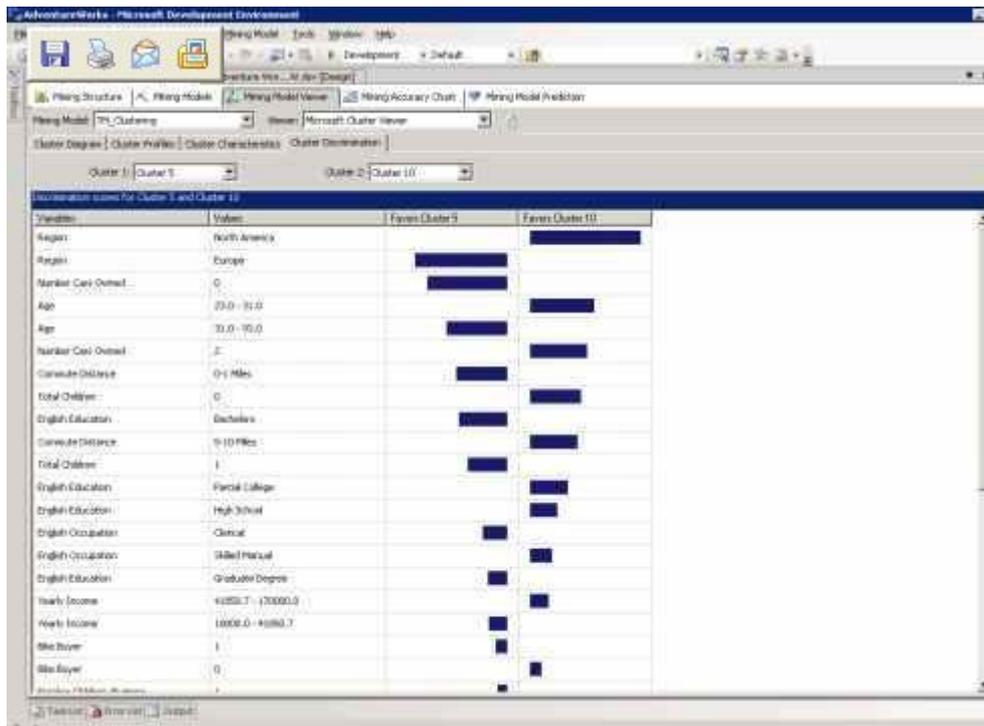


Рисунок 11 Страница Cluster Discrimination модели TM\_Clustering.

## Модель Microsoft Na?ve Bayes

Используйте выпадающий список Mining Model для перехода к модели TM\_NaiveBayes. Окно этой модели, Na?ve Bayes viewer, содержит четыре закладки: Dependency Network, Attribute Profiles, Attribute Characteristics, и Attribute Discrimination.

Для более подробной информации по использованию Na?ve Bayes viewer, обратитесь к разделу " Viewing with Na?ve Bayes viewer " в SQL Server Books Online.

### Сеть зависимостей (Dependency Network)

Страница Dependency Network в Na?ve Bayes viewer организована также как одноимённая страница в Tree viewer. Каждый узел представляет атрибут, а линии, связывающие узлы соответствуют отношениям. На рисунке 12

показаны все атрибуты, влияющие на состояние прогнозируемого атрибута, Bike Buyer.

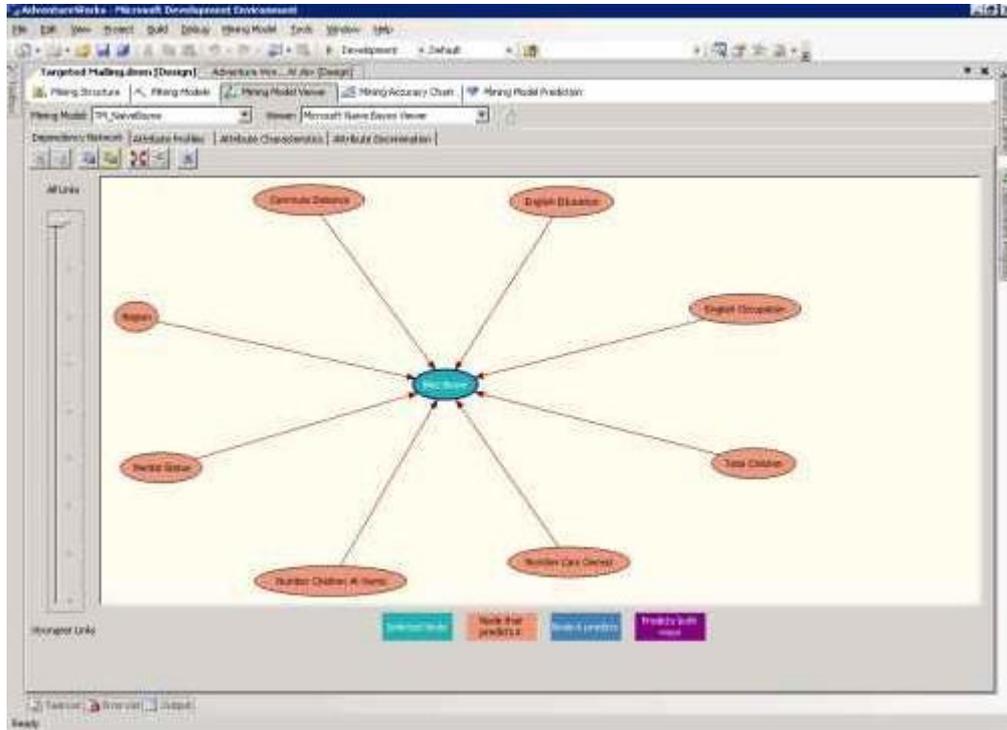


Рисунок 12 Страница Dependency Network для модели TM\_Na?veBayes.

По мере передвижении ползунка вниз, остаются атрибуты, наиболее сильно влияющие столбец Bike Buyer (являющийся прогнозируемым). Действуя таким образом, вы обнаружите, что самым сильным фактором, определяющим, купит ли человек велосипед, является число принадлежащих ему машин.

### Профили атрибута (Attribute Profiles)

Страница Attribute Profiles описывает как различные состояния входных атрибутов влияют на значение прогнозируемого атрибута.

В поле Predictable выберите Bike Buyer. Атрибуты, влияющие на состояние прогнозируемого атрибута, перечислены вместе со значениями каждого состояния входных атрибутов и их распределениями для каждого состояния прогнозируемого атрибута.

В поле Predictable выберите Bike Buyer. Атрибуты, влияющие на состояние прогнозируемого атрибута, перечислены для всех значений входных параметров совместно с их распределением для каждого состояния прогнозируемого атрибута.

На рисунке 13 показана страница Attribute Profiles для столбца Bike Buyer для модели TM\_NaiveBayes model.

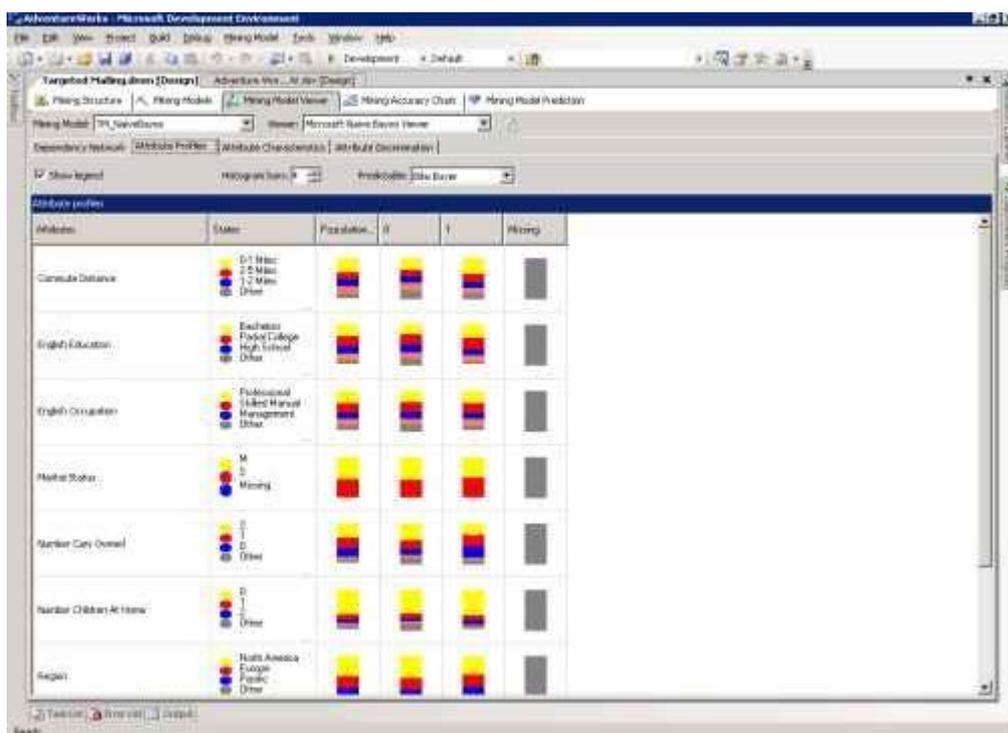


Рисунок 13 Страница Attribute Profiles для модели TM\_NaiveBayes.

### Характеристики атрибута (Attribute Characteristics)

Используя страницу Attribute Characteristics, вы можете выбрать атрибут и его значение, чтобы увидеть, значения каких атрибутов меняются при этом в большей или меньшей степени.

В поле атрибут выберите Bike Buyer, в поле значение - 1.

Например, из рисунка 14 видно, что люди, не имеющие детей, покупают наибольшее количество велосипедов.

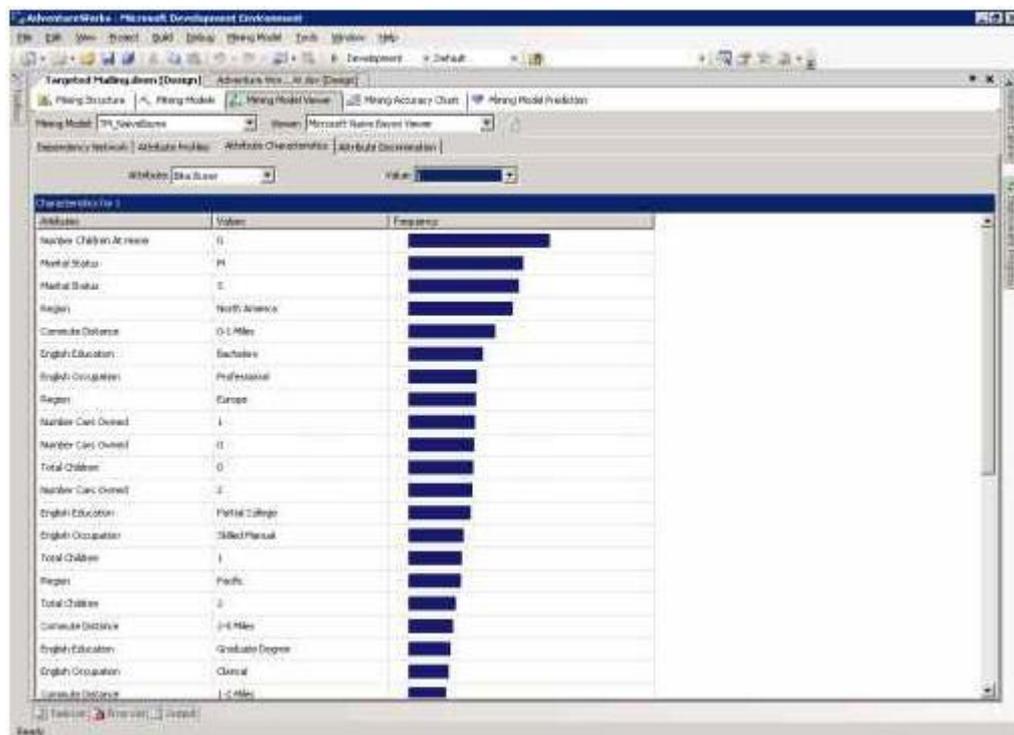


Рисунок 14 Страница Attribute Characteristics для модели TM\_Na?veBayes.

### Attribute Discrimination

На странице Attribute Discrimination вы можете исследовать отношение между двумя дискретными значениями выбранного прогнозируемого атрибута и значениями других атрибутов.

Поскольку модель TM\_NaiveBayes имеет всего два состояния, 1 и 0, вам не придётся совершать каких-либо дополнительных действий по настройке окна.

К примеру, из рисунка 15 видно, что те, у кого нет машины более склонны к покупке велосипеда, а люди, имеющие две машины скорее не будут этого делать.

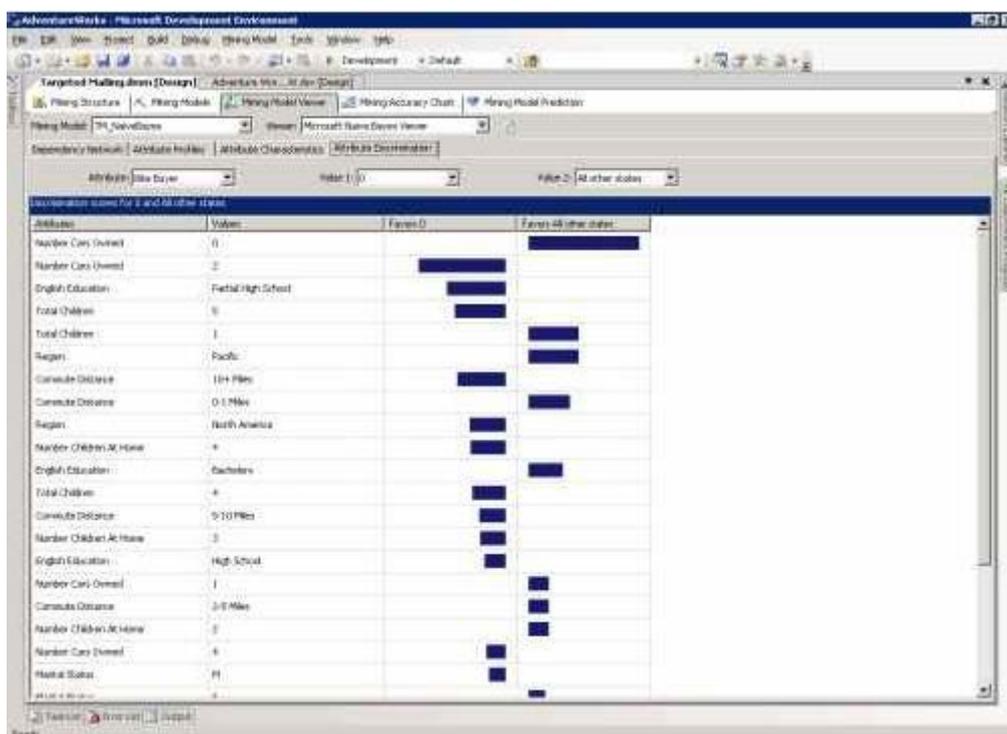


Рисунок 15 Страница Attribute Discrimination для модели TM\_NaiveBayes.

### Тестирование точности Data Mining-моделей

Теперь модели обработаны и исследованы. Но насколько хорошо они предсказывают ситуацию? Работает ли какая-нибудь из моделей лучше, чем другие?

Используя страницу Mining Accuracy Chart, вы можете вычислить, как хорошо каждая из моделей предсказывает результат и сравнить их между собой. Этот метод сравнения также иногда называется диаграммой роста (lift chart). На странице The Mining Accuracy Chart используются данные, отделённые от первоисточника, что позволяет сравнивать прогнозы с известными результатами. Затем результаты сортируются и отображаются на графике вместе с идеальной моделью, показывая, насколько хорошо каждая модель делает прогноз. График идеальной модели соответствует теоретической модели, предсказывающей результат со 100% точностью.

Диаграмма роста является важным инструментом, поскольку она помогает различать модели практически идентичные по структуре, показывая, какая делает лучший прогноз. Подобным образом, из её рассмотрения можно сделать вывод о том, какие типы алгоритмов делают лучшие предсказания в данной ситуации. Для более подробной информации по использованию Mining Accuracy Chart, обратитесь к разделу "Comparing Data Mining Models with the Lift Chart" в SQL Server Books Online.

Страница Mining Accuracy Chart показана на рисунке 16.

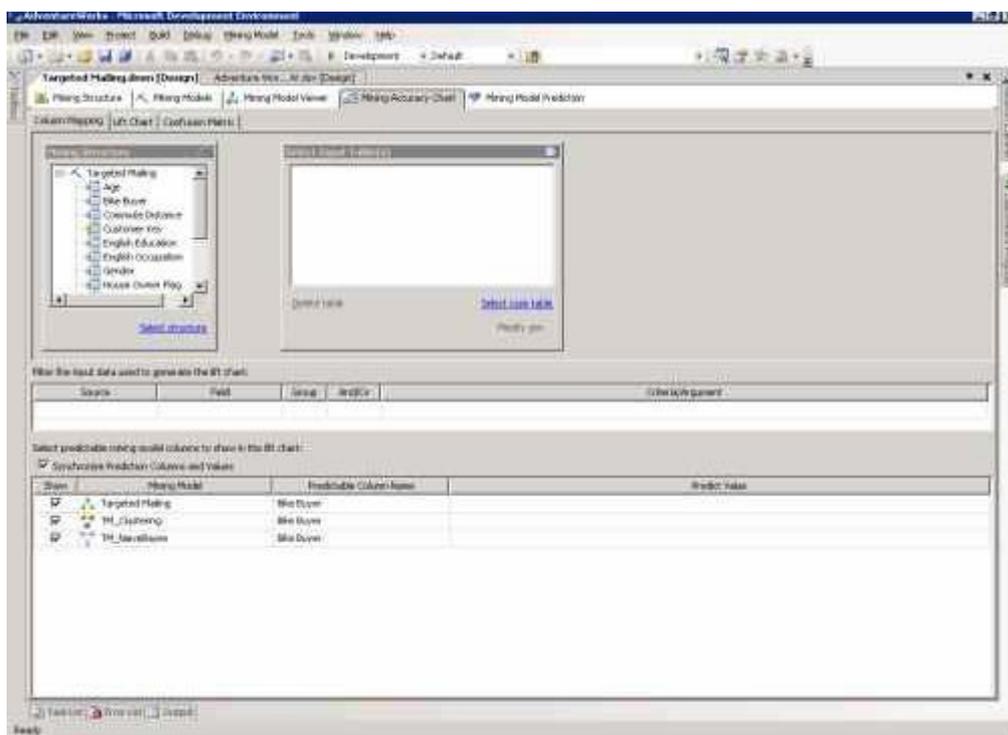


Рисунок 16 Страница Mining Accuracy Chart.

Для того чтобы создать новую диаграмму выполните следующие шаги:

1. Свяжите столбцы модели с соответствующими столбцами во входных данных.
2. Установите фильтр на входные данные.
3. Выберите сравниваемые модели, а также поля для которых делается прогнозирование и их значения.

**Примечание:** Для использования mining accuracy chart модели должны быть обработаны.

### Отображение входных столбцов

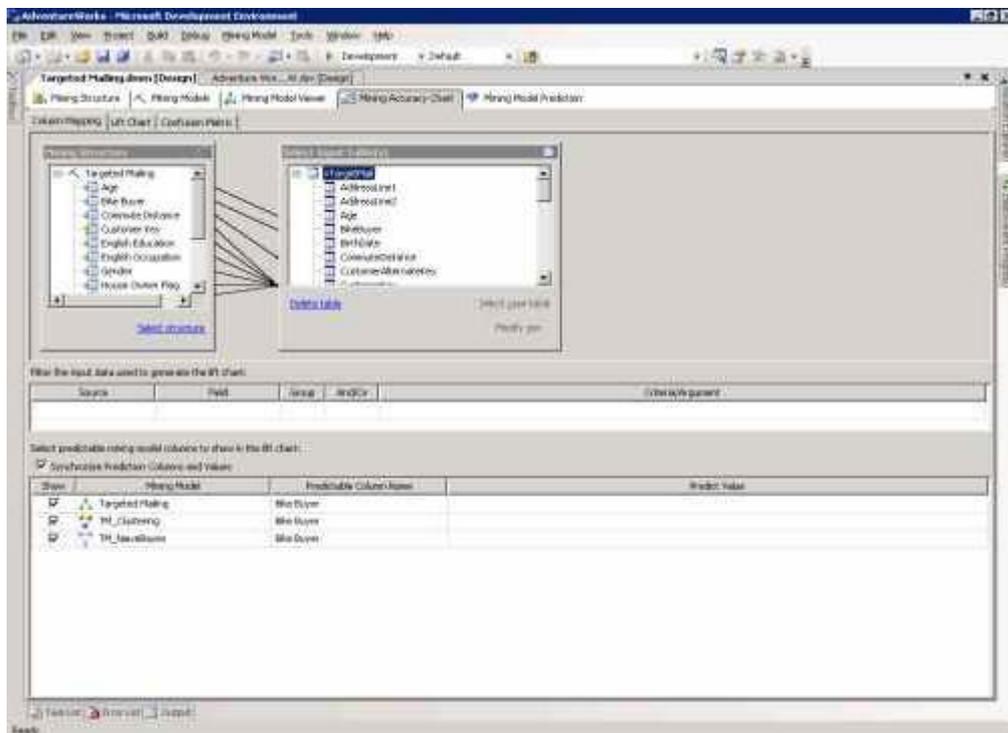
На первом шаге необходимо связать столбцы, используемые в модели со столбцами во входных данных. В том случае, если имена столбцов совпадают, связи создадутся автоматически.

Чтобы соотнести входные столбцы со структурой модели:

1. В поле Select Input Table(s) выберите Select case table. Откроется диалоговое окно выбора таблицы, где вы можете выбрать таблицу, содержащую данные, которые вы хотите использовать для прогнозирования и определения точности моделей. В этом руководстве используются те же данные, что и для обработки моделей.

2. В диалоговом окне Select table, раскройте базу AdventureWorks и выберите источник данных Adventure Works DW.

3. Нажмите **vTargetMail**. Столбцы data mining структуры автоматически будут соответствовать столбцам с теми же именами таблицы с исходными данными, как показано на рисунке 17.



## Рисунок 17 Связанные столбцы.

Запрос на прогноз генерируется для каждой модели Mining Structure на основании связанных столбцов. Вы можете удалить связь, выбрав линию, связывающую поля в Mining Structure и Select Input Table(s) и нажав DELETE. Также вы можете создать связь вручную, выбрав столбец в Select Input Table(s) и перетащив его на соответствующий столбец в Mining Structure.

### **Фильтрация входных столбцов**

Для установки фильтров на входные данные вы можете использовать таблицу Filter the input data used to generate the lift chart. Вы можете перетащить столбцы из Select Input Table(s) на таблицу с результатами или выбрать значения в выпадающих списках. Например, если необходимо учитывать только те строки в столбце Income, где его значение больше x, выберите Income в поле Field, а в поле Criteria/Argument введите >x.

Для выполнения данного упражнения вам не потребуется фильтровать данные.

### **Выбор моделей, прогнозируемых полей и их значений**

Следующий шаг заключается в выборе моделей, которые вы хотите включить в рассмотрение, и прогнозируемого поля, по которому они будут сравниваться. По умолчанию, выбраны все модели.

Вы можете создавать два типа диаграмм. Если указано значение прогнозируемого атрибута, вы увидите график вроде того, что показан на рисунке 18, из которого видно какая из моделей показывает лучший результат (lift chart). В том случае, если значение не указано, то диаграмма отобразит точность модели (см. рисунок 19).

Чтобы показать диаграмму lift chart:

1. Для каждой модели в поле **Predictable Column Name** выберите **Bike Buyer**.
2. Для каждой модели в поле **Predict Value** выберите 1.

Чтобы показать диаграмму точности моделей:

- В поле Predictable Column Name, выберите Bike Buyer. Оставьте поле Predict Value пустым.

Если установлен флаг Synchronize Prediction Columns and Values, прогнозируемый столбец синхронизируется для каждой модели в структуре Data Mining.

**Примечание:** Список столбцов модели Predictable Column Name, ограничен полями с типом Predict и Predict Only, содержащими дискретные или дискретизированные данные.

В некоторых более сложных случаях, вам может понадобиться создать диаграмму роста для двух моделей анализа, но в которых значения прогнозируемого атрибута берутся из разных столбцов структуры Data Mining. Если выключить флаг Synchronize Prediction Columns and Values, можно выбрать любой доступный прогнозируемый столбец и значение. Результаты отображаются вместе, вне зависимости от того как они получены.

### **Диаграмма роста**

Кликните на закладку **Lift Chart** чтобы просмотреть диаграмму роста. При выполнении этого действия на сервере выполняется запрос на прогнозирование. Полученные результаты прогноза сравниваются с известными значениями и сортируются по вероятности, после чего отображаются на графиче-

ке. Для более подробной информации по использованию диаграммы, обратитесь к разделу "Lift Chart" в SQL Server Books Online.

При условии указанного значения прогнозируемого столбца, диаграмма роста имеет вид показанный на рисунке 18.

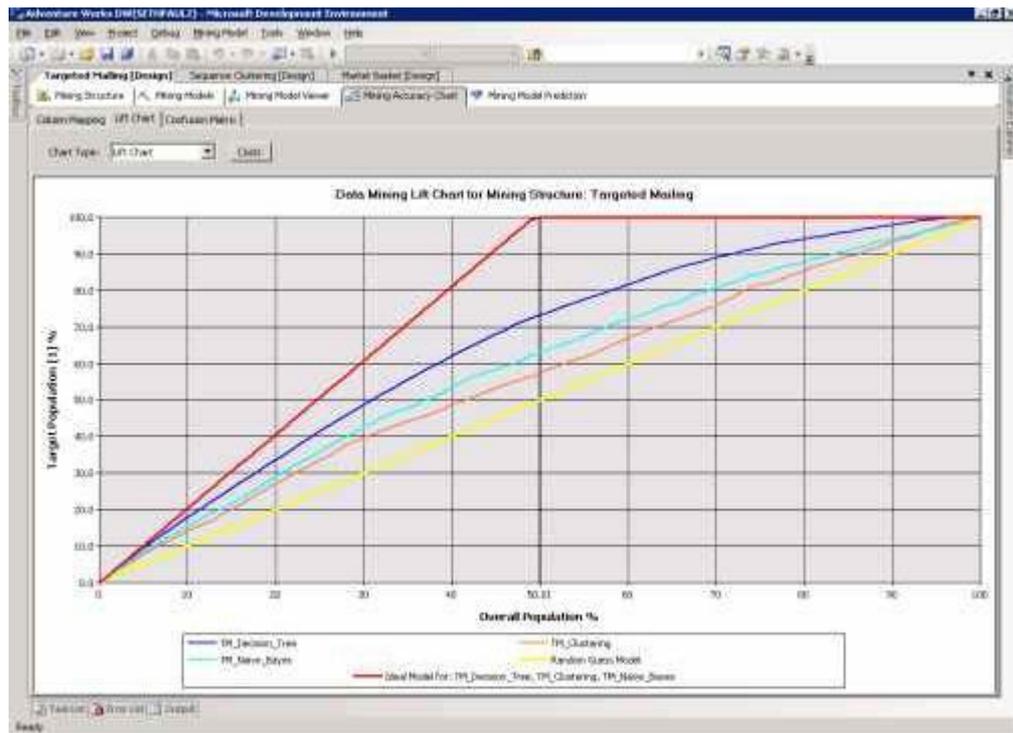


Рисунок 18 Рост для каждой модели в сравнении с идеальной моделью.

Если значение прогнозируемого атрибута не было указано, диаграмма роста показывает точность предсказаний моделей (см. рисунок 19).

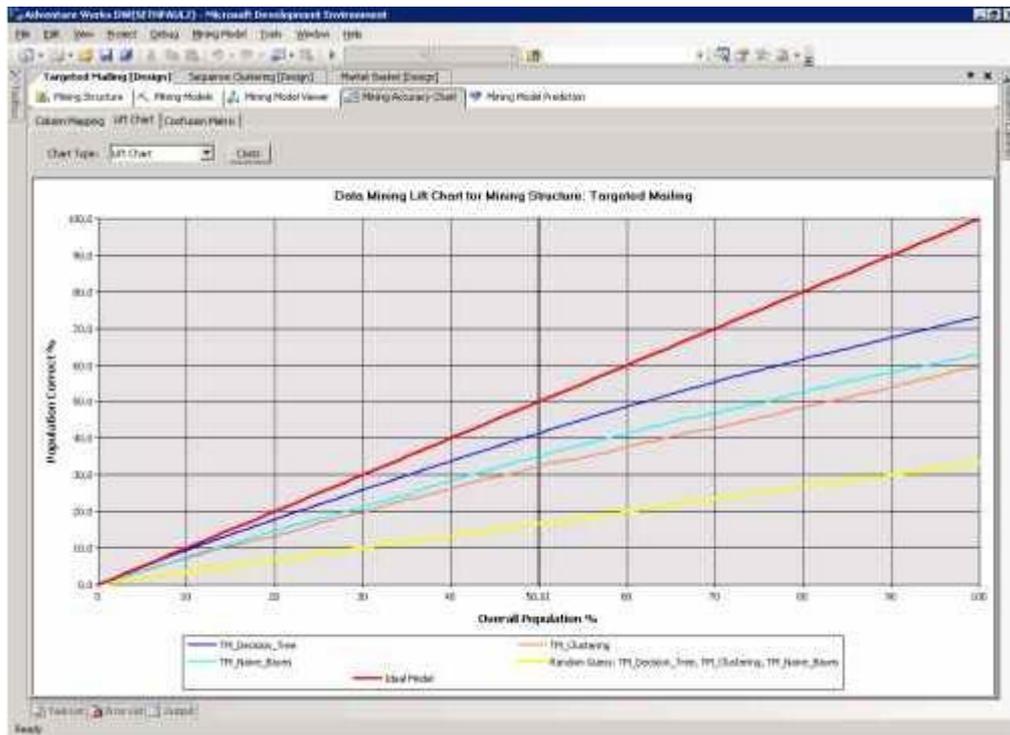


Рисунок 19 Точность каждой модели в сравнении с идеальной моделью.

## Создание прогнозов

Теперь, после того как вы закончили с моделями анализа, можете перейти к созданию DMX запросов, используя Prediction Query Builder. Prediction Query Builder аналогичен Access Query Builder, где вы можете использовать drag-and-drop для построения запросов. Инструмент содержит три окна:

- Design
- Query
- Result

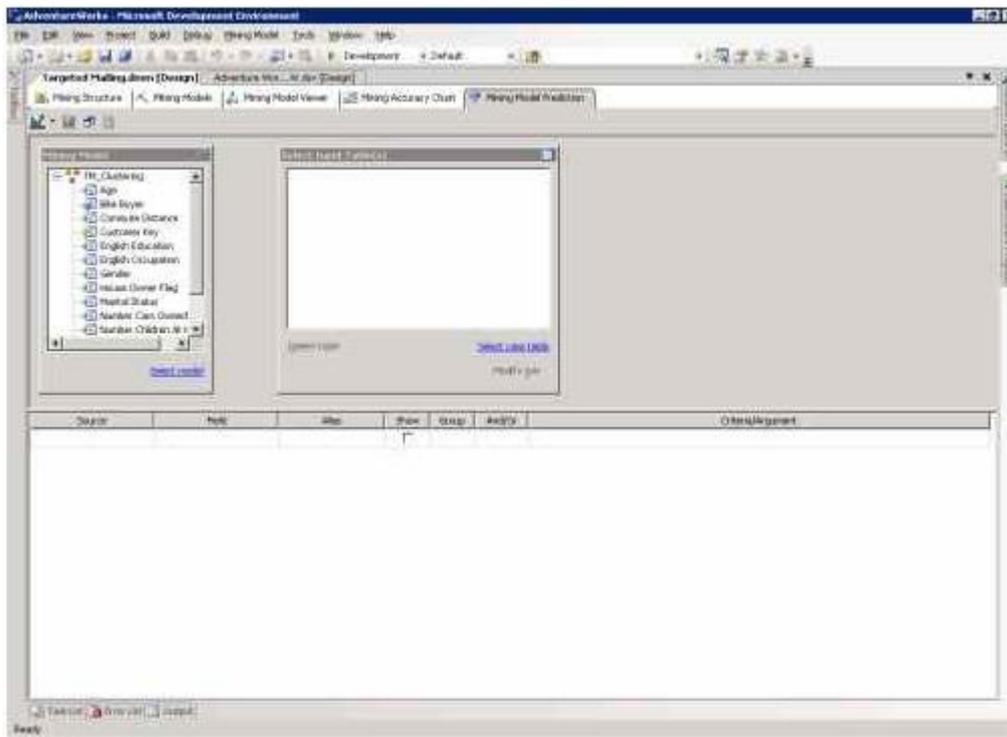


Рисунок 20 Prediction Query Builder - вид по умолчанию.

Используя окна Design и Query, вы можете создавать и просматривать запросы. Результаты выполнения отображаются на странице Result.

### Создание запросов

Первым шагом при создании запроса является выбор модели анализа и таблицы с исходными данными.

Чтобы выбрать модель и таблицу с исходными данными

1. В окне Mining Model нажмите Select model.

Откроется диалоговое окно Select Mining Model. По умолчанию выбрана первая модель в структуре.

2. Переместитесь по дереву к пункту Targeted Mailing и нажмите на него.

3. В поле Select Input Table(s) нажмите Select case table.

Откроется диалоговое окно Select Table.

4. Спуститесь по дереву к таблице Prospect из data source AdventureWorksDW.

После того, как вы указали таблицу с исходными данными, Prediction Query Builder создаст связи по умолчанию между моделью и таблицей, основанные на совпадении имён столбцов, как показано на рисунке 21.

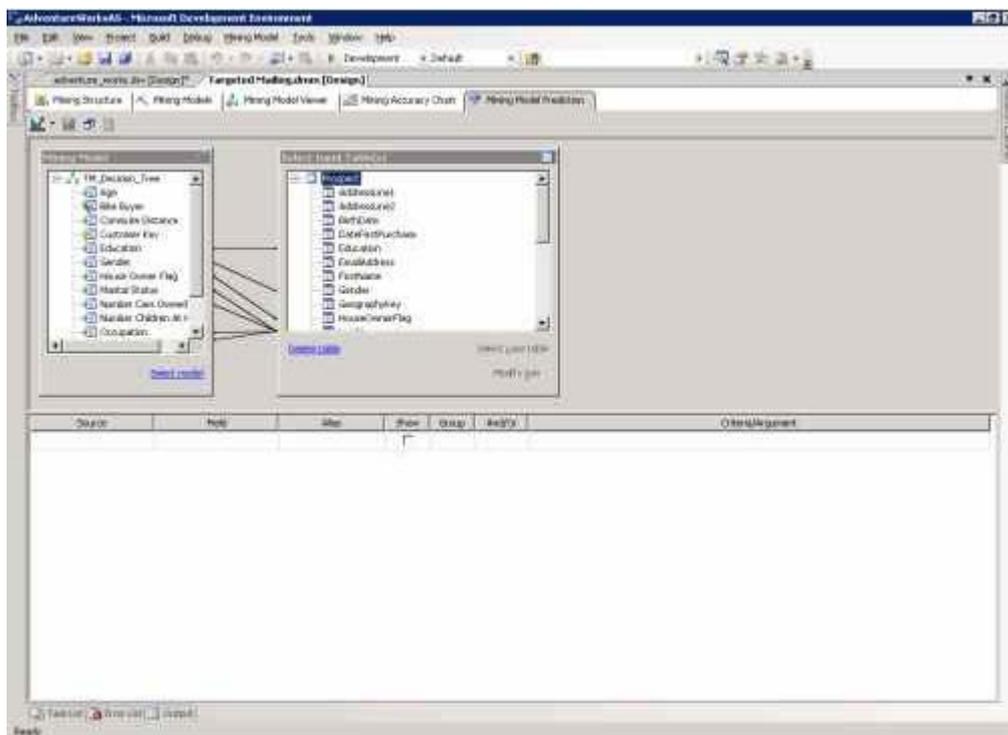


Рисунок 21 Соответствие столбцов на странице Mining Model Prediction.

Чтобы построить запрос на прогноз

1. В поле Source, нажмите на ячейку в первой пустой строке, затем кликните на Prospect table.

2. В поле Field, рядом с записью, созданной на первом шаге, кликните ProspectKey.

При этом добавится уникальный идентификатор к запросу, так что вы сможете определить кто склонен или наоборот, несклонен к покупке велосипеда.

3. Нажмите на следующей ячейке в столбце Source, затем кликните на Targeted Mailing.

4. В ячейке Field, выберите Bike Buyer. Таким образом, определяется, что для создания прогнозов будет использована модель Microsoft Clustering из структуры Targeted Mailing.

5. Нажмите на следующей ячейке в столбце Source, затем нажмите Prediction Function.

6. Рядом с Prediction Function, в поле Field, нажмите PredictProbability.

Функции прогнозирования дают информацию о том, о том как делается прогноз. Функция PredictProbability предоставляет информацию о вероятности правильного прогноза. Вы можете определить параметры функции в столбце Criteria/Argument.

7. В поле Criteria/Argument, наберите [Targeted Mailing].[Bike Buyer].

Этим действием назначается целевой столбец для функции PredictProbability.

Для более подробной информации о функциях, обратитесь к разделу "DMX Function Reference" в SQL Server Books Online.

На вашем экране должно отобразиться следующее (см. рисунок 22).

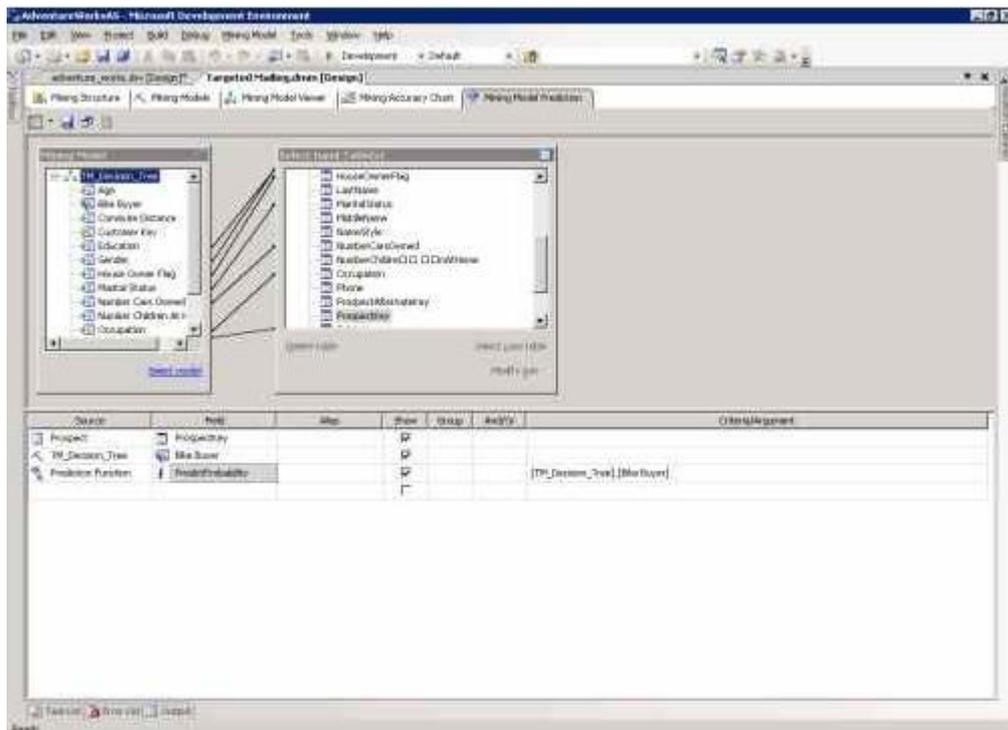


Рисунок 22 Prediction Query Builder на странице Mining Model Prediction.

Нажав на иконку в левом верхнем углу, можно переключиться в Query view и просмотреть DMX код, созданный Prediction Query Builder. Кроме того, вы можете исполнить запрос, изменить и затем исполнить его, однако изменённый запрос не отобразится при переходе назад в Design View.

### Просмотр результатов

Вы можете исполнить запрос, нажав на стрелку рядом с иконкой в левом верхнем углу страницы, затем нажмите Result. На рисунке 23 показаны результаты запроса.

| ProspectKey | BikeBuyer | Expression     |
|-------------|-----------|----------------|
| 21          | 0         | 0.79429180096  |
| 21          | 0         | 0.68152300006  |
| 43          | 0         | 0.65150300006  |
| 99          | 0         | 0.68170300006  |
| 99          | 0         | 0.68150300006  |
| 82          | 0         | 0.68150300006  |
| 95          | 1         | 0.55200011006  |
| 95          | 1         | 0.70210994596  |
| 96          | 1         | 0.61300000006  |
| 102         | 1         | 0.55200011006  |
| 111         | 1         | 0.55200011006  |
| 119         | 0         | 0.629825024106 |
| 140         | 0         | 0.79409150096  |
| 140         | 0         | 0.65150300006  |
| 192         | 0         | 0.68170300006  |
| 193         | 1         | 0.61300000006  |
| 196         | 0         | 0.55360474206  |
| 198         | 0         | 0.52805004106  |
| 212         | 0         | 0.702702654906 |
| 224         | 0         | 0.68150300006  |
| 238         | 0         | 0.523205674206 |
| 274         | 1         | 0.55200011006  |
| 305         | 0         | 0.597117002706 |
| 306         | 0         | 0.617380746106 |
| 321         | 0         | 0.65150300006  |
| 349         | 0         | 0.55360474206  |
| 391         | 0         | 0.629825024106 |
| 445         | 0         | 0.61300000006  |
| 503         | 1         | 0.633316700006 |
| 572         | 1         | 0.63290300006  |

Рисунок 23 Страница результатов.

Столбцы ProspectKey, BikeBuyer и Expression отвечают соответственно идентификатору потенциальных покупателей, купят ли они велосипед (1 или 0) и вероятности правильности прогноза. Вы можете использовать эти результаты, чтобы определить, кому следует разослать письма с предложением.

## **Сценарий учебника по службам Analysis Services**

Этот учебник основывается на вымышленной организации Компания Adventure Works Cycles. Компания Adventure Works Cycles — крупная транснациональная производственная компания, производящая и реализующая металлические и композитные велосипеды на рынках Северной Америки, Европы и Азии. Штаб-квартира компании Компания Adventure Works Cycles находится в городе Боселл, штат Вашингтон. В ней работают 500 сотрудников компании. Кроме того, компания Компания Adventure Works Cycles имеет в своем составе несколько групп сбыта на региональных рынках.

В последние годы компания Компания Adventure Works Cycles приобрела небольшой завод под названием Importadores Neptuno, расположенный в Мексике. Предприятие Importadores Neptuno производит несколько важных компонентов для производственной линии компании Компания Adventure Works Cycles. Эти компоненты поставляются в Боселл для окончательной сборки продукции. В 2005 году Importadores Neptuno стала единоличным производителем и поставщиком туристических велосипедов.

По окончании успешного финансового года Компания Adventure Works Cycles хочет расширить свое присутствие на рынке путем направленной рекламы для своих лучших клиентов, повысить доступность продукта при помощи внешнего веб-сайта и уменьшить стоимость продаж за счет снижения стоимости производства.

### **Текущая среда анализа**

Для поддержки нужд анализа данных групп маркетинга и сбыта, а также высшего руководства компания использует данные о транзакциях из базы данных AdventureWorks2012 и другие сведения, например целевые показатели объема продаж из электронных таблиц, и объединяет эту информацию в

реляционном хранилище данных **AdventureWorksDW2012**. Однако использование реляционного хранилища данных влечет следующие проблемы:

- Отчеты статичны. Пользователи не могут интерактивно просматривать данные отчетов для получения более подробной информации так, как это делается в сводной таблице в редакторе Microsoft Office Excel. Хотя существующий набор predefined отчетов достаточен для многих пользователей, более квалифицированным пользователям потребуется прямой доступ к запросам базы данных для выполнения интерактивных запросов и создания специализированных отчетов. Однако из-за сложности базы данных **AdventureWorksDW2012** таким пользователям потребуется много времени, чтобы освоить создание эффективных запросов.
- Скорость выполнения запросов меняется в широком диапазоне. Например, некоторые запросы возвращают результаты очень быстро, через считанные секунды, для выполнения других запросов требуется несколько минут.
- Таблицами статистических вычислений сложно управлять. Для сокращения времени обработки запросов команда разработчиков по хранилищам данных компании Adventure Works составила несколько таблиц статистических вычислений в базе данных **AdventureWorksDW2012**. Например, они создали таблицу, суммирующую продажи за месяц. Однако хотя такие таблицы существенно ускорили выполнение запросов, инфраструктура, поддерживающая эти таблицы, хрупка и подвержена ошибкам.
- Комплексная логика вычислений скрыта в определениях отчетов и сложна в передаче между отчетами. Так как бизнес-логика формируется отдельно для каждого отчета, данные сводки иногда различаются для разных отчетов. Поэтому руководство не слишком доверяет отчетам хранилища данных.

- Пользователи различных подразделений компании заинтересованы в различных представлениях данных. Пользователей отдельной группы приводят в замешательство и отвлекают элементы данных, не относящиеся к их группе.
- Логика вычислений особенно сложна для пользователей, которым требуются специализированные отчеты. Так как данные пользователи должны определять логику вычислений отдельно для каждого отчета, централизованного контроля за логикой вычислений нет. Например, некоторые пользователи знают, что им следовало бы использовать базовые статистические методы, такие как скользящее среднее, однако они не знают, как построить такие вычисления, и поэтому не используют эти методы.
- Трудно сочетать связанные наборы данных. Специализированные запросы, сочетающие два набора связанных данных, такие как продажи и квоты продаж, пользователям построить весьма непросто. Такие запросы перегружают базу данных, поэтому компания потребовала от пользователей запросить у команды разработчиков хранилища данных межпредметные наборы данных. В результате создано лишь небольшое количество predetermined отчетов, сочетающих данные из нескольких предметных областей. Более того, пользователи неохотно модифицируют эти отчеты из-за их сложности.
- Отчеты сконцентрированы в основном на деловых сведениях в Соединенных Штатах. Пользователи в филиалах за пределами США недовольны этим и хотят иметь возможность просматривать отчеты в разных денежных единицах и на разных языках.
- Сведения сложны для аудита. Финансовый отдел в данное время использует базу данных **AdventureWorksDW2012** только как источник данных для массовых запросов. Затем данные загружаются в отдельные электронные таблицы, что приводит к большим затратам времени на подготовку данных и на управление таблицами. Таким образом,

корпоративные финансовые отчеты трудно готовить, проверять и управлять ими.

## Решение

Команда разработчиков хранилища данных недавно провела исследование структуры текущей системы аналитики. Исследование включало анализ просчетов прогнозирования текущих проблем и будущих требований. Команда разработчиков хранилища данных определила, что база данных **AdventureWorksDW2012** — это хорошо сконструированная многомерная база данных с согласованными измерениями и суррогатными ключами. Согласованные измерения позволяют использовать измерение, например измерение времени или измерение продукта, в нескольких витринах данных. Суррогатные ключи — это искусственные ключи, связывающие измерения и таблицы фактов, они используются для обеспечения уникальности и повышения производительности. Кроме того, команда разработчиков хранилища данных определила, что в данный момент нет никаких значительных проблем с загрузкой базовых таблиц и управлением ими в базе данных **AdventureWorksDW2012**. Поэтому было решено с использованием базы данных служб Microsoft Службы Analysis Services выполнить следующие задачи:

- предоставить унифицированный доступ к данным через общий слой метаданных для анализа и отчетов;
- упростить просмотр данных пользователями, ускорив разработку интерактивных и предопределенных запросов, а также предопределенных отчетов;
- правильно составлять запросы, сочетающие данные из нескольких предметных областей;
- управлять статистическими вычислениями;
- хранить и повторно использовать сложные вычисления;

- предоставлять возможность локализации пользователям за пределами Соединенных Штатов.

Занятия в учебнике по службам Analysis Services позволяют изучить создание базы данных куба, которая удовлетворяет всем этим требованиям. Чтобы приступить к работе, перейдите к первому занятию: Занятие 1. Создание нового проекта табличной модели.

## **Занятие 1. Определение представления источников данных в проекте служб Analysis Services**

### **SQL Server 2014**

Разработка приложения бизнес-аналитики в SQL Server начинается с создания проекта служб Службы Analysis Services в среде SQL Server Data Tools (SSDT). В этом проекте будут определены все элементы решения, начиная с представления источников данных.

Это занятие содержит следующие задания.

#### 1. Создание проекта служб Analysis Services

В этой задаче будет создан проект Службы Analysis Services Tutorial на основе шаблона многомерной модели служб Службы Analysis Services.

#### 2. Определение источника данных

В этой задаче в качестве источника данных для измерений служб Службы Analysis Services и кубов, которые будут определяться на следующих занятиях, будет использоваться база данных **AdventureWorksDW2012**.

#### 3. Определение представления источников данных

В этой задаче будет определен единый вид представления метаданных из выбранных таблиц базы данных **AdventureWorksDW2012**.

#### 4. Изменение имен таблиц по умолчанию

В этой задаче будут изменены имена таблиц в представлении источников данных, чтобы сделать имена последующих определяемых объектов служб Службы Analysis Services более понятными для пользователей.

Сравните свои результаты с образцом файла проекта, созданным на этом занятии. Дополнительные сведения о загрузке образцов проектов для данного учебника см. в статье Проекты многомерных моделей служб SSAS для SQL Server 2012 на странице образцов продуктов на сайте CodePlex.

В следующей задаче будет использована среда SQL Server Data Tools (SSDT) для создания нового проекта служб Службы Analysis Services с именем **Учебник по службам Analysis Services** на основе шаблона проекта служб Службы Analysis Services. *Проект* представляет собой коллекцию связанных объектов. Проекты содержатся в решении, которое может включать несколько проектов. Дополнительные сведения см. в разделе Создание проекта служб Analysis Services (среда SSDT).

### 1. Создание нового проекта служб Analysis Services

1. Нажмите кнопку **Пуск**, укажите пункты **Все программы, Microsoft SQL Server 2012** и выберите пункт **SQL Server Data Tools**.

Откроется среда разработки Microsoft Visual Studio.

2. На начальной странице Visual Studio нажмите кнопку **Создание проекта**.
3. В диалоговом окне **Создание проекта** в области **Установленные шаблоны** разверните пункт **Бизнес-аналитика** и выберите **Службы Analysis Services**. Выберите шаблон **Проект многомерных данных и интеллектуального анализа данных служб Analysis Services**.

Обратите внимание, что имя проекта по умолчанию, расположение и имя решения по умолчанию генерируются в нижней части диалогового окна. По умолчанию для решения создается новый каталог.

4. Измените имя проекта на **Analysis Services Tutorial** (при этом также изменится поле **Имя решения**) и нажмите кнопку **ОК**.

Проект **Учебник по службам Analysis Services** был успешно создан на основе шаблона **Проект многомерных данных и интеллектуального анализа данных служб Analysis Services** в рамках нового решения, также названного **Учебник по службам Analysis Services**.

После создания проекта служб Службы Analysis Services работа с проектом обычно начинается с определения одного или нескольких источников данных, которые будут использоваться в этом проекте. Для определения источника данных нужно задать строку соединения, которая будет использована для подключения к этому источнику данных. Дополнительные сведения см. в разделе Создание источника данных (многомерные службы SSAS).

В следующей задаче предстоит определить образец базы данных AdventureWorksDWSQLServer2012 в качестве источника данных для проекта учебника по службам Службы Analysis Services. Хотя эта база данных в данном случае расположена на локальном компьютере, часто исходные базы данных размещаются на одном или нескольких удаленных компьютерах.

## 2. Определение нового источника данных

1. В обозревателе решений (справа от окна Microsoft Visual Studio) щелкните правой кнопкой мыши папку **Источники данных** и выберите **Создать новый источник данных**.

2. На странице **Мастер источников данных** в мастере источника данных нажмите кнопку **Далее**, чтобы открыть страницу **Выбор параметров соединения**.
3. На странице **Выбор метода определения соединения** можно определить источник данных на основе нового соединения, существующего соединения или предварительно определенного объекта источника данных. В этом учебнике будет определен источник данных на основе нового соединения. Убедитесь, что выбран параметр **Создать источник данных, основанный на существующем или на новом соединении**, а затем нажмите кнопку **Создать**.
4. В диалоговом окне **Диспетчер соединений** определяются свойства соединения для источника данных. Убедитесь, что в списке **Поставщик** выбран **Собственный поставщик данных OLE DB\Собственный клиент SQL Server 11.0**.

Службы Службы Analysis Services также поддерживают других поставщиков, которые доступны в списке **Поставщик**.

5. В текстовом поле **Имя сервера** введите **SERVER1**.

Чтобы подключиться к именованному экземпляру на локальном компьютере, введите **localhost\<имя\_экземпляра>**. Чтобы подключиться к конкретному компьютеру вместо локального, введите имя компьютера или его IP-адрес.

6. Убедитесь, что выбран параметр **Использовать проверку подлинности Windows**. В поле списка **Выберите или введите имя базы данных** выберите **AdventureWorksDW2012**.
7. Нажмите кнопку **Проверить соединение**, чтобы проверить соединение с базой данных.
8. Нажмите кнопку **ОК**, а затем нажмите кнопку **Далее**.

9. На странице мастера **Сведения об олицетворении** определяются учетные данные безопасности, которые будут использованы в службах Службы Analysis Services для подключения к источнику данных. Олицетворение влияет на учетную запись Windows, которая используется для подключения к источнику данных при выборе проверки подлинности Windows. Службы Службы Analysis Services не поддерживают олицетворение для обработки объектов OLAP. Выберите команду **Использовать учетную запись службы** и нажмите кнопку **Далее**.
10. На странице **Завершение работы мастера** примите имя по умолчанию, **Adventure Works DW 2012**, а затем нажмите кнопку **Готово**, чтобы создать новый источник данных.

### **3. Определение представления источников данных**

После определения источников данных, используемых в проекте служб Службы Analysis Services, на следующем этапе, как правило, определяется представление источника данных для этого проекта. Представление источника данных является отдельным единым представлением метаданных из указанных таблиц и представлений, определяемых источником данных для проекта. Хранение метаданных в представлении источника данных позволяет работать с метаданными в процессе разработки, не устанавливая соединений с базовыми источниками данных. Дополнительные сведения см. в разделе [Представления источников данных в многомерных моделях](#).

В следующей задаче будет определено представление источника данных, которое содержит пять таблиц из источника данных **AdventureWorksDW2012**.

#### **Определение нового представления источников данных**

1. В обозревателе решений (в правой части окна Microsoft Visual Studio) щелкните правой кнопкой мыши папку **Представления источников данных** и выберите **Создать представление источника данных**.

2. На странице **Вас приветствует мастер представления источников данных** нажмите кнопку **Далее**. Будет открыта страница **Выбор источника данных**.
3. В группе **Источник реляционных данных** выбран источник данных **Adventure Works DW 2012**. Нажмите кнопку **Далее**.

#### **Примечание**

Чтобы создать представление источника данных, в основе которого лежат несколько источников данных, необходимо предварительно определить представление источника данных, основанное на одном источнике данных. Этот источник данных впоследствии считается первичным источником данных. Затем можно добавить таблицы и представления из вторичного источника данных. При проектировании измерений с атрибутами, основанными на связанных таблицах нескольких источников данных, может потребоваться определить источник данных Microsoft SQL Server как основной источник данных, чтобы использовать его возможности обработки распределенных запросов.

4. На странице **Выбор таблиц и представлений** выберите таблицы и представления из списка объектов, доступных в выбранном источнике данных. Можно установить для этого списка фильтр, который поможет отобрать нужные таблицы и представления.

#### **Примечание**

Нажмите кнопку разворачивания в верхнем правом углу, чтобы окно заняло весь экран. Это облегчит просмотр полного списка доступных объектов.

5. В списке **Доступные объекты** выберите следующие объекты. Несколько таблиц можно выбрать, удерживая клавишу CTRL при выборе.
  - **DimCustomer (dbo)**

- **DimDate (dbo)**
  - **DimGeography (dbo)**
  - **DimProduct (dbo)**
  - **FactInternetSales (dbo)**
6. Нажмите кнопку **>**, чтобы добавить выбранные таблицы в список **Включенные объекты**.
  7. Нажмите кнопку **Далее**.
  8. Убедитесь, что в поле «Имя» введено **Adventure Works DW 2012**, и затем нажмите кнопку **Готово**.

Представление источника данных **Adventure Works DW 2012** будет выведено в папке **Представления источников данных** обозревателя решений. Содержимое представления источников данных также отображается в конструкторе представлений источников данных в среде SQL Server Data Tools (SSDT). В этом конструкторе содержатся следующие элементы:

- Панель **Диаграмма**, на которой представлены в графическом виде таблицы и их связи.
  - Панель **Таблицы**, на которой отображаются в виде дерева таблицы и их элементы схем.
  - Панель **Организатор схем**, на которой можно создавать вложенные диаграммы, позволяющие просматривать поднаборы этого представления источников данных.
  - Панель инструментов конструктора представлений источников данных.
9. Нажмите кнопку **Развернуть**, чтобы развернуть окно среды разработки Microsoft Visual Studio.
  10. На панели инструментов в верхней части конструктора представлений источников данных щелкните значок **Масштаб**, чтобы просмотреть

таблицы на панели **Диаграмма** в масштабе 50%. При этом будут скрыты подробные сведения в столбцах таблиц.

11. Чтобы скрыть обозреватель решений, нажмите кнопку **Автоматически скрыть**, на которой изображен значок кнопки, в заголовке обозревателя решений. Чтобы вернуться в обозреватель решений, установите указатель над вкладкой «Обозреватель решений» в правой части среды разработки. Чтобы раскрыть обозреватель решений, вновь нажмите кнопку **Автоматически скрыть**.

12. Если окна не скрыты по умолчанию, нажмите кнопку **Автоматически скрыть** в заголовке окна свойств или обозревателя решений.

Теперь все таблицы и связи между ними можно легко просматривать на панели **Диаграмма**. Обратите внимание, что между таблицами FactInternetSales и DimDate имеются три связи. С каждой продажей связано три даты: дата заказа, дата оплаты и дата отгрузки. Чтобы просмотреть подробные сведения по связи, дважды щелкните стрелку этой связи на панели **Диаграмма**.

#### 4. Изменение имен таблиц по умолчанию

##### SQL Server 2014

Значение свойства **FriendlyName** для объектов в представлении источника данных можно изменить, сделав его имя более понятным и более простым в использовании.

В следующей задаче будет изменено понятное имя каждой из таблиц в представлении источника данных посредством удаления из их имен префиксов **Dim** и **Fact**. Это сделает более понятными имена объектов куба и измерений, которые будут определены на следующем занятии.

## 📌 Примечание

Также в представлении источника данных можно присваивать понятные имена столбцам, определять вычисляемые столбцы, а также соединять таблицы и представления, чтобы сделать их более удобными в работе.

### Изменение имени таблицы по умолчанию

1. На панели **Таблицы конструктора представлений источников данных** щелкните правой кнопкой мыши таблицу **FactInternetSales** и выберите пункт **Свойства**.
2. Если окно свойств не отображено в правой части окна Microsoft Visual Studio, нажмите кнопку **Автоматически скрыть** в строке заголовка окна свойств, чтобы окно оставалось видимым.

Когда окно свойств открыто, проще изменять свойства каждой таблицы в представлении источника данных. Если окно не закреплено в открытом состоянии с помощью кнопки **Автоматически скрыть**, оно будет закрыто, как только выбран другой объект на панели **Диаграмма**.

3. Измените свойство **FriendlyName** объекта **FactInternetSales** на **InternetSales**.

Изменение будет применено, если щелкнуть вне ячейки свойства **FriendlyName**. На следующем занятии будет рассмотрено определение группы мер на основе этой таблицы фактов. Из-за сделанных на этом занятии изменений она будет называться не **FactInternetSales**, а **InternetSales**.

4. На панели **Таблицы** выберите таблицу **DimProduct**. В окне свойств задайте для свойства **FriendlyName** значение **Продукт**.

5. Точно таким же образом измените значения свойства **FriendlyName** для всех оставшихся таблиц в представлении источника данных, удалив префикс **Dim**.
6. Затем нажмите кнопку **Автоматически скрыть**, чтобы снова скрыть окно свойств.
7. В меню **Файл** выберите команду **Сохранить все** (или нажмите соответствующую кнопку на панели инструментов среды SQL Server Data Tools), чтобы сохранить изменения, внесенные в проект учебника по службам Службы Analysis Services. При необходимости здесь можно прервать работу с учебником, чтобы продолжить ее позже.

## Занятие 2. Определение и развертывание куба

После определения представления источников данных в проекте служб Microsoft Службы Analysis Services можно определить исходный куб служб Службы Analysis Services.

Кроме того, можно определить куб и его измерения за один проход с помощью мастера кубов. Также можно определить одно или несколько измерений, а затем с помощью мастера кубов определить куб, в котором они будут использоваться. Разработку сложного решения обычно начинают с определения измерений. Дополнительные сведения см. в разделах Измерения в многомерных моделях и Кубы в многомерных моделях.

### Примечание

Завершенные проекты для всех занятий в этом учебнике доступны через Интернет. Можно перейти вперед к любому занятию, используя в качестве отправной точки завершенный проект из предыдущего урока. Щелкните здесь

для загрузки образцов проектов, прилагаемых к этому учебнику.

Это занятие содержит следующие задания.

Определение измерения В этой задаче будет определено измерение с помощью мастера измерений.

## 1. Определение измерения

1. В обозревателе решений (справа в Microsoft Visual Studio) щелкните правой кнопкой мыши узел **Измерения** и выберите **Создать измерение**. Появится мастер измерений.
2. На странице **Вас приветствует мастер измерений** нажмите кнопку **Далее**.
3. На странице **Выбор метода создания** выберите параметр **Использовать существующую таблицу** и нажмите кнопку **Далее**.
4. На странице **Определение исходных сведений** убедитесь, что выбрано представление источника данных **Adventure Works DW 2012**.
5. В списке **Основная таблица** выберите таблицу **Date**.
6. Нажмите кнопку **Далее**.
7. На странице **Выбор атрибутов измерения** установите флажки для перечисленных ниже атрибутов:
  - **Ключ даты**
  - **Full Date Alternate Key**
  - **English Month Name**
  - **Calendar Quarter**
  - **Calendar Year**
  - **Calendar Semester**
8. Для атрибута **Резервный ключ полной даты** в столбце **Тип атрибута** вместо значения **Обычный** выберите **Дата**. Для этого щелкните значение **Обычный** в столбце **Тип атрибута**. Щелкните стрелку, чтобы раскрыть список параметров. Затем выберите значение **Дата &gt; Кален-**

**дарь &gt; Дата.** Нажмите кнопку **ОК**. Повторите шаги, чтобы изменить тип атрибута в других атрибутах следующим образом.

- **English Month Name** на **Месяц**
- **Calendar Quarter** на **Квартал**
- **Calendar Year** на **Год**
- **Calendar Semester** на **Полугодие**

9. Нажмите кнопку **Далее**.

10. На странице **Завершение работы мастера** на панели просмотра будет отображено измерение **Date** и его атрибуты.

11. Чтобы завершить работу мастера, нажмите кнопку **Готово**.

В обозревателе решений в проекте «Учебник по службам Службы Analysis Services» в папке **Измерения** появится измерение Date. В центральной части окна среды разработки это измерение отображается в конструкторе измерений.

12. В меню **Файл** выберите команду **Сохранить все**.

Определение куба В этой задаче с помощью мастера кубов будет определен исходный куб служб Службы Analysis Services.

Мастер кубов помогает определить для куба группы мер и измерения. В следующей задаче с помощью мастера кубов будет построен куб.

## **2. Определение куба и его свойств**

1. В обозревателе решений щелкните правой кнопкой мыши узел **Кубы** и выберите команду **Создать куб**. Появится мастер кубов.
2. На странице **Мастер кубов** нажмите кнопку **Далее**.

3. На странице **Выбор метода создания** убедитесь, что выбран параметр **Использовать существующие таблицы**, а затем нажмите кнопку **Далее**.
4. На странице **Выбор таблиц групп мер** убедитесь, что выбрано представление источника данных **Adventure Works DW 2012**.
5. Нажмите кнопку **Предложить**, чтобы мастер кубов предложил таблицы для создания групп мер.

Мастер просматривает все таблицы и предлагает **InternetSales** в качестве таблицы группы мер. Таблицы групп мер, также называемые таблицами фактов, содержат меры, которые могут представлять интерес (например, число проданных единиц).

6. Нажмите кнопку **Далее**.
7. На странице **Выбор мер** просмотрите выбранные меры в группе мер **Internet Sales** и снимите флажки для следующих мер:
  - **Promotion Key**
  - **Currency Key**
  - **Sales Territory Key**
  - **Revision Number**

По умолчанию мастер выбирает в качестве мер все числовые столбцы в таблице фактов, которые не привязаны к измерениям. Однако эти четыре столбца не являются фактическими мерами. Первые три представляют собой ключевые значения, связывающие таблицу фактов с таблицами измерений, которые не используются в первоначальной версии этого куба.

8. Нажмите кнопку **Далее**.
9. На странице **Выбор существующих измерений** убедитесь, что выбрано ранее созданное измерение **Date**, и нажмите кнопку **Далее**.

10. На странице **Выбор новых измерений** выберите новые измерения, которые необходимо создать. Для этого убедитесь, что установлены флажки **Customer**, **Geography** и **Product**, а флажок **InternetSales** снят.
11. Нажмите кнопку **Далее**.
12. На странице **Завершение работы мастера** измените имя куба на **Analysis Services Tutorial**. На панели просмотра будет отображена группа мер **InternetSales** и входящие в нее меры. Кроме того, будут показаны измерения **Date**, **Customer** и **Product**.
13. Чтобы завершить работу мастера, нажмите кнопку **Готово**.

В обозревателе решений в проекте «Учебник по службам Службы Analysis Services» появится куб «Учебник Службы Analysis Services» в папке **Кубы**, а измерения баз данных «Customer» и «Product» появятся в папке **Измерения**. Кроме того, в центре среды разработки на вкладке «Структура куба» будет отображен куб «Учебник Службы Analysis Services».

14. Чтобы лучше рассмотреть измерения и таблицы фактов в кубе, на панели инструментов вкладки «Структура куба» переключите **Масштаб** в значение 50 %. Обратите внимание, что таблицы фактов выделены желтым цветом, а таблицы измерений — синим.
15. В меню **Файл** выберите команду **Сохранить все**.

### **3. Добавление атрибутов в измерение «Заказчик»**

Добавление атрибутов к измерениям В этой задаче в созданные измерения будут добавлены атрибуты.

Теперь после определения измерений можно наполнить их атрибутами, которые представляют все элементы данных в измерении. Атрибуты обычно основаны на полях из представления источников данных. При добавлении ат-

рибутов в измерение можно включить поля из любой таблицы в представлении источника данных.

В этой задаче с помощью конструктора измерений в измерения «Заказчик» и «Продукт» будут добавлены атрибуты. Измерение «Заказчик» будет включать атрибуты, основанные на полях из таблиц как «Заказчик», так и «География».

### Добавление атрибутов

1. Откройте в конструкторе измерений измерение «Заказчик». Для этого щелкните дважды по измерению **Клиент** в узле **Измерения** обозревателя решений.
2. На панели **Атрибуты** обратите внимание на атрибуты «Customer Key» и «Geography Key», созданные мастером кубов.
3. На панели инструментов вкладки **Структура измерения** щелкните значок «Масштаб», чтобы просмотреть таблицы области **Представление источника данных** в масштабе 100 %.
4. Перетащите следующие столбцы из таблицы **Customer** в области **Представление источника данных** в область **Атрибуты**.
  - **BirthDate**
  - **MaritalStatus**
  - **Gender**
  - **EmailAddress**
  - **YearlyIncome**
  - **TotalChildren**
  - **NumberChildrenAtHome**
  - **EnglishEducation**
  - **EnglishOccupation**
  - **HouseOwnerFlag**

- **NumberCarsOwned**
  - **Phone**
  - **DateFirstPurchase**
  - **CommuteDistance**
5. Перетащите следующие столбцы из таблицы **Geography** в области **Представление источника данных** в область **Атрибуты**.
- **City**
  - **StateProvinceName**
  - **EnglishCountryRegionName**
  - **PostalCode**
6. В меню «Файл» выберите команду **Сохранить все**.

### **Добавление атрибутов в измерение «Продукт»**

#### **Добавление атрибутов**

1. Откройте в конструкторе измерений измерение «Продукт». Дважды щелкните измерение **Продукт** в обозревателе решений.
2. На панели **Атрибуты** обратите внимание на атрибут «Ключ продукта», созданный мастером кубов.
3. На панели инструментов вкладки **Структура измерения** щелкните значок «Масштаб», чтобы просмотреть таблицы области **Представление источника данных** в масштабе 100 %.
4. Перетащите следующие столбцы из таблицы **Продукт** в области **Представление источника данных** в область **Атрибуты**.
  - **StandardCost**
  - **Color**
  - **SafetyStockLevel**

- **ReorderPoint**
- **ListPrice**
- **Size**
- **SizeRange**
- **Weight**
- **DaysToManufacture**
- **ProductLine**
- **DealerPrice**
- **Class**
- **Style**
- **ModelName**
- **StartDate**
- **EndDate**
- **Status**

5. В меню «Файл» выберите команду **Сохранить все**.

Просмотр свойств куба и измерения В этой задаче будет просмотрена структура куба, определенная с помощью мастера кубов.

После определения свойств куба можно посмотреть результаты в конструкторе кубов. В следующей задаче предстоит посмотреть структуру куба проекта Службы Analysis Services Tutorial.

#### **4. Просмотр свойств куба и измерений в конструкторе кубов**

1. Чтобы открыть конструктор кубов, в обозревателе решений в узле **Кубы** дважды щелкните куб **Analysis Services Tutorial**.
2. Чтобы посмотреть определенные меры, на панели **Меры** вкладки **Структура куба** конструктора кубов раскройте группу мер **Internet Sales**.

Можно изменить порядок отображения этих мер, перетаскивая их и выстраивая в желаемой последовательности. Такой порядок влияет на очередность использования этих мер определенными клиентскими приложениями. Эта группа мер и каждая содержащаяся в ней мера имеют свойства, которые можно изменять в окне свойств.

3. На панели **Измерения** вкладки **Структура куба** конструктора кубов просмотрите измерения куба Службы Analysis Services Tutorial.

Обратите внимание, что в кубе Службы Analysis Services Tutorial доступно пять измерений, хотя на уровне базы данных было создано только три измерения, что отражено в обозревателе решений. Куб имеет больше измерений, чем база данных, потому что измерение Date в базе данных служит основой для трех отдельных измерений куба, связанных с датами. Эти измерения основаны на разных связанных с датами фактах из таблицы фактов. Эти измерения даты называются также *ролевыми измерениями*. Три измерения даты куба дают пользователям возможность разделить куб по трем отдельным фактам, связанным с каждой продажей: дате заказа товара, сроку выполнения заказа и дате отгрузки заказа. Повторно используя одно измерение базы данных для нескольких измерений куба, службы Службы Analysis Services упрощают управление измерениями, используют меньше места на диске и уменьшают общее время обработки данных.

4. На панели **Измерения** вкладки **Структура куба** раскройте измерение **Customer**, а затем нажмите кнопку **Изменить Customer**, чтобы открыть это измерение в конструкторе измерений.

Конструктор измерений содержит следующие вкладки: **Структура измерения**, **Связи атрибутов**, **Переводы** и **Браузер**. Обратите внимание, что вкладка **Структура измерения** содержит три панели: **Атрибуты**, **Иерархии** и **Представление источника данных**. Атрибуты, содержа-

щиеся в измерении, отображаются на панели **Атрибуты**. Дополнительные сведения см. в разделе Справочник по свойствам атрибута измерения, Создание пользовательских иерархий.

5. Чтобы переключиться в конструктор кубов, щелкните правой кнопкой мыши **Учебник Analysis Services** в узле **Кубы** в обозревателе решений и выберите пункт **Конструктор представлений**.
6. В конструкторе кубов перейдите на вкладку **Использование измерений**.

В этом представлении куба Службы Analysis Services Tutorial показаны измерения куба, используемые группой мер Internet Sales. Также можно определить тип связи между каждым измерением и каждой группой мер, в которой оно используется.

7. Перейдите на вкладку **Секции**.

Мастер кубов определяет единственную секцию для куба с использованием режима хранения результатов многомерной интерактивной аналитической обработки данных (MOLAP) без статистических выражений. Для обработки MOLAP все данные конечного уровня и все статистические выражения хранятся в кубе, чтобы обеспечить максимальную производительность. Статистические выражения представляют собой предварительно вычисленные сводные данные, которые содержат ответы на еще не заданные вопросы, что позволяет сократить время до получения ответа на запрос. На вкладке **Секции** можно определять дополнительные секции, параметры хранения и настройки обратной записи. Дополнительные сведения см. в разделе Секции (службы Analysis Services — многомерные данные), Агрегаты и статистические схемы.

8. Перейдите на вкладку **Браузер**.

Обратите внимание, что куб нельзя просмотреть, поскольку он еще не был развернут на экземпляре служб Службы Analysis Services. На данном этапе куб в проекте Службы Analysis Services Tutorial представляет собой лишь определение куба, его можно развернуть на любом экземпляре служб Службы Analysis Services. Во время развертывания и обработки куба создаются определенные объекты в экземпляре служб Службы Analysis Services, которые затем заполняются данными из базовых источников данных.

9. В обозревателе решений щелкните правой кнопкой мыши **Учебник Analysis Services** в узле **Кубы** и выберите пункт **Просмотр кода**. Возможно, потребуется подождать.

XML-код для куба «Учебник по Службы Analysis Services» отображается на вкладке **Службы Analysis Services Tutorial.cube [XML]**. Это фактический код, используемый для создания куба в экземпляре служб Службы Analysis Services в процессе развертывания. Дополнительные сведения см. в разделе [Просмотр XML-кода проекта служб Analysis Services \(среда SSDT\)](#).

10. Закройте вкладку XML-кода.

Развертывание проекта служб Analysis Services В этой задаче будет выполнено развертывание проекта служб Службы Analysis Services в локальном экземпляре служб Службы Analysis Services, и будут изучены некоторые свойства такого развертывания.

Чтобы просмотреть куб и данные измерения для объектов куба «Учебник по службам Службы Analysis Services» проекта «Учебник по службам Службы Analysis Services», необходимо развернуть проект на указанном экземпляре служб Службы Analysis Services, а затем выполнить обработку куба и его из-

мерений. В процессе *развертывания* проекта служб Службы Analysis Services в экземпляре служб Службы Analysis Services создаются определенные объекты. При *обработке* объектов в экземпляре Службы Analysis Services данные из базовых источников данных копируются в объекты кубов. Дополнительные сведения см. в разделах Развертывание проектов служб Analysis Services (среда SSDT) и Настройка свойств проекта служб Analysis Services (среда SSDT).

На этой стадии процесса разработки развертывание куба обычно производится в экземпляре служб Службы Analysis Services на сервере разработки. После завершения разработки проекта бизнес-аналитики обычно производится развертывание данных на рабочем сервере с помощью мастера развертывания служб Службы Analysis Services. Дополнительные сведения см. в разделах Развертывание решений многомерных моделей и Развертывание решений модели с использованием мастера развертывания.

Для выполнения следующей задачи необходимо проверить свойства развертывания проекта Tutorial служб Службы Analysis Services, а затем развернуть его для локального экземпляра служб Службы Analysis Services.

## **Развертывание проекта служб Analysis Services**

1. В обозревателе решений щелкните правой кнопкой мыши проект **Учебник по службам Analysis Services** и выберите команду **Свойства**.

Откроется диалоговое окно **Страницы свойств учебника по службам Analysis Services**, содержащее свойства конфигурации Active(Development). Можно определить несколько конфигураций, каждая из которых будет иметь собственные значения свойств. Например, разработчик может по-разному настроить один и тот же проект для его развертывания на разных компьютерах с разными свойствами среды разработки, такими как имена баз данных или свойства обработ-

ки. Обратите внимание на значение свойства **Путь вывода**. Это свойство указывает расположение, в котором сохраняются XMLE-скрипты развертывания проекта при его сборке. Это скрипты, используемые для развертывания содержащихся в проекте объектов для экземпляра служб Службы Analysis Services.

2. В узле **Свойства конфигурации** на левой панели щелкните элемент **Развертывание**.

Просмотрите свойства развертывания проекта. По умолчанию шаблон проекта служб Службы Analysis Services настроен таким образом, что выполняется последовательное развертывание всех проектов служб Службы Analysis Services для указанного по умолчанию экземпляра служб Службы Analysis Services на локальном компьютере, создается база данных Службы Analysis Services с именем проекта, а также выполняется обработка объектов с параметрами по умолчанию после завершения развертывания. Дополнительные сведения см. в разделе [Настройка свойств проекта служб Analysis Services \(среда SSDT\)](#).

#### **Примечание**

Если необходимо развернуть проект в именованном экземпляре служб Службы Analysis Services на локальном компьютере или в экземпляре, расположенном на удаленном сервере, задайте для свойства **Сервер** соответствующее имя экземпляра, например `<ИмяСервера>\<ИмяЭкземпляра>`.

3. Нажмите кнопку **ОК**.
4. В обозревателе решений щелкните правой кнопкой мыши проект **Учебник Analysis Services** и выберите пункт **Развертывание**. Возможно, потребуется подождать.

## 📌 Примечание

Если при развертывании возникают ошибки, проверьте разрешения на базу данных с помощью среды SQL Server Management Studio. У учетной записи, указанной для соединения с источником данных, должно быть имя входа в экземпляр SQL Server. Дважды щелкните имя пользователя для просмотра свойств сопоставления пользователей. Эта учетная запись должна иметь разрешения `db_datareader` для базы данных **AdventureWorksDW2012**.

5. В среде SQL Server Data Tools (SSDT) собирается и развертывается проект учебника по службам Службы Analysis Services на указанном экземпляре служб Службы Analysis Services с использованием скрипта развертывания. Ход выполнения развертывания отображается в двух окнах: **Выход** и **Выполнение развертывания — учебник по службам Analysis Services**.
6. Откройте окно вывода, если необходимо, выбрав пункт **Выход** в меню **Вид**. В окне **Вывод** отображается общий ход выполнения развертывания. Окно **Выполнение развертывания — учебник по службам Analysis Services** отображаются подробные сведения о каждом шаге, выполняемом в процессе развертывания. Дополнительные сведения см. в разделах Построение проектов служб Analysis Services (среда SSDT) и Развертывание проектов служб Analysis Services (среда SSDT).
7. Просмотрите содержимое окон **Вывод** и **Выполнение развертывания — учебник по службам Analysis Services**, чтобы убедиться, что куб построен, развернут и обработан без ошибок.
8. Чтобы скрыть окно **Выполнение развертывания — учебник по службам Analysis Services**, нажмите значок **Автоматически скрыть** (значок похож на канцелярскую кнопку) на панели инструментов этого окна.

9. Чтобы скрыть окно **Вывод**, нажмите значок **Автоматически скрыть** на панели инструментов этого окна.

Развертывание куба учебника по службам Службы Analysis Services для локального экземпляра Службы Analysis Services и его обработка успешно завершены.

Просмотр куба В этой задаче будут просмотрены данные куба и измерения с помощью Excel и конструктора запросов многомерных выражений.

### **Просмотр куба**

После развертывания куба данные куба отображаются на вкладке **Обозреватель** конструктора кубов, а данные измерений отображаются на вкладке **Браузер** конструктора измерений. Просмотр данных куба и измерения позволяет постепенно проверять свою работу. Можно видеть, что небольшие изменения свойств, связей и других объектов оказывают нужный эффект после обработки объекта. Вкладка «Обозреватель» используется как для данных куба, так и для данных измерения, однако в зависимости от просматриваемого объекта она предлагает разные возможности.

Для измерений вкладка «Обозреватель» позволяет просматривать элементы или осуществлять навигацию по иерархии вплоть до конечных узлов. Данные измерения можно просматривать на разных языках, если в модель были добавлены переводы.

Для кубов вкладка «Обозреватель» предлагает два подхода для просмотра данных. С помощью встроенного конструктора запросов многомерных выражений можно создавать запросы, которые возвращают плоский набор строк из многомерной базы данных. Кроме того, можно использовать ярлык Excel. При запуске из среды SQL Server Data Tools Excel открывается со

сводной таблицей на листе и уже определенным подключением к базе данных рабочей области модели.

Как правило, Excel предлагает более широкие возможности просмотра, поскольку позволяет исследовать данные куба в интерактивном режиме с использованием горизонтальной и вертикальной оси для анализа связей в данных. Напротив, конструктор запросов многомерных выражений ограничен одной осью. Более того, поскольку набор строк является плоским, получить углубленную детализацию, обеспечиваемую сводными таблицами Excel, невозможно. При добавлении новых измерений и иерархий в куб, что предстоит на следующих занятиях, Excel будет предпочтительным решением для просмотра данных.

### **Просмотр развернутого куба**

1. В **конструкторе измерений** среды SQL Server Data Tools (SSDT) откройте измерение «Продукт». Для этого дважды щелкните измерение **Продукт** в узле **Измерения** обозревателя решений.
2. Перейдите на вкладку **Браузер**, чтобы просмотреть элемент **Все** иерархии атрибутов **Product Key**. На третьем занятии будет определена пользовательская иерархия для измерения «Продукты», позволяющая просматривать это измерение.
3. Перейдите в **конструктор кубов** среды SQL Server Data Tools (SSDT). Для этого в обозревателе решений в узле **Кубы** дважды щелкните куб **Учебник по службам Analysis Services**.
4. Перейдите на вкладку **Браузер** и на панели инструментов конструктора нажмите значок **Повторное соединение**.

Левая панель конструктора отображает объекты куба учебника по службам Службы Analysis Services. В правой части вкладки **Обозреватель** находятся две панели: верхняя панель — панель **Фильтр**, а ниж-

няя — панель **Данные**. На следующем занятии будет выполнен анализ с помощью средства просмотра кубов.

### Занятие 3. Изменение мер, атрибутов и иерархий

После определения начального куба можно увеличить его полезность и удобство использования. Это можно сделать, добавив иерархии, поддерживающие навигацию и агрегаты на разных уровнях, применив форматы к конкретной мере и определив вычисления и связи.

#### Примечание

Завершенные проекты для всех занятий в этом учебнике доступны через Интернет. Можно перейти вперед к любому занятию, используя в качестве отправной точки завершенный проект из предыдущего урока. [Щелкните здесь](#) для загрузки образцов проектов, прилагаемых к этому учебнику.

Это занятие содержит следующие задачи.

#### Изменение мер

В этой задаче предстоит указать свойства форматирования мер валюты и процентов в кубе учебника по службам Службы Analysis Services.

Свойство **FormatString** позволяет определить параметры форматирования, управляющие способом отображения мер для пользователей. В этой задаче предстоит указать свойства форматирования мер валюты и процентов в кубе учебника по службам Службы Analysis Services.

#### **Изменение мер куба**

1. Перейдите на вкладку **Структура куба** конструктора кубов для куба учебника по службам Службы Analysis Services, раскройте группу мер **Internet Sales** на панели **Меры**, щелкните правой кнопкой мыши элемент **Order Quantity** и выберите команду **Свойства**.
2. В окне свойств нажмите значок канцелярской кнопки **Автоматически скрыть**, чтобы оставить окно свойств постоянно открытым.

Если окно свойств остается постоянно открытым, изменять свойства нескольких элементов куба проще.

3. В окне свойств щелкните список **FormatString** и введите **#,#**.
4. На панели инструментов вкладки **Структура куба** нажмите значок **Показывать сетку мер** слева.

Сетка просмотра позволяет выбрать несколько мер одновременно.

5. Выберите следующие меры. Можно выбрать несколько мер. Для этого щелкните каждую из них, удерживая нажатой клавишу CTRL.
  - **Unit Price**
  - **Extended Amount**
  - **Discount Amount**
  - **Product Standard Cost**
  - **Total Product Cost**
  - **Sales Amount**
  - **Tax Amt**
  - **Freight**
6. В окне свойств в списке **FormatString** выберите параметр **Валюта**.
7. В раскрывающемся списке в верхней части окна свойств (под строкой названия) выберите меру **Процент скидки от стоимости единицы товара**, а затем выберите значение **Процент** в списке **FormatString**.
8. В окне свойств измените свойство **Имя** меры **Unit Price Discount Pct** на **Unit Price Discount Percentage**.

9. На панели **Меры** щелкните **Tax Amt** и измените имя меры на **Tax Amount**.
10. В окне свойств нажмите значок **Автоматически скрыть**, чтобы скрыть окно свойств, а затем нажмите кнопку **Показывать дерево мер** на вкладке панели инструментов **Структура куба**.
11. В меню **Файл** выберите команду **Сохранить все**.

### Изменение измерения «Заказчик»

В этой задаче определяется пользовательская иерархия, создаются именованные вычисления, изменяются атрибуты для использования именованных вычислений, группируются атрибуты и пользовательские иерархии в папки отображения.

Существует много способов повысить удобство использования и функциональность измерений куба. При выполнении задач этого раздела будет изменено измерение «Заказчик».

### **Переименование атрибутов**

Имена атрибутов можно изменить на вкладке **Структура измерения** конструктора измерений.

### **Переименование атрибута**

1. Откройте измерение «Клиент» в **конструкторе измерений** среды SQL Server Data Tools (SSDT). Для этого дважды щелкните измерение **Клиент** в узле **Измерения** обозревателя решений.
2. На панели **Атрибуты** щелкните правой кнопкой мыши атрибут **Название региона Англии** и выберите команду **Переименовать**. Измените имя атрибута на **Country-Region**.
3. Аналогичным образом измените имена следующих атрибутов.
  - Имя атрибута **English Education** замените на **Education**.
  - Имя атрибута **English Occupation** замените на **Occupation**.
  - Имя атрибута **State Province Name** замените на **State-Province**.
4. В меню **Файл** выберите команду **Сохранить все**.

### Создание иерархии

Новую иерархию можно создать, перетащив атрибут из панели **Атрибуты** на панель **Иерархии**.

### Создание иерархии

1. Перетащите атрибут **Country-Region** из панели **Атрибуты** на панель **Иерархии**.
2. Перетащите атрибут **State-Province** из панели **Атрибуты** в ячейку **<новый уровень>** на панели **Иерархии** под уровнем **Country-Region**.
3. Перетащите атрибут **City** из панели **Атрибуты** в ячейку **<новый уровень>** на панели **Иерархии** под уровнем **State-Province**.
4. На панели **Иерархии** на вкладке **Структура измерения** щелкните правой кнопкой мыши строку заголовка окна **Иерархия**, выберите команду **Переименовать** и введите **География клиентов**.

Теперь иерархия называется **Customer Geography**.

5. В меню **Файл** выберите команду **Сохранить все**.

### **Добавление именованного вычисления**

К таблице в представлении источника данных можно добавить именованное вычисление, которое является выражением SQL и представляет собой вычисляемый столбец. Это выражение имеет вид и функции столбца таблицы. Именованные вычисления позволяют расширять реляционную схему таблиц, существующих в представлении источника данных, не изменяя структуру таблиц в базовом источнике данных. Дополнительные сведения см. в разделе [Определение именованных вычислений в представлении источника данных \(службы Analysis Services\)](#).

### **Добавление именованного вычисления**

1. Откройте представление источника данных **Adventure Works DW 2012**, дважды щелкнув его в папке **Представления источников данных** обозревателя решений.
2. На панели **Таблицы** слева щелкните правой кнопкой мыши таблицу **Клиент** и выберите команду **Создать именованное вычисление**.
3. В диалоговом окне **Создание именованного вычисления** в поле **Имя столбца** введите **FullName**, затем введите или скопируйте и вставьте в поле **Выражение** следующую инструкцию **CASE**:
4. CASE
5.     WHEN MiddleName IS NULL THEN
6.     FirstName + ' ' + LastName
7.     ELSE
8.     FirstName + ' ' + MiddleName + ' ' + LastName
9. END
- 10.

Инструкция **CASE** объединяет столбцы **FirstName**, **MiddleName** и **LastName** в один столбец, который затем может использоваться в измерении **Customer** в качестве отображаемого имени для атрибута **Customer**.

11.Нажмите кнопку **ОК**, а затем раскройте таблицу **Customer** на панели **Таблицы**.

Именованное вычисление **FullName** появляется в списке столбцов таблицы **Customer** со значком, обозначающим, что это именованное вычисление.

12.В меню **Файл** выберите команду **Сохранить все**.

13.На панели **Таблицы** щелкните правой кнопкой мыши таблицу **Клиент** и выберите пункт **Просмотр данных**.

14.Просмотрите последний столбец в представлении **Просмотр таблицы Customer**.

Обратите внимание, что в представлении источника данных появился столбец **FullName**, содержащий сцепление данных из других столбцов источника данных; при этом исходный источник данных изменять не пришлось.

15.Закройте вкладку **Просмотр таблицы Customer**.

### **Использование именованных вычислений в качестве имен элементов**

После создания именованного вычисления в представлении источника данных это вычисление можно использовать в качестве свойства атрибута.

### **Использование именованного вычисления в качестве имен элементов**

1. В конструкторе измерений откройте измерение Customer.
2. На панели **Атрибуты** вкладки **Структура измерения** выберите атрибут **Customer Key**.
3. Откройте окно «Свойства» и нажмите в строке заголовка кнопку **Автоматически скрыть**, чтобы оно оставалось открытым.
4. В поле свойства **Name** введите **Full Name**.
5. Щелкните поле свойства **NameColumn** внизу и нажмите кнопку обзора (...), чтобы открыть диалоговое окно **Столбец имени**.
6. Выберите пункт **FullName** внизу списка **Исходный столбец** и нажмите кнопку **ОК**.
7. На вкладке «Структура измерений» перетащите атрибут **Полное имя** с панели **Атрибуты** в ячейку <новый уровень> области **Иерархии** под уровень **Город**.
8. В меню **Файл** выберите команду **Сохранить все**.

## Определение папок отображения

Папки отображения позволяют группировать пользовательские иерархии и иерархии атрибутов в структуры папок, удобные для пользователя.

## Определение папок отображения

1. Откройте вкладку **Структура измерения** для измерения Customer.
2. На панели **Атрибуты** выберите следующие атрибуты (щелкните каждый из них, удерживая нажатой клавишу CTRL).
  - **City**
  - **Country-Region**
  - **Postal Code**
  - **State-Province**

3. В окне «Свойства» щелкните поле свойства **AttributeHierarchyDisplayFolder** сверху (может потребоваться указать на элемент, чтобы увидеть его полностью) и введите **Расположение**.
4. На панели **Иерархии** щелкните элемент **География клиентов** и в окне «Свойства» справа выберите пункт **Расположение** в качестве значения свойства **DisplayFolder**.
5. На панели **Атрибуты** выберите следующие атрибуты (щелкните каждый из них, удерживая нажатой клавишу CTRL):
  - **Commute Distance**
  - **Education**
  - **Gender**
  - **House Owner Flag**
  - **Marital Status**
  - **Number Cars Owned**
  - **Number Children At Home**
  - **Occupation**
  - **Total Children**
  - **Yearly Income**
6. В окне свойств щелкните поле свойства **AttributeHierarchyDisplayFolder** сверху и введите **Демография**.
7. На панели **Атрибуты** выберите следующие атрибуты (щелкните каждый из них, удерживая нажатой клавишу CTRL).
  - **Email Address**
  - **Phone**
8. В окне свойств щелкните поле свойства **AttributeHierarchyDisplayFolder** и введите **Contacts**.
9. В меню **Файл** выберите команду **Сохранить все**.

## Определение составных ключевых столбцов

Свойство **KeyColumns** содержит столбец или столбцы, представляющие ключ для атрибута. На этом занятии будет создан составной ключ для атрибутов **City** и **State-Province**. Составные ключи могут оказаться полезными для уникальной идентификации атрибута. Например, при определении связи атрибутов (см. далее в этом учебнике) атрибут **City** должен уникально идентифицировать атрибут **State-Province**. Однако в разных областях бывают города с одинаковыми названиями. Поэтому для атрибута **Город** необходимо создать составной ключ, состоящий из столбцов **StateProvinceName** и **City**. Дополнительные сведения см. в разделе Изменение свойства KeyColumn атрибута.

### Определение составного свойства KeyColumns для атрибута City

1. Откройте вкладку **Структура измерения** для измерения Customer.
2. На панели **Атрибуты** щелкните атрибут **City**.
3. В окне **Свойства** щелкните в поле **KeyColumns** ближе к нижней части и нажмите кнопку обзора (...).
4. В диалоговом окне **Ключевые столбцы** в списке **Доступные столбцы** выберите столбец **StateProvinceName** и нажмите кнопку **&gt;**.  
Теперь в списке **Ключевые столбцы** отображаются столбцы **City** и **StateProvinceName**.
5. Нажмите кнопку **ОК**.
6. Чтобы задать свойство **NameColumn** атрибута **Город**, щелкните в окне свойств поле **NameColumn** в окне «Свойства» и нажмите кнопку обзора (...).
7. В диалоговом окне **Столбец имени** в списке **Исходный столбец** выберите **City** и нажмите кнопку **ОК**.

8. В меню **Файл** выберите команду **Сохранить все**.

### **Определение составного свойства KeyColumns для атрибута «Республика, область или край»**

1. Убедитесь, что вкладка **Структура измерения** для измерения «Клиент» открыта.
2. На панели **Атрибуты** щелкните атрибут **State-Province**.
3. В окне **Свойства** щелкните в поле **KeyColumns** и нажмите кнопку обзора (...).
4. В диалоговом окне **Ключевые столбцы** в списке **Доступные столбцы** выберите столбец **EnglishCountryRegionName** и нажмите кнопку **>**.

Теперь в списке **Ключевые столбцы** отображаются столбцы **EnglishCountryRegionName** и **StateProvinceName**.

5. Нажмите кнопку **ОК**.
6. Чтобы задать свойство **NameColumn** атрибута **Штат или провинция**, щелкните в окне свойств поле **NameColumn** в окне «Свойства» и нажмите кнопку обзора (...).
7. В диалоговом окне **Столбец имени**, в списке **Исходный столбец**, выберите **StateProvinceName** и нажмите кнопку **ОК**.
8. В меню **Файл** выберите команду **Сохранить все**.

### **Определение связей атрибутов**

Необходимо определять связи между атрибутами, если базовые данные это поддерживают. Определение связей между атрибутами ускоряет обработку измерений, секций и запросов. Дополнительные сведения см. в разделах [Определение связей атрибутов](#) и [Связи атрибутов](#).

## Определение связей атрибутов

1. В конструкторе измерений для измерения Customer откройте вкладку **Связи атрибутов**. Возможно, потребуется подождать.
2. На диаграмме щелкните правой кнопкой мыши атрибут **Город** и выберите команду **Создать связь атрибутов**.
3. В диалоговом окне **Создание связи атрибутов** поле **Исходный атрибут** имеет значение **City**. Для поля **Связанный атрибут** задайте значение **State-Province**.
4. В списке **Тип связи** выберите тип связи **Жесткая**.

Связь имеет тип **Жесткая**, поскольку связи между элементами не будут меняться с течением времени. Например, переход города под юрисдикцию другого штата или провинции — явление крайне редкое.

5. Нажмите кнопку **ОК**.
6. На диаграмме щелкните правой кнопкой мыши атрибут **State-Province** и выберите команду **Создать связь атрибутов**.
7. В диалоговом окне **Создание связи атрибутов** поле **Исходный атрибут** имеет значение **State-Province**. Задайте для поля **Связанный атрибут** значение **Country-Region**.
8. В списке **Тип связи** выберите тип связи **Жесткая**.
9. Нажмите кнопку **ОК**.
10. В меню **Файл** выберите команду **Сохранить все**.

## Развертывание изменений, обработка объектов и просмотр изменений

После изменения атрибутов и иерархий необходимо произвести развертывание произведенных изменений и повторную обработку связанных объектов, прежде чем эти изменения можно будет просмотреть.

## Развертывание изменений, обработка объектов и просмотр изменений

1. В меню **Построение** среды SQL Server Data Tools выберите команду **Развернуть учебник по службам Analysis Services**.
2. Получив сообщение **Развертывание выполнено успешно**, в конструкторе измерений перейдите на вкладку **Браузер** измерения «Клиент» и щелкните на панели инструментов значок повторного соединения.
3. Убедитесь, что в списке **Иерархия** выделена иерархия **География клиентов**, на панели браузера последовательно раскройте узлы **Все**, **Australia (Австралия)**, **New South Wales (Новый Южный Уэльс)** и наконец **Coffs Harbour (Коффс Харбор)**.

В браузере отображаются заказчики, живущие в этом городе.

4. Переключитесь в **конструктор кубов** для куба учебника по службам Службы Analysis Services». Чтобы сделать это, дважды щелкните куб **Учебник по службам Analysis Services**, который находится в узле **Кубы** в дереве **обозревателя решений**.
5. Перейдите на вкладку **Браузер** и на панели инструментов конструктора нажмите кнопку повторного соединения.
6. На панели **Группа мер** разверните узел **Заказчик**.

Обратите внимание, что вместо длинного списка атрибутов для измерения «Заказчик» отображаются только папки отображения и те атрибуты, для которых не указана папка отображения.

7. В меню **Файл** выберите команду **Сохранить все**.

## Изменение измерения Product

В этой задаче определяется пользовательская иерархия, создаются именованные вычисления, определяются имя элемента «Все» и папки отображения.

При выполнении задач этого раздела будут использованы именованные вычисления, чтобы предоставить понятные имена для линий товаров, определена иерархия в измерении «Продукт» и указано имя элемента «(Все)» для иерархии. Также атрибуты будут сгруппированы в папки отображения.

### **Добавление именованного вычисления**

К таблице в представлении источника данных может быть добавлено именованное вычисление. В следующей задаче будет создано именованное вычисление, которое отображает полное наименование линейки продуктов.

### **Добавление именованного вычисления**

1. Чтобы открыть представление источника данных **Adventure Works DW 2012**, дважды щелкните **Adventure Works DW 2012** в папке **Представления источников данных** в обозревателе решений.
2. В нижней части панели диаграмм щелкните правой кнопкой мыши по заголовку таблицы **Продукт** и выберите команду **Создать именованное вычисление**.
3. В диалоговом окне **Создание именованного вычисления** в поле **Имя столбца** введите **ProductLineName**.
4. В поле **Выражение** введите или скопируйте и вставьте следующую инструкцию **CASE**:
5. CASE ProductLine
6. WHEN 'M' THEN 'Mountain'
7. WHEN 'R' THEN 'Road'
8. WHEN 'S' THEN 'Accessory'
9. WHEN 'T' THEN 'Touring'

10. ELSE 'Components'
11. END
- 12.

Эта инструкция для каждой строки товара в кубе создает понятные имена.

13. Нажмите кнопку **ОК**, чтобы создать именованное вычисление **ProductLineName**. Возможно, потребуется подождать.
14. В меню **Файл** выберите команду **Сохранить все**.

### Изменение свойства **NameColumn** атрибута

В конструкторе измерений откройте измерение **Product**. Для этого дважды щелкните измерение **Продукт** в узле **Измерения обозревателя решений**.

1. На панели **Атрибуты** вкладки **Структура измерения** выберите **Product Line**.
2. В окне свойств с правой стороны экрана щелкните поле свойства **NameColumn** в нижней части окна, затем щелкните по кнопке обзора (...), чтобы открыть диалоговое окно **Столбец имени**. Возможно, потребуется щелкнуть по вкладке **Свойства** в правой части окна, чтобы открыть окно «Свойства».
3. Выберите пункт **ProductLineName** внизу списка **Исходный столбец** и нажмите кнопку **ОК**.

Теперь поле **NameColumn** содержит текст **Product.ProductLineName (WChar)**. После этого элементы иерархии атрибута **Product Line** будут содержать не сокращенное, а полное наименование линейки продуктов.

4. На панели **Атрибуты** вкладки **Структура измерения** выберите **Product Key**.

5. В окне свойств щелкните поле свойства **NameColumn** и нажмите кнопку обзора (...), чтобы открыть диалоговое окно **Столбец имени**.
6. Выберите в списке **Исходный столбец** значение **EnglishProductName** и нажмите кнопку **ОК**.

Теперь поле **NameColumn** содержит текст **Product.EnglishProductName (WChar)**.

7. В окне свойств прокрутите вверх, щелкните поле свойства **Имя** и введите **Имя продукта**.

## Создание иерархии

## Создание иерархии

1. Перетащите атрибут **Product Line** из панели **Атрибуты** на панель **Иерархии**.
2. Перетащите атрибут **Model Name** из панели **Атрибуты** в ячейку **<новый уровень>** панели **Иерархии** ниже уровня **Product Line**.
3. Перетащите атрибут **Название продукта** из панели **Атрибуты** в ячейку **<новый уровень>** панели **Иерархии** ниже уровня **Имя модели**. («Ключ продукта» был переименован в «Имя продукта» в предыдущем разделе.)
4. На панели **Иерархии** вкладки **Структура измерения** щелкните правой кнопкой мыши строку заголовка окна **Иерархия**, выберите команду **Переименовать** и введите **Линейки моделей продукции**.

Теперь иерархия называется **Product Model Lines**.

5. В меню **Файл** выберите команду **Сохранить все**.

## Определение имен папок и имени элемента «Все»

### Указание имен папок и элементов

1. На панели **Атрибуты** выберите следующие атрибуты (щелкните каждый из них, удерживая нажатой клавишу CTRL).
  - **Class**
  - **Color**
  - **Days To Manufacture**
  - **Reorder Point**
  - **Safety Stock Level**
  - **Size**
  - **Size Range**
  - **Style**
  - **Weight**
2. В окне свойств в поле свойства **AttributeHierarchyDisplayFolder** введите **Stocking**.

Атрибуты сгруппированы в единую папку отображения.

3. На панели **Атрибуты** выберите следующие атрибуты.
  - **Dealer Price**
  - **List Price**
  - **Standard Cost**
4. В ячейке свойства **AttributeHierarchyDisplayFolder** окна свойств введите значение **Financial**.

Атрибуты сгруппированы во вторую папку отображения.

5. На панели **Атрибуты** выберите следующие атрибуты.
  - **End Date**
  - **Start Date**
  - **Status**
6. В ячейке свойства **AttributeHierarchyDisplayFolder** окна свойств введите **History**.

Атрибуты сгруппированы в третью папку отображения.

7. На панели **Иерархии** выберите иерархию **Product Model Lines** и для свойства **AllMemberName** в окне свойств задайте значение **All Products**.
8. Щелкните открытую область панели **Иерархии** и измените свойство **AttributeAllMemberName** в верхней части окна свойств на **Все продукты**.

Щелкнув рабочую область, можно изменять свойства самого измерения **Product**. Также можно щелкнуть значок измерения **Продукт** в начале списка атрибутов на панели **Атрибуты**.

9. В меню **Файл** выберите команду **Сохранить все**.

## **Определение связей атрибутов**

Необходимо определять связи между атрибутами, если базовые данные это поддерживают. Определение связей между атрибутами ускоряет обработку измерений, секций и запросов. Дополнительные сведения см. в разделах [Определение связей атрибутов](#) и [Связи атрибутов](#).

## **Определение связей атрибутов**

1. В окне **Конструктор измерений** для измерения Product откройте вкладку **Связи атрибутов**.
2. На диаграмме щелкните правой кнопкой мыши атрибут **Имя модели** и выберите команду **Создать связь атрибутов**.
3. В диалоговом окне **Создание связи атрибутов** поле **Исходный атрибут** имеет значение **Model Name**. Задайте для поля **Связанный атрибут** значение **Product Line**.

В списке **Тип связи** оставьте выбранным тип **Гибкая**, поскольку связи между элементами могут измениться с течением времени. Например, модель товара со временем могла быть перенесена в другую линию товаров.

4. Нажмите кнопку **ОК**.
5. В меню **Файл** выберите команду **Сохранить все**.

## **Просмотр изменений в измерении Product**

## **Просмотр изменений в измерении Product**

1. В меню **Построение** среды SQL Server Data Tools (SSDT) выберите команду **Развернуть Analysis Services Tutorial**.
2. Получив сообщение **Развертывание выполнено успешно**, перейдите на вкладку **Браузер** окна **Конструктор измерений** для измерения **Продукт** и нажмите на панели инструментов кнопку повторного соединения.
3. Убедитесь, что в списке **Иерархия** выбрана вкладка **Product Model Lines** и раскройте узел **All Products**.

Обратите внимание, что элемент **Все** отображается как **Все продукты**. Причина этого заключается в том, что свойство **AllMemberName**

иерархии ранее на этом занятии было заменено на **All Products**. Кроме того, все элементы уровня **Линейка продуктов** теперь имеют понятные имена, а не однобуквенные сокращения.

Изменение измерения DateВ этой задаче определяется пользовательская иерархия, изменяются имена элементов атрибута и используются составные ключи, чтобы указать уникальные элементы атрибутов.

В ходе выполнения задач этого раздела мы создадим определяемую пользователем иерархию и изменим имена элементов, отображаемых для атрибутов «Дата», «Месяц», «Календарный квартал» и «Календарное полугодие». Также мы определим составные ключи для атрибутов, настроим порядок сортировки элементов измерения и зададим связи атрибутов.

### **Добавление именованного вычисления**

К таблице в представлении источника данных можно добавить именованное вычисление, которое является выражением SQL и представляет собой вычисляемый столбец. Это выражение имеет вид и функции столбца таблицы. Именованные вычисления позволяют расширять реляционную схему таблиц, существующих в представлении источника данных, не изменяя таблицы в базовом источнике данных. Дополнительные сведения см. в разделе Определение именованных вычислений в представлении источника данных (службы Analysis Services).

### **Добавление именованного вычисления**

1. Чтобы открыть представление источника данных **Adventure Works DW 2012**, дважды щелкните его в папке **Представления источников данных** в обозревателе решений.
2. Ближе к нижней части панели **Таблицы** щелкните правой кнопкой мыши таблицу **Дата** и выберите команду **Создать именованное вычисление**.
3. В диалоговом окне **Создание именованного вычисления** в поле **Имя столбца** введите **SimpleDate**, а затем введите или скопируйте и вставьте следующую инструкцию **DATENAME** в поле **Выражение**:
  4. DATENAME(mm, FullDateAlternateKey) + ' ' +
  5. DATENAME(dd, FullDateAlternateKey) + ', ' +
  6. DATENAME(yy, FullDateAlternateKey)
  - 7.

Инструкция **DATENAME** извлекает из столбца FullDateAlternateKey значения числа, месяца и года. Этот новый столбец будет содержать отображаемое имя для атрибута FullDateAlternateKey.

8. Нажмите кнопку **ОК**, а затем раскройте таблицу **Date** на панели **Таблицы**.

Именованное вычисление **SimpleDate** отображается в списке столбцов таблицы «Дата» со значком, обозначающим, что это именованное вычисление.

9. В меню **Файл** выберите команду **Сохранить все**.
10. На панели **Таблицы** щелкните правой кнопкой мыши таблицу **Дата** и выберите пункт **Просмотр данных**.
11. Прокрутите вправо и просмотрите последний столбец в представлении **Просмотр таблицы «Дата»**.

Обратите внимание, что в представлении источника данных появился столбец **SimpleDate**, содержащий объединенные данные из нескольких столбцов базового источника данных. При этом исходный источник данных изменять не пришлось.

12. Закройте представление **Просмотр таблицы «Дата»**.

### **Использование именованных вычислений в качестве имен элементов**

После создания именованного вычисления в представлении источника данных это вычисление можно использовать в качестве свойства атрибута.

### **Использование именованного вычисления в качестве имен элементов**

1. Откройте измерение **Date** в **конструкторе измерений** среды SQL Server Data Tools (SSDT). Для этого дважды щелкните измерение **Date** в узле **Измерения** в **Обозревателе решений**.
2. На панели **Атрибуты** на вкладке **Структура измерения** выберите атрибут **Date Key**.
3. Если окно «Свойства» не открыто, откройте его и нажмите в строке заголовка кнопку **Автоматически скрыть**, чтобы оно оставалось открытым.
4. Ближе к нижней части окна щелкните поле свойства **NameColumn** и нажмите кнопку обзора (...), чтобы открыть диалоговое окно **Столбец имени**.
5. Выберите пункт **SimpleDate** внизу списка **Исходный столбец** и нажмите кнопку **ОК**.
6. В меню **Файл** выберите команду **Сохранить все**.

### **Создание иерархии**

Новую иерархию можно создать, перетащив атрибут с панели **Атрибуты** на панели **Иерархии**.

### Создание иерархии

1. На вкладке **Структура измерений** конструктора измерений для измерения **Дата** перетащите атрибут **Календарный год** из панели **Атрибуты** в панель **Иерархии**.
2. Перетащите атрибут **Calendar Semester** из панели **Атрибуты** в ячейку **<новый уровень>** на панели **Иерархии** под уровнем **Calendar Year**.
3. Перетащите атрибут **Calendar Quarter** с панели **Атрибуты** в ячейку **<новый уровень>** на панели **Иерархии** под уровнем **Calendar Semester**.
4. Перетащите атрибут **English Month Name** из панели **Атрибуты** в ячейку **<новый уровень>** на панели **Иерархии** под уровнем **Calendar Quarter**.
5. Перетащите атрибут **Date Key** из панели **Атрибуты** в ячейку **<новый уровень>** на панели **Иерархии** под уровнем **English Month Name**.
6. На панели **Иерархии** щелкните правой кнопкой мыши строку заголовка окна **Иерархия**, выберите команду **Переименовать** и введите **Календарная дата**.
7. Щелкнув правой кнопкой мыши по контекстному меню, в иерархии **Календарная дата** измените имя уровня **Имя месяца на английском** на **Календарный месяц**, а имя уровня **Ключ даты** — на **Дата**.
8. Удалите атрибут **Резервный ключ полной даты** на панели **Атрибуты**, поскольку он больше не понадобится. Нажмите кнопку **ОК** в окне подтверждения **Удалить объекты**.
9. В меню **Файл** выберите команду **Сохранить все**.

### Определение связей атрибутов

Необходимо определять связи между атрибутами, если базовые данные это поддерживают. Определение связей между атрибутами ускоряет обработку измерений, секций и запросов.

### Определение связей атрибутов

1. В конструкторе измерений в измерении **Date** перейдите на вкладку **Связи атрибутов**.
2. На диаграмме щелкните правой кнопкой мыши атрибут **Имя месяца на английском** и выберите команду **Создать связь атрибутов**.
3. В диалоговом окне **Создание связи атрибутов** свойство **Исходный атрибут** имеет значение **English Month Name**. Задайте для поля **Связанный атрибут** значение **Calendar Quarter**.
4. В списке **Тип связи** выберите тип связи **Жесткая**.

Связь имеет тип **Жесткая**, поскольку связи между элементами не будут меняться с течением времени.

5. Нажмите кнопку **ОК**.
6. На диаграмме щелкните правой кнопкой мыши атрибут **Календарный квартал** и выберите команду **Создать связь атрибутов**.
7. В диалоговом окне **Создание связи атрибутов** поле **Исходный атрибут** имеет значение **Calendar Quarter**. Задайте для поля **Связанный атрибут** значение **Calendar Semester**.
8. В списке **Тип связи** выберите тип связи **Жесткая**.
9. Нажмите кнопку **ОК**.
10. На диаграмме щелкните правой кнопкой мыши атрибут **Календарное полугодие**, а затем выберите команду **Создать связь атрибутов**.

11. В диалоговом окне **Создание связи атрибутов** поле **Исходный атрибут** имеет значение **Calendar Semester**. Задайте для поля **Связанный атрибут** значение **Calendar Year**.
12. В списке **Тип связи** выберите тип связи **Жесткая**.
13. Нажмите кнопку **ОК**.
14. В меню **Файл** выберите команду **Сохранить все**.

### Создание уникальных имен для элементов измерения

В этой задаче будут созданы столбцы понятных имен, которые будут использованы для атрибутов **EnglishMonthName**, **CalendarQuarter** и **CalendarSemester**.

### Создание уникальных имен для элементов измерения

1. Чтобы переключиться в представление источника данных **Adventure Works DW 2012**, дважды щелкните его в папке **Представления источников данных** в обозревателе решений.
2. На панели **Таблицы** щелкните правой кнопкой мыши таблицу **Date** и выберите команду **Создать именованное вычисление**.
3. В диалоговом окне **Создание именованного вычисления** введите **MonthName** в поле **Имя столбца**, а в поле **Выражение** введите или скопируйте и вставьте следующую инструкцию:
4. `EnglishMonthName+' '+ CONVERT(CHAR (4), CalendarYear)`
- 5.

Эта инструкция объединяет месяц и год для каждого месяца в таблице в новый столбец.

6. Нажмите кнопку **ОК**.

7. На панели **Таблицы** щелкните правой кнопкой мыши таблицу **Date** и выберите команду **Создать именованное вычисление**.
8. В диалоговом окне **Создание именованного вычисления** введите **CalendarQuarterDesc** в поле **Имя столбца**, а затем в поле **Выражение** введите или скопируйте и вставьте следующий скрипт SQL:
9. 'Q' + CONVERT(CHAR (1), CalendarQuarter) + ' ' + 'CY ' +
10. CONVERT(CHAR (4), CalendarYear)
- 11.

Этот скрипт SQL помещает в новый столбец объединение календарного квартала и года для каждого квартала в таблице.

- 12.Нажмите кнопку **ОК**.
- 13.На панели **Таблицы** щелкните правой кнопкой мыши таблицу **Date** и выберите команду **Создать именованное вычисление**.
- 14.В диалоговом окне **Создание именованного вычисления** введите **CalendarSemesterDesc** в поле **Имя столбца**, а затем в поле **Выражение** введите или скопируйте и вставьте следующий скрипт SQL:
15. CASE
16. WHEN CalendarSemester = 1 THEN 'H1' + ' ' + 'CY' + ' '
17.       + CONVERT(CHAR(4), CalendarYear)
18. ELSE
19. 'H2' + ' ' + 'CY' + ' ' + CONVERT(CHAR(4), CalendarYear)
20. END
- 21.

Этот скрипт SQL помещает в новый столбец объединение календарного полугодия и года для каждого полугодия в таблице.

- 22.Нажмите кнопку **ОК**.
- 23.В меню **Файл** выберите команду **Сохранить все**.

## Определение составного свойства **KeyColumns** и задание столбца имени

Свойство **KeyColumns** содержит столбец или столбцы, представляющие ключ для атрибута. В этой задаче будет определено составное свойство **KeyColumns**.

### Определение составного свойства **KeyColumns** для атрибута **English Month Name**

1. Откройте измерение **Date** и перейдите на вкладку **Структура измерения**.
2. На панели **Атрибуты** щелкните атрибут **English Month Name**.
3. В окне **Свойства** щелкните в поле **KeyColumns** и нажмите кнопку обзора (...).
4. В диалоговом окне **Ключевые столбцы** выберите из списка **Доступные столбцы** столбец **CalendarYear**, а затем нажмите кнопку **&gt;**.
5. Столбцы **EnglishMonthName** и **CalendarYear** отображаются в списке **Ключевые столбцы**.
6. Нажмите кнопку **ОК**.
7. Чтобы задать свойство **NameColumn** атрибута **EnglishMonthName**, щелкните поле **NameColumn** в окне свойств и нажмите кнопку обзора (...).
8. В диалоговом окне **Столбец имени**, в списке **Исходный столбец**, выберите **MonthName**, а затем нажмите кнопку **ОК**.
9. В меню **Файл** выберите команду **Сохранить все**.

### Определение составного свойства **KeyColumns** для атрибута **Calendar Quarter**

1. На панели **Атрибуты** щелкните атрибут **Calendar Quarter**.

2. В окне **Свойства** щелкните в поле **KeyColumns** и нажмите кнопку обзора (...).
3. В диалоговом окне **Ключевые столбцы** выберите из списка **Доступные столбцы** столбец **CalendarYear**, а затем нажмите кнопку **>**.

Столбцы **CalendarQuarter** и **CalendarYear** отображаются в списке **Ключевые столбцы**.

4. Нажмите кнопку **ОК**.
5. Чтобы задать свойство **NameColumn** атрибута **Calendar Quarter**, щелкните поле **NameColumn** в окне свойств и нажмите кнопку обзора (...).
6. В диалоговом окне **Столбец имени**, в списке **Исходный столбец**, выберите **CalendarQuarterDesc**, а затем нажмите кнопку **ОК**.
7. В меню **Файл** выберите команду **Сохранить все**.

### **Определение составного свойства KeyColumns для атрибута Calendar Semester**

1. На панели **Атрибуты** щелкните атрибут **Calendar Semester**.
2. В окне **Свойства** щелкните в поле **KeyColumns** и нажмите кнопку обзора (...).
3. В диалоговом окне **Ключевые столбцы** выберите из списка **Доступные столбцы** столбец **CalendarYear**, а затем нажмите кнопку **>**.

Столбцы **CalendarSemester** и **CalendarYear** отображаются в списке **Ключевые столбцы**.

4. Нажмите кнопку **ОК**.
5. Чтобы задать свойство **NameColumn** атрибута **Calendar Semester**, щелкните поле **NameColumn** в окне свойств и нажмите кнопку обзора (...).

6. В диалоговом окне **Столбец имени**, в списке **Исходный столбец**, выберите **CalendarSemesterDesc**, а затем нажмите кнопку **ОК**.
7. В меню **Файл** выберите команду **Сохранить все**.

### **Развертывание и просмотр изменений**

После изменения атрибутов и иерархий необходимо произвести развертывание произведенных изменений и повторную обработку связанных объектов, прежде чем эти изменения можно будет просмотреть.

### **Развертывание и просмотр изменений**

1. В меню **Построение** среды SQL Server Data Tools выберите команду **Развернуть Analysis Services Tutorial**.
2. Получив сообщение **Развертывание выполнено успешно**, перейдите на вкладку **Браузер** окна **Конструктор измерений** для измерения **Дата** и щелкните на панели инструментов кнопку повторного соединения.
3. В списке **Иерархия** выберите значение **Calendar Quarter**. Просмотрите элементы иерархии атрибута **Calendar Quarter**.

Обратите внимание, что имена элементов иерархии атрибута **Календарный квартал** выглядят более понятно и проще в использовании, поскольку для их отображения создано именованное вычисление. Теперь в иерархии атрибута **Календарный квартал** есть элементы для каждого квартала каждого года. Элементы не отсортированы в хронологическом порядке. Вместо этого они отсортированы сначала по кварталам, а затем по годам. В следующей задаче этого раздела требуется

изменить поведение этой иерархии атрибута, чтобы отсортировать ее элементы сначала по годам, а затем по кварталам.

4. Просмотрите элементы иерархий атрибутов **English Month Name** и **Calendar Semester**.

Обратите внимание, что элементы этих иерархий также не отсортированы в хронологическом порядке. Вместо этого они отсортированы сначала соответственно по месяцам или полугодиям, а затем — по годам. В следующей задаче этого раздела поведение этих иерархий атрибутов будет изменено, чтобы изменить порядок сортировки.

### **Изменение порядка сортировки путем изменения порядка элементов составных ключей**

В этой задаче будет изменен порядок сортировки посредством изменения порядка ключей, составляющих составной ключ.

### **Изменение порядка элементов составных ключей**

1. Открыв измерение **Дата** в конструкторе измерений, перейдите на вкладку **Структура измерения** и выберите **Календарное полугодие** на панели **Атрибуты**.
2. В окне свойств просмотрите значение свойства **OrderBy**. Оно имеет значение **Ключ**.

Элементы иерархии атрибута **Calendar Semester** отсортированы по значению ключа. При составном ключе сортировка ключей элементов ведется упорядочение сначала по значению первого ключа элемента, а затем — по значению второго ключа элемента. Иными словами, эле-

менты иерархии атрибута **Календарное полугодие** отсортированы сначала по полугодиям, а затем по годам.

3. В окне свойств нажмите кнопку обзора (...), чтобы изменить значения свойства **KeyColumns**.

4. В списке **Ключевые столбцы** в диалоговом окне **Ключевые столбцы** выберите столбец **CalendarSemester**, а затем нажмите стрелку вниз, чтобы изменить порядок элементов составного ключа. Нажмите кнопку **ОК**.

Теперь элементы этой иерархии атрибута отсортированы сначала по годам, а затем по полугодиям.

5. Выберите **Календарный квартал** на панели **Атрибуты** и нажмите кнопку обзора (...) для свойства **KeyColumns** в окне свойств.

6. В списке **Ключевые столбцы** диалогового окна **Ключевые столбцы** выделите столбец **CalendarQuarter**, а затем нажмите стрелку вниз, чтобы изменить порядок элементов составного ключа. Нажмите кнопку **ОК**.

Теперь элементы этой иерархии атрибута отсортированы сначала по годам, а затем по кварталам.

7. На панели **Атрибуты** выберите элемент **English Month Namea**, а затем в окне свойств нажмите кнопку с многоточием (...) для свойства **KeyColumns**.

8. В списке **Ключевые столбцы** диалогового окна **Ключевые столбцы** выделите столбец **EnglishMonthName**, а затем нажмите стрелку вниз, чтобы изменить порядок элементов составного ключа. Нажмите кнопку **ОК**.

Теперь элементы этой иерархии атрибута отсортированы сначала по годам, а затем по месяцам.

9. В меню **Построение** среды SQL Server Data Tools выберите команду **Развернуть Analysis Services Tutorial**. После успешного завершения развертывания в конструкторе измерений перейдите на вкладку **Браузер** и выберите измерение **Дата**.
10. На панели инструментов вкладки **Браузер** нажмите кнопку повторного соединения.
11. Просмотрите элементы иерархий атрибутов **Calendar Quarter** и **Calendar Semester**.

Обратите внимание, что теперь элементы этих иерархий атрибутов сортируются в хронологическом порядке, по годам, а затем по полугодиям или кварталам соответственно.

12. Просмотрите элементы иерархии атрибута **English Month Name**.

Обратите внимание, что теперь элементы этой иерархии атрибута сортируются сначала по годам, а затем по месяцам в алфавитном порядке. Причина этого заключается в том, что столбец EnglishCalendarMonth в представлении источника данных имеет строковый тип данных, который основан на типе данных nvarchar базовой реляционной базы данных. Сведения о сортировке по месяцам хронологически внутри года см. в разделе [Сортировка элементов атрибута по вторичному атрибуту](#).

Просмотр развернутого куба В этой задаче с помощью обозревателя в конструкторе кубов выполняется просмотр данных куба.

В этой задаче будет проводиться просмотр куба Службы Analysis Services Tutorial. Так как при проведении анализа сравниваются несколько измерений, для просмотра данных воспользуемся сводной таблицей Excel. Разместим в сводной таблице информацию о клиенте, дате и продукте на разных осях,

чтобы было видно, как изменяются продажи через Интернет в различные периоды времени, для различных групп клиентов и продуктовых линий.

### Просмотр развернутого куба

1. Чтобы перейти в конструктор кубов в среде SQL Server Data Tools, дважды щелкните куб «Учебник по **Службы Analysis Services**» в папке **Кубы** в обозревателе решений.
2. Перейдите на вкладку **Браузер** и на панели инструментов конструктора нажмите кнопку **Повторное соединение**.
3. Щелкните значок Excel, чтобы запустить программу Excel, используя в качестве источника базу данных рабочей области. Нажмите кнопку **Включить** при появлении запроса на подключение к данным.
4. В списке полей сводной таблицы разверните группу **Продажи через Интернет**, а затем перетащите показатель **Объем продаж** в область **Значения**.
5. В списке полей сводной таблицы разверните узел **Продукт**.
6. Перетащите пользовательскую иерархию **Product Model Lines** в область **Столбцы**.
7. В списке полей сводной таблицы разверните узел **Клиент**, затем узел **Местоположение** и перетащите иерархию **Customer Geography** из папки отображения расположения в область **Строки**.
8. В списке полей сводной таблицы разверните узел **Дата заказа** и перетащите иерархию **Order Date.Calendar Date** в область **Фильтр отчета**.
9. Нажмите стрелку справа от фильтра **Order Date.Calendar Date** на панели данных, снимите флажок для уровня (**Все**), последовательно разверните узлы **2006**, **H1 CY 2006** и **Q1 CY 2006**, установите флажок **Февраль 2006** и нажмите кнопку **ОК**.

На экран будут выведены продажи через Интернет по регионам и линейкам продуктов в феврале 2006 г., как показано на следующем рисунке.

The screenshot shows Microsoft Excel with a PivotTable and the PivotTable Field List. The PivotTable is structured as follows:

|                      | Mountain        | Road             | Road-150         | Road-650          | Road Итор         | Общий итор |
|----------------------|-----------------|------------------|------------------|-------------------|-------------------|------------|
| 5 Названия строк     |                 |                  |                  |                   |                   |            |
| 6 Australia          | 6799,98         | 153865,61        | 3495,491         | 157361,101        | 164161,081        |            |
| 7 Canada             | 10149,97        | 150287,34        | 699,0982         | 150986,4382       | 161136,4082       |            |
| 8 France             | 6749,98         | 21469,62         | 2097,2946        | 23566,9146        | 30316,8946        |            |
| 9 Germany            | 3374,99         | 42939,24         | 1398,1964        | 44337,4364        | 47712,4264        |            |
| 10 United Kingdom    | 13574,96        | 32204,43         | 1398,1964        | 33602,6264        | 47177,5864        |            |
| 11 United States     | 20274,94        | 75143,67         | 4893,6874        | 80037,3574        | 100312,2974       |            |
| 12 <b>Общий итор</b> | <b>60924,82</b> | <b>475909,91</b> | <b>13981,964</b> | <b>489891,874</b> | <b>550816,694</b> |            |

The PivotTable Field List on the right is configured with the following settings:

- Выберите поля для добавления в отчет: Customer, Due Date, Order Date, Order Date.Calendar Date (checked), Calendar Year.
- Фильтр отчета: Order Date.Ca...
- Названия столбцов: Product Model...
- Названия строк: Customer Geo...
- Значения: Sales Amount.
- Отложить обновление макета:  (unchecked).
- Кнопка: Обновить.