

Введение в ВІ

Методичка к уроку 6

Алгоритмы математических и
статистических функций для расчета





Оглавление

Введение	4
Связи	4
Кратность и кроссфилترация	7
Основные характеристики связей	7
Кратность	7
Многие ко многим	7
Кроссфилтерация	7
Однонаправленная	7
Двунаправленная	8
Однонаправленная кроссфилтерация	8
Двунаправленная кроссфилтерация	8
DAX - Data analysis expressions	8
Что такое DAX?	9
Синтаксис DAX	11
Быстрые меры	13



Введение

На этой лекции вы познакомитесь с методами создания связей в разных таблицах с целью объединять данные по общему признаку в столбцах. В рамках работы в Power BI настройка связей позволяет пользователю не тратить время на сведение данных в одну таблицу, то есть осуществлять набор данных. И также вы познакомитесь с базовым функционалом DAX: синтаксис, популярные формулы, быстрые меры.

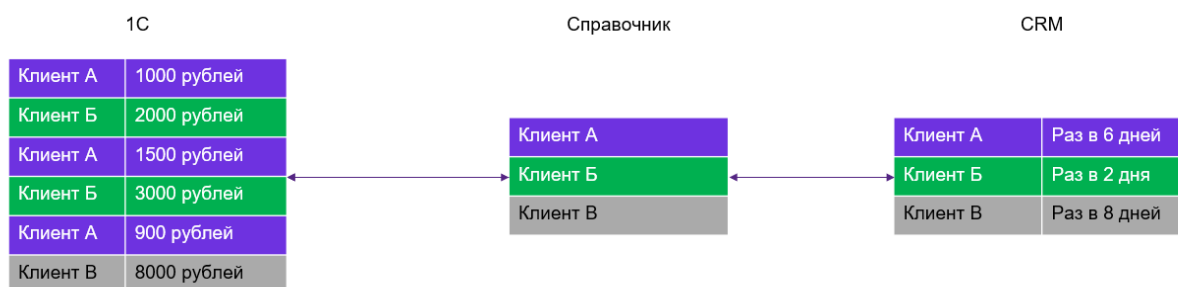
Связи

Связи позволяют объединять таблицы по общему признаку. Настраиваются для конкретного столбца в каждой из связанных таблиц.

Пример: перед вами задача — построить простой прогноз продаж на будущий год по текущим клиентам. У вас есть данные о фактических продажах из 1С и клиентские данные из CRM. Для задачи вы возьмёте из 1С данные по среднему чеку за период, а из CRM — данные о частоте покупок.

Чтобы связать эти данные, проще всего сделать отдельную таблицу с уникальными названиями клиентов и связать эту таблицу с двумя другими.

У нас получится единая база для анализа из трёх таблиц, где одна вспомогательная.



Важно! После настройки связей вы сможете работать с данными в двух таблицах так, как будто это одна таблица.

Связи в Power BI:

- автоматически ищутся;
- найденная автоматически связь делается активной;
- чем-то похоже на функцию ВПР / VLOOKUP в Excel.

В параметрах можно настроить автоматический поиск связей:

Параметры

ГЛОБАЛЬНЫЕ

- Загрузка данных
- Редактор Power Query
- DirectQuery
- Создание R-скриптов
- Создание скриптов Python
- Безопасность
- Конфиденциальность
- Региональные настройки
- Обновления
- Данные об использовании
- Диагностика
- Предварительные версии во...
- Автовосстановление
- Параметры отчетов

ТЕКУЩИЙ ФАЙЛ

- Загрузка данных
- Региональные настройки
- Конфиденциальность
- Автовосстановление

Определение типов

- ☒ Определять типы и заголовки столбцов для неструктурированных источников

Связи

- ☒ Импорт связей из источников данных при первой загрузке ⓘ
- ☐ Обновление или удаление связей при обновлении данных ⓘ
- ☒ Автоматически искать новые связи после загрузки данных ⓘ

[Дополнительные сведения](#)

Логика операций со временем

- ☒ Автоматические дата и время ⓘ [Дополнительные сведения](#)

Фоновые данные

- ☒ Разрешить скачивание в фоновом режиме для предварительного просмотра данных

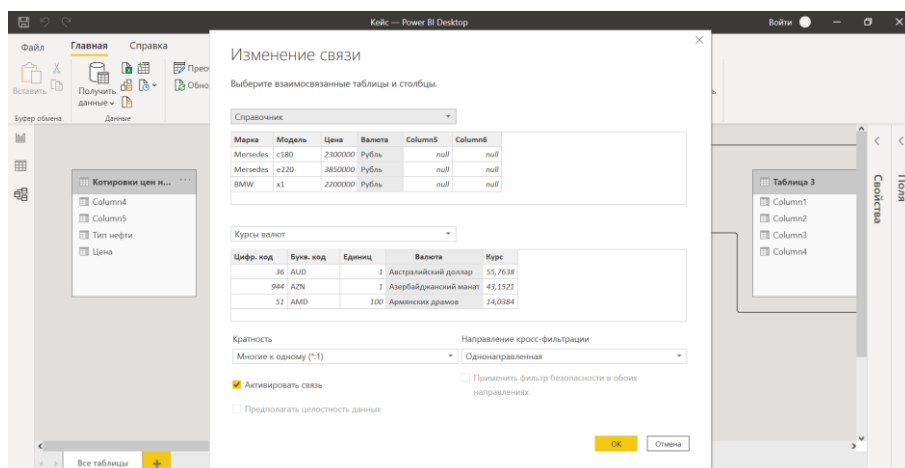
Параллельная загрузка таблиц

- ☒ Включить параллельную загрузку таблиц ⓘ

Вопросы и ответы

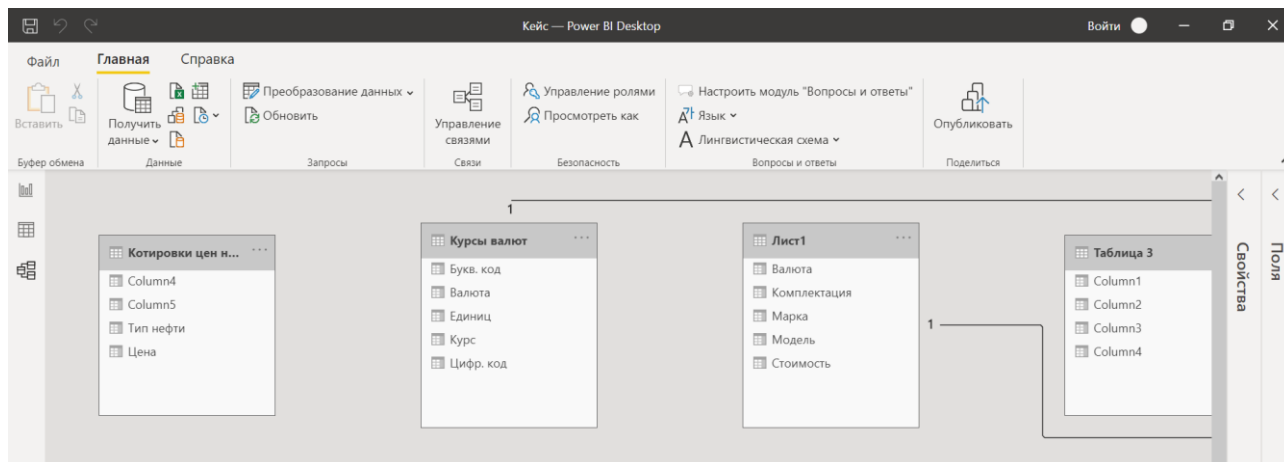
- ☒ Включить функцию "Вопросы и ответы", чтобы задавать вопросы о данных на естественном языке ⓘ [Дополнительные сведения](#)

Настройка связи выглядит следующим образом:





Готовая настроенная связь выглядит так:





Кратность и кроссфилтрация

Основные характеристики связей

Кратность

Многие к одному (*:1) — наиболее распространённый тип кратности. У первой таблицы несколько значений, у второй — только уникальное значение в нужном столбце

Один к одному (1:1) — у обеих таблиц в связанных столбцах только уникальные значения

Один ко многим (1:*) — аналогичная первой связи, только наоборот

Многие ко многим (*:*) — в обоих столбцах неуникальные значения.

Многие ко многим

Сценарий применения связи "многие ко многим" предполагает связывание двух таблиц фактов. Две таблицы фактов можно связать напрямую. Такой прием может быть полезен для быстрого изучения данных.

Основной минус многие ко многим - это отсутствие гибкости.

Как правило многие ко многим применяется редко и не рекомендуется связывать две таблицы фактов напрямую, используя кратность "многие ко многим". Главная причина в том, что модель будет недостаточно гибкой в плане фильтрации и группировки визуальных элементов в отчете.

Вместо того чтобы связывать таблицы фактов напрямую, мы рекомендуем применять принципы проектирования на основе схемы типа "звезда". Для этого нужно добавить таблицы измерений. Затем они связываются с таблицами фактов связями "один ко многим". Такой подход к проектированию эффективнее, так как делает отчеты более гибкими. Он позволяет выполнять фильтрацию и группирование по любым столбцам измерений, а также суммировать любую связанную таблицу фактов.



Кроссфилترация

Однонаправленная

Стандартный метод, когда филтрация происходит в таблице, где значения агрегируются

По умолчанию лучше делать связь однонаправленной

Быстрее работает

Двунаправленная

Обе таблицы обрабатываются как одна

Актуально для таблиц измерений

Двунаправленную связь нужно настраивать осознанно, когда вы понимаете зачем

Эта настройка влияет на то, какая из таблиц будет филтровать связанную таблицу на уровне визуала.

Однонаправленная кроссфилтрация

Это наиболее распространенное стандартное направление означает, что филтрация вариантов в подключенных таблицах выполняется в таблице, где значения агрегируются. При импорте модели данных Power Pivot в Excel или более ранней версии все связи будут однонаправленными. Из SQL импорт будет синхронным с источником.

Двунаправленная кроссфилтрация

При филтрации обе таблицы обрабатываются так же, как одна.

Параметр Оба хорошо работает с одной таблицей, которая содержит множество таблиц подстановки, которые его окружают. Примером может служить таблица фактических данных по продажам с таблицей подстановки для отделов. Такая конфигурация часто называется схемой типа "звезда" (центральная таблица с несколькими таблицами подстановки).



При наличии нескольких таблиц, у которых также есть таблицы подстановки (некоторые из которых являются общими), не следует использовать параметр Двухнаправленная. Продолжая предыдущий пример, предположим, что в этом случае также есть таблица бюджета, содержащая целевой бюджет для каждого отдела. Таблица отделов подключена к таблице продаж и таблице бюджета. Избегайте направления связи Двухнаправленная для конфигурации такого вида.

DAX - Data analysis expressions

Что такое DAX?

DAX - это Data Analysis Expressions (Выражения анализа данных)

- Это библиотека функций и операторов, которые можно комбинировать для создания формул и выражений в моделях данных Power BI, Analysis Services и Power Pivot в Excel. Это коллекция функций, операторов и констант, которые можно использовать в формуле или выражении для вычисления и возврата одного или нескольких значений.
- Алгоритм функций, похожий на Excel и SQL
- По DAX существует много качественной документации, каждая формула разобрана на сайте Майкрософт
- Многие функции доступны из коробки
- Одно или несколько выражений DAX, используемых для определения вычисления модели, называется формулой (запросом).
- В DAX можно делать как столбцы, так и виртуальные меры

Что такое Меры и чем они отличаются от столбцов?

Что такое меры?



Мера - это виртуальная расчетная единица, которая отображается в представлении отчетов или представлении данных, но не видна в ячейке или столбца таблицы. Меры, которые создали пользователи, отображаются в списке «Поля» со значком калькулятора:

Разница между столбцами и мерами:

Вычисляемые столбцы	Меры
хранятся в физической и оперативной памяти	Рассчитываются в моменте использования
Рассчитывается только один раз, на значение не влияют примененные фильтры	Рассчитываются в зависимости от контекста и фильтров
Можно использовать для фильтров	Нельзя (ну почти нельзя)
Нужны для качественных метрик	Нужны для количественных метрик

- У мер и столбцов можно задавать формат для отображения данных
- Аналоги ВПР в Excel — это RELATED и LOOKUPVALUE
- RELATED — возвращает значение из указанного столбца связанной таблицы.

Особенности: требует наличия связи между таблицами.

Плюсы: очень проста в использовании, не перегружает синтаксис. Удобна для поиска по единственному значению.

Минусы:

- сканирует всю таблицу, соответственно, на больших объёмах замедляет быстроедействие
- при поиске значения пренебрегает наложенными фильтрами
- требует однозначный контекст строки
- не может использоваться для поиска данных в таблицах с ограниченной связью



- LOOKUPVALUE — полноценный ВПР, ищет по широкому набору условий, позволяет задавать альтернативный результат.
- CALCULATE — аналог СУММЕСЛМН. Одна из самых популярных функций, невероятно гибкая и обманчиво простая в использовании. Позволяет задавать вычисление и указывать, какие фильтры применять.

Поддерживает типы фильтров:

- логические
- фильтры таблиц
- модификацию фильтров

Особенности: позволяет переводить контекст строки в контекст фильтра, это необходимо для использования в вычислениях-итераторах.

- DISTINCT — выводит только уникальные значения из столбца
- DISTINCTCOUNT — считает количество уникальных значений в столбце

На этой лекции вы начнете знакомство с библиотекой DAX (Data Analysis Expressions) - коллекцией функций и операторов, которые можно комбинировать для создания формул и выражений в моделях данных Power BI, Analysis Services и Power Pivot в Excel.

Синтаксис DAX

Синтаксис

Синтаксис — это правила записи формулы, включает различные элементы, которые составляют формулу.

1 — Имя меры, куда будет записан ответ.

2 — Оператор знака равенства (=), который обозначает начало формулы.

3 — Функция DAX.



4 — В скобки () заключается выражение, содержащее один или несколько аргументов.

5 — Имя таблицы, из которой берутся значения.

6 — Имя столбца, из которого берутся данные для вычислений.

Редактор DAX в Power BI Desktop помогает создавать синтаксически верные формулы, подсказывая подходящие элементы.

Функции

Функции являются предопределёнными формулами, которые выполняют вычисления с использованием специальных значений, именуемых аргументами, в определённом порядке или структуре.

Аргументами могут быть другие функции, другие формулы (запросы), выражения, ссылки на столбцы, числа, текст, логические значения, такие как TRUE и FALSE или константы.

Чтобы добавить в формулу функцию, нужно начать вводить её название.

Формулы DAX могут содержать до 64 вложенных функций.

Категории функций

В DAX имеются следующие категории функций:

- └ Дата и время.

- └ Логика операций со временем.

- └ Информационные.

- └ Логические.



- └ Математические.
- └ Статистические.
- └ Текстовые.
- └ Родительские/дочерние.
- └ Прочие.

Если вы знакомы с функциями в формулах Excel или SQL, многие из функций в DAX будут казаться вам аналогичными.

Операторы DAX

В DAX операторы используются для создания выражений, которые сравнивают значения, выполняют арифметические вычисления или работают со строками.

Существует четыре типа вычислительных операторов:

- └ арифметические: +, −, *, /, ^
- └ сравнения: >, <, >=, <=, <>, ==, =
- └ объединения текста: &
- └ логические: &&, ||, IN.

VAR

Служебное слово или специальная функция.

Сохраняет результат выражения как именованную переменную, которую затем можно передать в качестве аргумента в другие выражения мер. Вычисленные итоговые значения для выражения переменной не изменяются, даже если на переменную существует ссылка в другом выражении.



Также часто взаимодействует со служебным словом Return, которая заканчивает создание блока переменных и после нее уже идет основной код формулы.

Быстрые меры

Быстрые меры - это преднастроенные меры, которые можно сделать в формате по code/self-service и не писать формулу. При этом формула генерируется автоматически после создания быстрой меры.

Среди быстрых мер есть как часто используемые, так и практически не имеющие широкого применения.

Быстрые меры поделены на 6 категорий: агрегировать по категориям, Фильтры, Операции со временем, Итоги, Математические операции и Текст.