

Глава 17. Множественные таблицы данных Power Pivot

Это продолжение перевода книги Роб Колли. Формулы DAX для Power Pivot. Главы не являются независимыми, поэтому рекомендую начать сначала.

[Предыдущая глава](#) [Содержание](#) Следующая глава

До сих пор мы работали только с одной таблицей данных – Продажи [Sales]. Таблицы данных обычно представляют бизнес-процессы, и в этом случае таблица Sales описывает транзакции продаж. Очевидно, что реальный бизнес включает более одного бизнес-процесса, и транзакции собираются в несколько таблиц.

Добавим еще один бизнес-процесс – Сервисные звонки [Service Calls]. Колл-центр принимает звонки от клиентов, и фиксирует: ID клиента, дату и продолжительность звонка, ID продукта, в связи с которым был звонок, а также субъективную оценку того, как была решена проблема. Эти данные доступны нам в конце каждого дня в формате CSV (рис. 17.1; чтобы получить картинку, как на рис. 17.1 я импортировал данные из CSV-файла в Excel; для этого создал новый файл Excel и прошел по меню Данные → Получить и преобразовать данные → Из текстового/CSV-файла).

	A	B	C	D	E
1	CallDate	ProductID	CustomerID	Minutes	Resolution
2	12/18/2001	214	11116	59	Positive
3	8/11/2001	214	24917	28	Positive
4	10/15/2001	214	11369	66	Positive
5	10/24/2001	214	11210	37	Negative
6	8/21/2001	214	29005	63	Negative
7	10/4/2001	214	18599	88	Negative
8	8/23/2001	214	26583	39	Negative
9	10/27/2001	214	12143	58	Neutral

Рис. 17.1. Записи звонков

Загрузим эти данные в нашу модель. Откройте файл ch17_MultipleDataTables.xlsx, пройдите по меню Power Pivot → Управление и в окне Power Pivot пройдите по меню Главная → Получить внешние данные → Из других источников → Текстовый файл.

В нашей модели появилась таблица Service Calls (рис. 17.2).

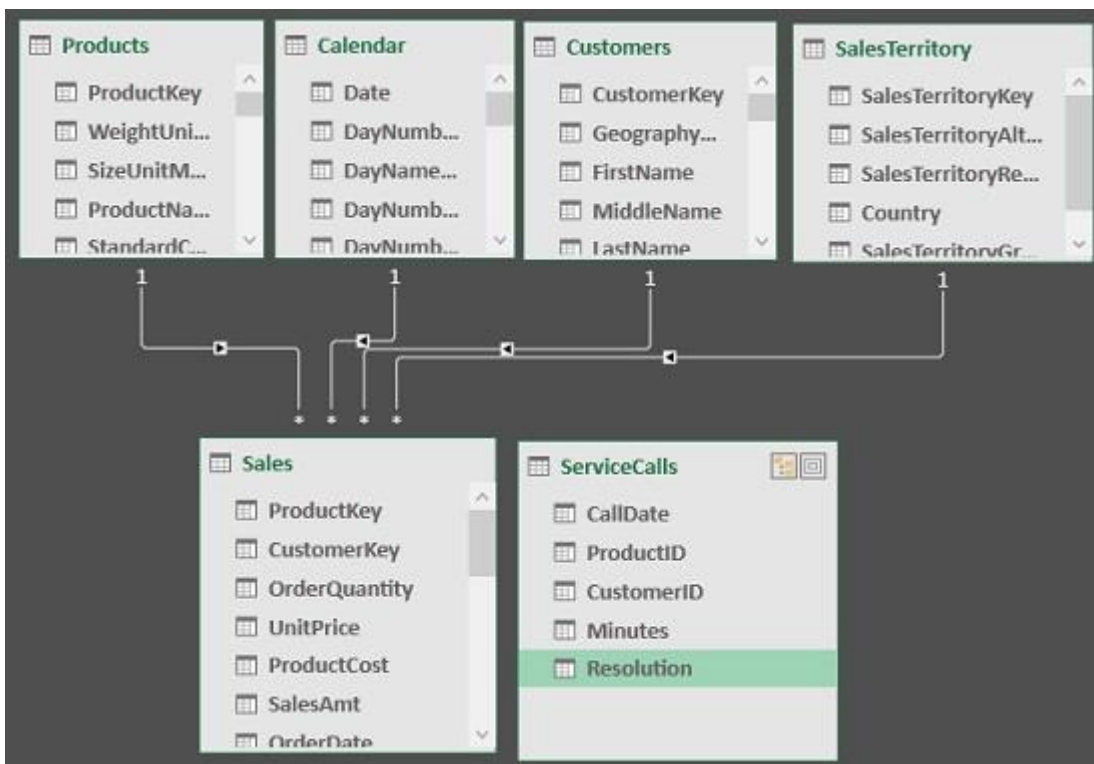


Рис. 17.2. Таблица Service Calls, импортированная в модель данных

Свяжите таблицы в верхней части модели с Service Calls (рис. 17.3).

Управление связями

Создать Изменить Удалить

Активно	Таблица 1	Мощность	Направление фильтра	Таблица 2
Да	Sales [CustomerKey]	Многие к одному (*:1)	<< K Sales	Customers [CustomerKey]
Да	Sales [OrderDate]	Многие к одному (*:1)	<< K Sales	Calendar [Date]
Да	Sales [ProductKey]	Многие к одному (*:1)	<< K Sales	Products [ProductKey]
Да	Sales [SalesTerritoryKey]	Многие к одному (*:1)	<< K Sales	SalesTerritory [SalesTerritor...]
Да	ServiceCalls [CallDate]	Многие к одному (*:1)	<< K ServiceCalls	Calendar [Date]
Да	ServiceCalls [CustomerID]	Многие к одному (*:1)	<< K ServiceCalls	Customers [CustomerKey]
Да	ServiceCalls [ProductID]	Многие к одному (*:1)	<< K ServiceCalls	Products [ProductKey]

Рис. 17.3. Связи, созданные для таблицы Service Calls

Обратите внимание, мы не создали связи между таблицами Sales и Service Calls. Вы никогда не связываете таблицы данных друг с другом! Это связано с тем, что в каждом столбце ключа таблицы данных есть дубликаты. Например, мы принимаем несколько вызовов в один день и продаем несколько продуктов в один день, поэтому таблицы Sales и Service Calls содержат несколько строк для данной даты. Всегда связывайте таблицы данных только с набором таблиц поиска, но не друг с другом (рис. 17.4).

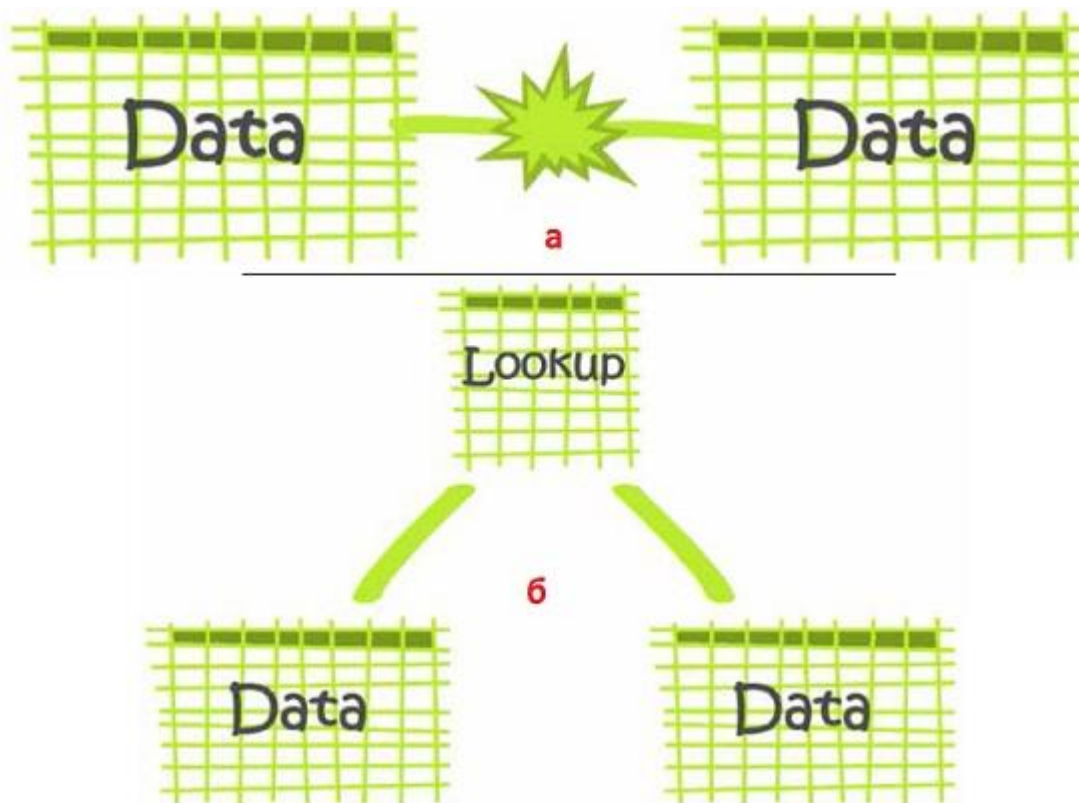


Рис. 17.4. (а) Никогда не связывайте таблицы данных друг с другом; (б) таблицы данных связывайте только с набором таблиц поиска

Определим некоторые меры:

[Calls] = COUNTROWS(ServiceCalls)

[Total Call Minutes] = SUM(ServiceCalls[Minutes])

[Avg Call Length] = DIVIDE([Total Call Minutes];[Calls])

Директор колл-центра сообщил, что минута работы колл-центра стоит \$1,75. Подсчитаем стоимость обслуживания звонков:

[Cost of Calls] = [Total Call Minutes] * 1,75

С новыми мерами наш анализ сервисных звонков становится наглядным (рис. 17. 5).

Month	Calls	Total Call Minutes	Avg Call Length	Cost of Calls
Jan	186	6 687	36,0	\$11 702
Feb	173	6 511	37,6	\$11 394
Mar	197	7 318	37,1	\$12 807
Apr	174	6 236	35,8	\$10 913
May	188	7 057	37,5	\$12 350
Jun	167	6 813	40,8	\$11 923
Jul	188	6 613	35,2	\$11 573
Aug	244	10 081	41,3	\$17 642
Sep	173	6 532	37,8	\$11 431
Oct	209	7 710	36,9	\$13 493
Nov	187	6 529	34,9	\$11 426
Dec	190	6 262	33,0	\$10 959
Grand Total	2 276	84 349	37,1	\$147 611

Рис. 17.5. Сводная таблица сервисных звонков

Объединение мер из разных таблиц данных в одной сводной

В традиционном Excel, анализ нескольких таблиц данных сводится к отображению двух сводных таблиц бок о бок (рис. 17.6).

CalendarYear	Cost of Calls	CalendarYear	Total Sales
2001	\$78 906	2001	\$3 266 374
2002	\$147 611	2002	\$6 530 344
2003	\$148 166	2003	\$9 791 060
2004	\$99 152	2004	\$9 770 900
Grand Total	\$473 834	Grand Total	\$29 358 677

Рис. 17.6. Традиционный анализ двух таблиц данных в Excel

Такое представление сталкивается с трудностями, когда строки относятся к разным продуктам/датам/покупателям (рис. 17.7).

Product Name	Cost of Calls	Product Name	Total Sales
All-Purpose Bike Stand	\$1 062	All-Purpose Bike Stand	\$20 670
Bike Wash - Dissolver	\$607	AWC Logo Cap	\$11 732
Classic Vest, L	\$635	Bike Wash - Dissolver	\$4 174
Classic Vest, S	\$1 106	Classic Vest, L	\$8 128
Fender Set - Mountain	\$382	Classic Vest, M	\$8 065
Half-Finger Gloves, L	\$310	Classic Vest, S	\$6 477
Half-Finger Gloves, M	\$578	Fender Set - Mountain	\$27 211

Рис. 17.7. Некоторые элементы отображаются в одной сводной, но не в другой

Мы часто говорим, что для пользователей Excel изучение Power Pivot – это скорее отучение, чем изучение чего-то нового. И одна из важных заповедей – не объединяйте данные в одну таблицу (что является стандартным подходом при построении обычных сводных таблиц). Несколько таблиц данных обычно представляют различные бизнес-процессы и очень непохожи друг на друга.

В модели данных меры из разных таблиц могут отражаться в одной сводной! Определенные один раз, меры используются в любом месте. Перестройка сводной не потребует переделки мер.

Последние автоматически отреагируют на новые срезы. В нашем примере вы можете разместить меры [Service Calls] и меры [Sales] в одной сводной (рис. 17.8).

	A	B	C	D
1				
2		CalendarYear	2004	
3				
4		ProductName	Total Sales	Cost of Calls
5		All-Purpose Bike Stand	\$20 670	\$1 062
6		AWC Logo Cap	\$11 732	
7		Bike Wash - Dissolver	\$4 174	\$607
8		Classic Vest, L	\$8 128	\$635
9		Classic Vest, M	\$8 065	
10		Classic Vest, S	\$6 477	\$1 106
11		Fender Set - Mountain	\$27 211	\$382
12		Half-Finger Gloves, L	\$6 661	\$310
13		Half-Finger Gloves, M	\$7 445	\$578
14		Half-Finger Gloves, S	\$6 686	\$613
15		Hitok Book - 4 Bikes	\$22 000	\$670

Рис. 17.8. Две меры, принадлежащие разным таблицам данных в одной сводной

Просто измените фильтры, и данные автоматически подстроятся (рис. 17.9).

	H	I	J	K	L	M
1						
2		Product Category		CalendarYear	2004	
3		Accessories				
4		Bikes		ProductName	Total Sales	Cost of Calls
5		Clothing		January	\$42 293	\$2 084
6		Components		February	\$40 781	\$1 573
7				March	\$43 830	\$1 078
8				April	\$47 218	\$767
9		Customer Gender		May	\$52 670	\$1 218
10		F	M	June	\$49 272	\$1 624
11				July	\$25 820	\$1 768
12				August		\$1 722
13				Grand Total	\$301 883	\$11 834

Рис. 17.9. Меры подстраиваются под изменение фильтров

Гибридные меры

Иногда мы хотим использовать меру, основанную на двух наборах данных. Например, мы можем уточнить прибыль, уменьшив ее на стоимость обслуживания звонков (рис. 17.10):

$$[\text{Net Profit}] = [\text{Profit}] - [\text{Cost of Calls}]$$

	H	I	J	K	L
1					
2	CalendarYear		2004		
3					
4	ProductName	Profit	Cost of Calls	Net Profit	
5	All-Purpose Bike Stand	\$12 939	\$1 062	\$11 877	
6	AWC Logo Cap	\$2 698		\$2 698	
7	Bike Wash - Dissolver	\$2 613	\$607	\$2 006	
8	Classic Vest, L	\$5 088	\$635	\$4 453	
9	Classic Vest, M	\$5 048		\$5 048	
10	Classic Vest, S	\$4 055	\$1 106	\$2 949	
11	Fender Set - Mountain	\$17 034	\$382	\$16 653	
12	Half-Finger Gloves, L	\$4 170	\$310	\$3 860	
13	Half-Finger Gloves, M	\$4 661	\$578	\$4 083	
14	Half-Finger Gloves, S	\$4 185	\$613	\$3 573	
15	Hitch Rack - A-Bike	\$14 348	\$679	\$13 669	

Рис. 17.10. Чистая прибыль

Теперь, когда у нас есть новая мера – [Net Profit], вам не нужны бок о бок [Profit] и [Cost of Calls]. Вы можете убрать их и сосредоточиться только на чистой прибыли. Вы можете использовать в качестве инструмента визуализации Power View, чтобы собрать быстрый отчет о прибыли (рис. 17.11).



Рис. 17.11. Чистая прибыль на дашборде, выполненном в Power View (см. файл. ch17_MultipleDataTables_Excel2013_Dashboard; к сожалению, у меня файл работал некорректно)

Обратите внимание: Power Pivot является сердцем всех продуктов Power BI, а связи и меры – ядро Power Pivot. Используя их, вы находитесь на пути к овладению миром данных. Последующая визуализация на основе Power Pivot становится простой и удобной. Используя меры, вы можете нарезать кубы, используя любой из доступных атрибутов, и быстро построить любой отчет.

Проблемы нескольких таблиц данных

Если вы создадите сводную на основе таблицы [Sales], а затем добавите из таблицы [Service Calls] меру [Calls], то результат будет неожиданным (рис. 17.12).

Названия строк	Total Sales	Calls
01.07.2001	\$14 477	7 374
02.07.2001	\$13 932	7 374
03.07.2001	\$15 012	7 374
04.07.2001	\$7 157	7 374
05.07.2001	\$15 012	7 374
06.07.2001	\$14 313	7 374
07.07.2001	\$7 856	7 374
08.07.2001	\$7 856	7 374

Рис. 17.12. Мера [Calls] возвращает неожиданный результат

Мера [Calls] выводит во всех ячейках одно и то же число – 7374, что является общей суммой всех звонков. Что случилось с обещанием «определить один раз, использовать везде»? Проблема связана с тем, что мы в область сводной таблицы *Строки* поместили столбец [Order Date] из таблицы [Sales]. Но это значение из таблицы данных [Sales], а не из таблицы поиска [Calendar] (рис. 17.13). Поэтому для меры [Calls] никакие фильтры по дате не были установлены.

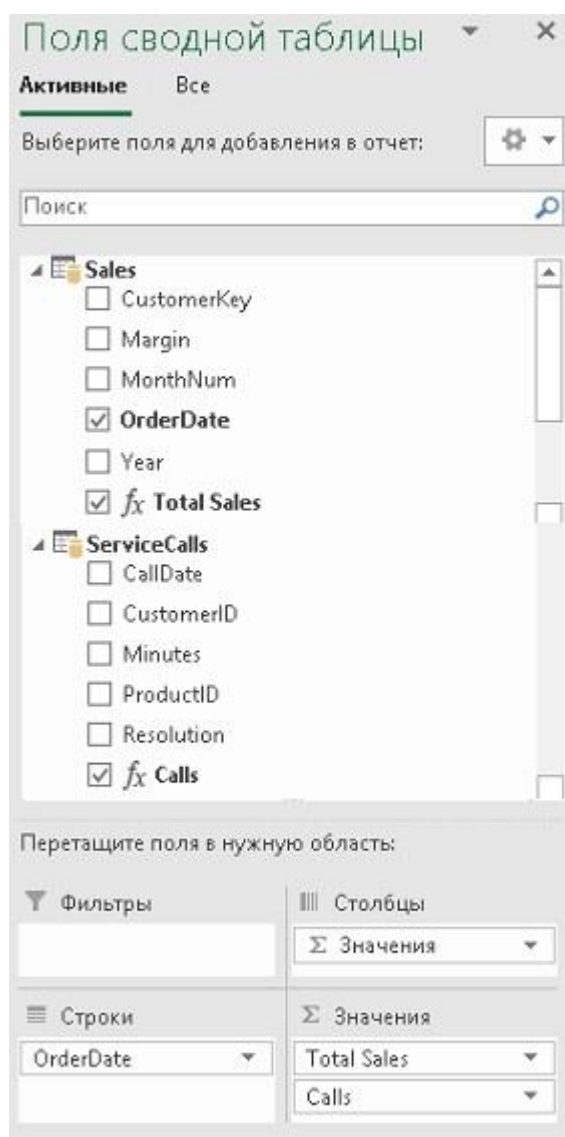


Рис. 17.13. Неверное поле в области *Строки* сводной таблицы

Запомните хорошее правило: используйте поля из таблиц поиска (а не таблиц данных) в строках, столбцах, фильтрах и срезах при анализе мер из разных таблиц данных. Это хорошая практика даже в моделях, которые имеют одну таблицу данных. Замените [Order Date] из таблицы [Sales] на [Date] из таблицы [Calendar]:

Calendar[Date]	Total Sales	Calls
01.07.2001	\$14 477	7
02.07.2001	\$13 932	4
03.07.2001	\$15 012	11
04.07.2001	\$7 157	4
05.07.2001	\$15 012	8
06.07.2001	\$14 313	14
07.07.2001	\$7 856	4
08.07.2001	\$7 856	4
09.07.2001	\$20 910	6
10.07.2001	\$10 557	4

Рис. 17.14. Мера [Calls] правильно отображаются, когда в троках находится поле [Date] из таблицы [Calendar]

Лучший способ избежать попадания в эту ловушку – скрыть некоторые из столбцов таблицы данных. Для этого войдите в модель данных, перейдите в режим *Диаграмма*, щелкните правой кнопкой мыши на поле и выберите *Скрыть от клиентских средств* (рис. 17.15).

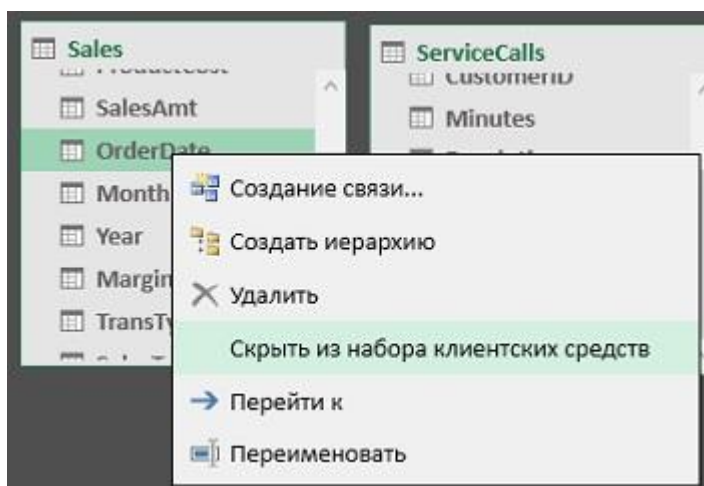


Рис. 17.15. Спрячьте поле [Order Date] из таблицы [Sales], чтобы оно стало невидимым в сводной таблице

Вот несколько рекомендаций:

- В идеале ваши **таблицы данных** должны содержать только числовые столбцы (например, количество и сумма), а также столбцы-ключи или ID (Date, ProductKey, CustomerKey), которые используются для подключения к справочникам.
- Все остальные столбцы должны «жить» в связанных **таблицах поиска**.
- Визуально проверьте таблицы данных, нет ли в них текстовых столбцов? Как правило, следует перенести такие столбцы в **таблицу поиска**.
- Таким образом, в идеальной **таблице данных** вы можете скрыть все столбцы! Ключевые столбцы (которые формируют связи) можно скрыть, так как у вас уже есть их «близнецы» в таблицах поиска. Числовые столбцы также можно скрыть, так как вы определили явные меры для них.

Таблица данных, подключенная к некоторым, но не ко всем таблицам поиска

Если вы захотите посмотреть на звонки в разрезе территорий продаж, то не увидите связи (рис. 17.16, а). И это не удивительно, так как в модели данных таблицы [Sales Territory] и [Service Calls] не связаны (рис. 17.16, б). Если наша бизнес-модель изменится, и мы начнем управлять сервисными центрами на территориальной основе, мы добавим столбец [Territory] в таблицу [ServiceCalls]. Но сейчас у нас работает единый колл-центр, который обслуживает всех наших клиентов.

	A	B	C
1			
2		Country	Calls
3		Australia	7 374
4		Canada	7 374
5		France	7 374
6		Germany	7 374
7		NA	7 374
8		United Kingd	7 374
9		United States	7 374
10		Общий итог	7 374

a

b

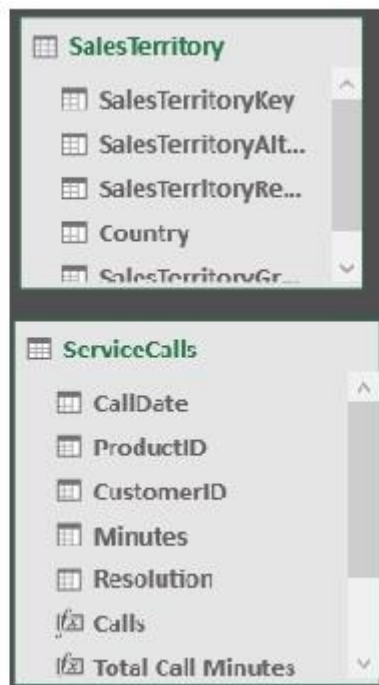


Рис. 17.16. Звонки не связаны с территориями продаж: а) ни в сводной таблице, б) ни в модели данных