

УДК 004.424
ББК 32.372
М314

М314 Даниил Маслюк

Анализируем данные в Microsoft Power BI. Подготовка к экзамену DA-100/
пер. с англ. А. Ю. Гинько – М.: ДМК Пресс, 2022. – 332 с.: ил.

ISBN 978-5-97060-849-4

В книге рассматриваются все основные темы для подготовки к сертификационному экзамену Microsoft DA-100, посвященному анализу данных при помощи Microsoft Power BI. Материал удобно структурирован: каждая глава охватывает одну тему, в каждом разделе рассматривается конкретный навык. Освещаются вопросы получения данных, их подготовки, преобразования, моделирования и визуализации средствами Power BI Desktop, создания отчетов и дашбордов, углубленного анализа данных и др.

Издание пригодится профессионалам в области бизнес-аналитики, аналитикам данных и специалистам по проектированию и созданию отчетов, желающим проверить или подтвердить свои навыки.

УДК 004.424
ББК 32.372

Original English language edition published with the authorization of Microsoft Corporation by: Pearson Education, Inc. Copyright © 2021 by Pearson Education, Inc. All rights reserved. Russian-language edition copyright © 2021 by DMK Press. All rights reserved.

Все права защищены. Любая часть этой книги не может быть воспроизведена в какой бы то ни было форме и какими бы то ни было средствами без письменного разрешения владельцев авторских прав.

Материал, изложенный в данной книге, многократно проверен. Но, поскольку вероятность технических ошибок все равно существует, издательство не может гарантировать абсолютную точность и правильность приводимых сведений. В связи с этим издательство не несет ответственности за возможные ошибки, связанные с использованием книги.

ISBN 978-0-13-681968-4 (англ.)
ISBN 978-5-97060-849-4 (рус.)

© Pearson Education, Inc., 2021
© Оформление, перевод на русский язык,
издание, ДМК Пресс, 2022

Оглавление

Предисловие от издательства	13
Благодарности	14
Об авторе	15
Введение	16
Структура книги	16
Подготовка к экзамену	17
Сертификация Microsoft	17
Сопроводительные файлы	17
Список ссылок на источники в сети	18
Ошибки, обновления и поддержка	18
Оставайтесь на связи	18
Глава 1. Подготовка данных	19
Навык 1.1. Получение данных из различных источников	19
Распознавание источника данных и подключение к нему	20
Редактор Power Query	23
Зависимости запросов	24
Изменение настроек источника данных	25
Выбор между использованием общего набора данных и созданием локального	26
Выбор режима хранения	28
Режим импорта данных	28
Режим DirectQuery	29
Режим Live Connection	30
Составные модели	30
Выбор подходящего типа запроса	32
Последствия использования режима DirectQuery	32
Поиск проблем с производительностью запросов	35
Просмотр машинных запросов	35
Диагностика запросов	37
Добавочное (инкрементное) обновление	39
Использование облачного хранилища Microsoft Dataverse	40
Применение параметров	40
Создание параметров	41
Использование параметров	42
Редактирование параметров	44
Создание функций	45
Использование и создание файла PBIDS	46

Использование и создание потока данных	47
Подключение к набору данных с использованием конечных точек XMLA.....	48
Навык 1.2. Профилирование данных.....	50
Распознавание аномалий в данных.....	50
Оценка структуры данных и извлечение свойств столбцов	52
Получение статистических сведений о данных	53
Навык 1.3. Очистка, преобразование и загрузка данных.....	55
Поиск проблем с качеством информации и противоречивостью данных и их устранение путем замены значений.....	56
Замена значений.....	56
Удаление строк.....	58
Поиск первопричины ошибки	58
Определение и преобразование типов данных столбцов.....	59
Использование локали	62
Определение и создание подходящих ключей для объединения	63
Ключи для объединения таблиц в Power Query	63
Ключи для создания связей.....	64
Применение преобразований к структуре таблиц	65
Работа с шагами запроса	67
Уменьшение количества строк и столбцов.....	68
Сведение, отмена свертывания и транспонирование	71
Добавление столбцов.....	73
Комбинирование запросов.....	78
Добавление.....	78
Объединение.....	80
Соглашение об именовании столбцов и запросов	83
Использование расширенного редактора для написания кода на языке M в Power Query	84
Строка формул.....	86
Настройка загрузки данных	88
Решение проблем с импортом данных.....	89
Заключение	91
Мысленный эксперимент.....	93
Ответы.....	95
Глава 2. Моделирование данных	96
Навык 2.1. Проектирование модели данных.....	96
Определение таблиц.....	97
Плоская схема	97
Схема «звезда»	98
Схема «снежинка»	99
Настройка свойств таблиц и столбцов.....	101
Свойства таблиц.....	101
Свойства столбцов.....	102

Определение быстрых мер.....	103
Выравнивание иерархий типа родитель–потомок.....	107
Определение ролевых измерений.....	110
Определение кратности связей и настройка двунаправленной перекрестной фильтрации.....	114
Кратность связей.....	115
Направление кросс-фильтрации.....	116
Оптимизация модели данных.....	119
Режим хранения данных.....	119
Организация связей.....	120
Создание агрегаций.....	120
Настройка кратности связей.....	120
Установка связей типа «многие ко многим».....	121
Прямые связи типа «многие ко многим».....	122
Связи типа «многие ко многим» с использованием таблиц-мостов.....	123
Создание общей таблицы дат.....	125
Создание календаря в Power Query.....	126
Создание календаря при помощи DAX.....	127
Определение приемлемого уровня гранулярности данных.....	128
Навык 2.2. Разработка модели данных.....	131
Установка направления кросс-фильтрации и применение фильтра безопасности.....	131
Создание вычисляемых таблиц.....	132
Клонирование таблиц.....	132
Создание таблицы на основе данных из разных источников.....	133
Предварительный расчет мер для повышения эффективности отчетов.....	134
Создание иерархий.....	134
Создание вычисляемых столбцов.....	135
Применение ролей безопасности на уровне строк.....	138
Создание ролей в Power BI Desktop.....	138
Выбор роли в Power BI Desktop.....	141
Динамическая безопасность на уровне строк.....	142
Настройка инструмента Вопросы и ответы (Q&A)	145
Обучение визуального элемента Вопросы и ответы	147
Синонимы.....	149
Навык 2.3. Создание мер с использованием языка DAX.....	150
Использование языка DAX для написания сложных мер.....	151
Использование функции CALCULATE для манипулирования фильтрами.....	155
Добавление фильтров.....	155
Удаление фильтров.....	156
Изменение фильтров.....	157
Преобразование контекста.....	158
Логика операций со временем с применением DAX.....	161
Замена числовых столбцов на меры.....	164

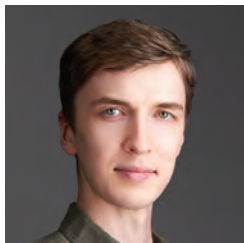
Использование базовых статистических функций для обогащения данных	165
Создание полуаддитивных мер	166
Навык 2.4. Оптимизация производительности модели данных.....	169
Удаление ненужных строк и столбцов	169
Удаление ненужных строк.....	169
Удаление ненужных столбцов.....	170
Определение мер, визуальных элементов и связей, негативно влияющих на производительность.....	170
Улучшение ситуации с кратностью столбцов путем изменения типов данных.....	172
Улучшение ситуации с кратностью столбцов при помощи суммирования данных	173
Создание и управление агрегациями.....	174
Заключение	176
Мысленный эксперимент.....	178
Ответы.....	181
Глава 3. Визуализация данных	184
Навык 3.1. Создание отчетов	184
Добавление элементов визуализации в отчет.....	185
Выбор подходящего типа элемента визуализации	186
Линейчатая диаграмма.....	187
График и диаграмма с областями.....	188
Комбинированная диаграмма	189
Ленточная диаграмма	190
Каскадная диаграмма	191
Воронка	192
Точечная диаграмма	193
Круговая и кольцевая диаграммы	194
Диаграмма дерева.....	195
Карта	196
Датчик	197
Карточка, многострочная карточка и KPI.....	198
Визуальные элементы с искусственным интеллектом.....	199
Форматирование и настройка визуализаций	200
Импортирование пользовательских визуализаций.....	200
AppSource.....	201
Из файла	202
Настройка условного форматирования.....	202
Применение срезов и фильтров.....	204
Срезы.....	204
Панель фильтров	206
Добавление визуальных элементов R и Python.....	208
Настройка страницы отчета.....	211
Настройка специальных возможностей в отчетах.....	212
Имена страниц, заголовки и метки.....	213

Маркеры.....	214
Альтернативный текст.....	214
Порядок табуляции.....	214
Тема отчета и выбор цвета.....	215
Настройка автоматического обновления страниц.....	216
Фиксированный интервал.....	217
Обнаружение изменений.....	217
Создание отчетов с разбивкой на страницы.....	219
Навык 3.2. Создание дашбордов.....	221
Управление плитками на дашборде.....	221
Настройка мобильного отображения.....	223
Конфигурирование оповещений о данных.....	225
Использование визуального элемента Вопросы и ответы	226
Добавление темы дашборда.....	227
Закрепление динамической страницы отчета на дашборде.....	228
Навык 3.3. Повышение удобства использования отчетов.....	229
Настройка закладок.....	230
Создание пользовательских подсказок.....	233
Настройка и конфигурирование взаимодействий между визуальными элементами.....	236
Настройка навигации по отчету.....	237
Применение сортировки.....	238
Синхронизация срезов.....	240
Использование панели Выбор	242
Применение детализации и кросс-филтрации.....	243
Углубление в данные с использованием интерактивных визуальных элементов.....	245
Экспорт данных отчета.....	247
Разработка отчетов для мобильных устройств.....	247
Заключение.....	249
Мысленный эксперимент.....	252
Ответы.....	253
Глава 4. Анализ данных.....	255
Навык 4.1. Повышение эффективности отчетов за счет выделения важных аналитических выводов.....	255
Применение условного форматирования.....	256
Таблицы и матрицы.....	256
Другие визуальные элементы.....	260
Анализ ведущих N элементов.....	261
Фильтр уровня визуализации.....	261
Вопросы и ответы.....	261
DAX.....	262

Исследование статистической сводки	263
Добавление результатов краткой аналитики (Quick Insights) на дашборд	265
Создание опорных линий с помощью панели Аналитика (Analytics).....	266
Использование оси воспроизведения (Play Axis) и проведение анализа на основе временных рядов	268
Персонализация визуальных элементов.....	270
Навык 4.2. Выполнение углубленного анализа.....	271
Определение выбросов в данных.....	272
Группирование данных и разделение на ячейки	274
Группирование данных.....	274
Разделение на ячейки	276
Использование элемента Ключевые факторы влияния для анализа по измерениям	277
Применение дерева декомпозиции для анализа мер	279
Применение аналитики ИИ	280
Заключение	282
Мысленный эксперимент.....	284
Ответы.....	285
Глава 5. Развертывание и поддержка	287
Навык 5.1. Управление наборами данных.....	287
Настройка запланированного обновления набора данных	288
Доступ к локальным данным.....	288
Настройка запланированного обновления	290
Настройка членства в группах безопасности на уровне строк	291
Присвоение ролей в службе Power BI.....	291
Просмотр в качестве роли в службе Power BI	293
Обеспечение доступа к наборам данных	294
Доступ через рабочую область.....	294
Доступ через приложение.....	294
Управлением разрешениями наборов данных	295
Анализ влияния.....	297
Настройка инкрементного обновления наборов данных.....	297
Создание параметров RangeStart и RangeEnd.....	298
Фильтрация с использованием созданных параметров.....	299
Определение политики инкрементного обновления.....	301
Свертывание запросов	302
Рекомендация и сертификация содержимого в Power BI.....	303
Настройка формата хранения крупных наборов данных.....	305
Навык 5.2. Создание и управление рабочими областями.....	307
Создание и настройка рабочей области	308
Рекомендованная стратегия жизненного цикла разработки	310
Параметры	310
Конвейеры развертывания	311

Присвоение ролей рабочей области.....	313
Настройка и публикация приложения рабочей области.....	314
Установка.....	315
Переходы.....	316
Разрешения.....	317
Вид приложения.....	317
Обновление опубликованного приложения.....	318
Отмена публикации приложения.....	319
Публикация, импорт и обновление элементов рабочей области.....	319
Применение меток конфиденциальности к содержимому рабочей области.....	320
Настройка подписок.....	321
Подписка на содержимое.....	321
Управление подписками.....	323
Заключение.....	323
Мысленный эксперимент.....	325
Ответы.....	327
Предметный указатель.....	328

Об авторе



Даниил Маслюк (Daniil Maslyuk) является независимым консультантом в области бизнес-аналитики, а также занимается преподавательской деятельностью и выступает в качестве спикера на мероприятиях, посвященных Microsoft Power BI. Даниил ведет блог на сайте xxlbi.com и Twitter: [@DMaslyuk](https://twitter.com/DMaslyuk).

Введение

Экзамен DA-100: Analyzing Data with Microsoft Power BI, как понятно из названия, посвящен анализу данных при помощи Microsoft Power BI. Примерно четверть экзамена отводится на подготовку данных, включая их получение из различных источников, профилирование, очистку, преобразование и загрузку. Около 30% вопросов на экзамене касается моделирования данных. Сюда относится проектирование модели, ее разработка и оптимизация. Еще треть экзамена посвящена визуализации и анализу данных, включая создание отчетов и дашбордов, а также проведение углубленного анализа данных. Ну, и остаток экзамена и книги, которую вы держите в руках, будет отведен на обсуждение вопросов, связанных с управлением наборами данных и рабочими областями в службе Power BI.

Экзамен DA-100 нацелен на профессионалов в области бизнес-аналитики, аналитиков данных и специалистов по проектированию и созданию отчетов, желающих проверить и подтвердить свои навыки в анализе данных при помощи Power BI. Для успешной сдачи экзамена необходимо хорошо разбираться в вопросах получения данных, их подготовки, преобразования, моделирования и визуализации средствами Power BI Desktop, а также уметь размещать отчеты и дашборды в общем доступе.

В данной книге мы подробно обсудим все основные темы для подготовки к экзамену. В то же время вы не встретите в ней конкретных экзаменационных вопросов, доступ к которым есть только у экзаменационной команды Microsoft. К тому же список актуальных вопросов постоянно обновляется и дополняется, в связи с чем осветить все из них просто не представляется возможным. Эту книгу скорее стоит рассматривать в качестве дополнения к другим обучающим материалам и вашему жизненному опыту в области анализа данных. При необходимости самостоятельно узнать больше о той или иной теме вы можете обращаться к соответствующим ссылкам в секциях «Дополнительная информация». Кроме того, массу полезной информации можно найти на сайтах MSDN и TechNet, а также в блогах и на форумах.

Структура книги

Данная книга организована согласно списку проверяемых на экзамене навыков. Со списком навыков для каждого экзамена можно ознакомиться на сайте Microsoft Learn по адресу <http://aka.ms/examlist> в разделе **Skills measured**. Каждая глава книги посвящена отдельной теме из этого списка, а разделы внутри главы описывают технические нюансы реализации того или иного аспекта темы. К примеру, если экзамен включает в себя шесть главных тем, в книге будет шесть глав.

Подготовка к экзамену

Успешная сдача сертификационного экзамена Microsoft – это прекрасный способ дополнить свое резюме и заявить окружающим о своем уровне знаний в конкретной области. Экзамен в полной мере подтверждает ваши знания и опыт. И хотя практический опыт ничто заменить не может, подготовка к экзамену путем освоения материала в теории и проверки его на практике вполне может принести успех. Данная книга не ставит себе целью научить вас новым навыкам.

Лучше всего готовиться к экзамену комплексно, изучая все доступные материалы и проходя курсы. Например, вы можете читать данную книгу и другие пособия и одновременно с этим пройти обучение на курсах *Microsoft Official Curriculum*. Выберите оптимальную для себя комбинацию обучающих материалов. Подробнее о курсах и обучающих мероприятиях от Microsoft вы можете узнать по адресу <http://microsoft.com/learn>. Практические тесты для сертификационных экзаменов от Microsoft по различным дисциплинам можно найти по следующей ссылке: <http://aka.ms/practicetests>.

Обратите внимание, что данная книга написана на основании информации, находящейся в общем доступе, и личного опыта автора. Из соображений профессиональной этики автор книги не имел доступа к актуальным экзаменационным вопросам.

Сертификация Microsoft

Наличие сертификата Microsoft выделяет вас из общей массы специалистов, свидетельствуя об обширном наборе знаний и опыта в работе с текущей линейкой продуктов и технологий от Microsoft. Экзамены и соответствующие им сертификаты разработаны с целью подтвердить ваш статус разработчика локальных и облачных решений на базе продуктов и технологий Microsoft. Сертификат дает большие преимущества как сотруднику, так и работодателю и компании в целом.

ИНФОРМАЦИЯ. Все сертификаты Microsoft

Полный список сертификатов Microsoft и дополнительную информацию о них можно найти на странице www.microsoft.com/learn.

Список доступных сертификатов Microsoft постоянно обновляется, следите за информацией на сайте.

Сопроводительные файлы

Большинство глав из этой книги включает в себя практические упражнения на закрепление полученных знаний. Сопроводительный файлы для упраж-

нений можно скачать по адресу <http://MicrosoftPressStore.com/ExamRefDA100PowerBI/downloads>.

К книге прилагаются два типа файлов.

1. Файлы источников, необходимые для работы в Power Query:
 - папка **Targets**;
 - файл *Inventory.xlsx*;
 - файл *WideWorldImporters.xlsx*.
2. Папка Power BI, содержащая файлы с расширением *PBIX*.

В упражнениях мы будем предполагать, что вы загрузили сопроводительные файлы в папку *C:\DA-100*.

Список ссылок на источники в сети

На протяжении книги автор рекомендует обращаться за дополнительной информацией в интернете. Ссылки на некоторые из этих источников очень длинные, и мы решили сократить их для удобства использования. Кроме того, мы собрали весь список в один файл PDF, к которому вы можете обращаться в процессе чтения книги.

Скачать список можно по ссылке <http://MicrosoftPressStore.com/ExamRefDA100PowerBI/downloads> в разделе **Downloads**.

Ссылки организованы по главам и заголовкам. Каждый раз, когда вы встречаете ссылку в книге, вы можете найти в списке соответствующий адрес и открыть его.

Ошибки, обновления и поддержка

Мы сделали все возможное, чтобы ошибки и опечатки не пробрались в книгу и сопроводительные файлы. Вы можете получить доступ ко всем исправлениям в книге по ссылке <http://MicrosoftPressStore.com/ExamRefDA100PowerBI/errata>. Если вы обнаружите в книге ошибку, не присутствующую в данном списке, просьба сообщить о ней на той же странице.

Для поддержки и дополнительной информации о книге вы можете перейти на страницу <http://www.MicrosoftPressStore.com/Support>.

Обратите внимание, что Microsoft не осуществляет поддержку программного и аппаратного обеспечения по приведенным выше ссылкам. Для получения помощи обращайтесь по адресу <http://support.microsoft.com>.

Оставайтесь на связи

Давайте продолжим общение! Наш Twitter: <http://twitter.com/MicrosoftPress>.

Глава 1

Подготовка данных

За последние пять лет Microsoft Power BI превратился из новичка в мире бизнес-аналитики в один из самых мощных аналитических инструментов для визуализации данных. Но, перед тем как начать анализировать информацию, мы должны подготовить ее, представить в виде модели и визуализировать. В данной главе мы будем говорить о первом шаге, а именно о подготовке данных, – пройдемся по всем навыкам, необходимым для загрузки данных в программу Power BI Desktop.

Начнем с действий, которые нужно выполнить для подключения к различным источникам данных. После этого подробно поговорим о техниках профилирования данных, которые помогают вам, как бы это сказать, почувствовать информацию. Далее мы перейдем к описанию процесса очистки и преобразования данных при помощи Power Query – во многих проектах, связанных с анализом информации, этот этап занимает непропорционально большое количество времени. В завершение главы посмотрим, как можно отлавливать и исправлять ошибки, возникшие при загрузке.

Навыки, описываемые в данной главе.

- 1.1. Получение данных из различных источников.
- 1.2. Профилирование данных.
- 1.3. Очистка, преобразование и загрузка данных.

Навык 1.1. Получение данных из различных источников

Вне зависимости от того, с каким источником данных вы имеете дело, прежде чем начать работать с информацией в Power BI, ее нужно получить. Power BI умеет подключаться к самым разнообразным *источникам данных* (data source), количество которых с каждым месяцем увеличивается. Кроме того, вы можете создавать собственные *коннекторы* (connector) к данным из Power BI, что фактически снимает любые ограничения на возможности получения информации извне.

В основе процесса получения данных лежит хорошее понимание бизнес-требований и особенностей вашего источника данных. К примеру, если вам необходимо работать с *данными почти в реальном времени* (near-real-time data), процедура получения информации будет значительно отличаться от ситуации, когда данные должны обновляться с некоторой периодичностью. Как вы совсем скоро узнаете, разные источники данных поддерживают разные режимы подключения.

Основные составляющие навыка:

- распознавание источника данных и подключение к нему;
- изменение настроек источника данных;
- выбор между использованием общего набора данных и созданием локального;
- выбор режима хранения;
- выбор подходящего типа запроса;
- поиск проблем с производительностью запросов;
- использование облачного хранилища Microsoft Dataverse;
- применение параметров;
- использование и создание файла PBIDS;
- использование и создание потока данных;
- подключение к набору данных с использованием конечных точек XMLA.

Распознавание источника данных и подключение к нему

В Power BI Desktop встроено более ста родных коннекторов к различным источникам данных, и команда разработчиков регулярно пополняет этот список. При этом самыми популярными источниками данных для Power BI остаются текстовые файлы, базы данных и веб-службы.

ДОПОЛНИТЕЛЬНАЯ ИНФОРМАЦИЯ. Источники данных в Power BI

Полный список источников данных, доступных в Power BI, можно увидеть по адресу <https://docs.microsoft.com/ru-ru/power-bi/power-bi-data-sources>.

Для выбора подходящего коннектора вы должны четко представлять, что из себя представляет источник. К примеру, вы не можете использовать коннектор Oracle для подключения к базе данных SQL Server, несмотря на то что оба коннектора предназначены для соединения с базами данных.

ПРИМЕЧАНИЕ. Сопроводительные файлы

В примерах из этой книги мы будем использовать сопроводительные файлы, данные в которых основаны на результатах деятельности вымышленной компании *Wide World Importers*. При этом мы будем исходить из предположения о том, что вы предварительно скопировали файлы в папку *C:\DA-100*.

Для проверки навыков, необходимых для подключения из Power BI к различным источникам, давайте начнем с открытия файла *WideWorldImporters.xlsx*, находящегося в списке сопроводительных файлов к книге.

1. На вкладке **Главная** (Home) в Power BI выберите пункт **Excel**.
2. В диалоговом окне **Открытие** (Open) найдите файл *WideWorldImporters.xlsx* и нажмите на кнопку **Открыть** (Open).
3. В открывшемся окне **Навигатор** (Navigator) отметьте все восемь флажков слева. После этого окно должно выглядеть так, как показано на рис. 1.1.

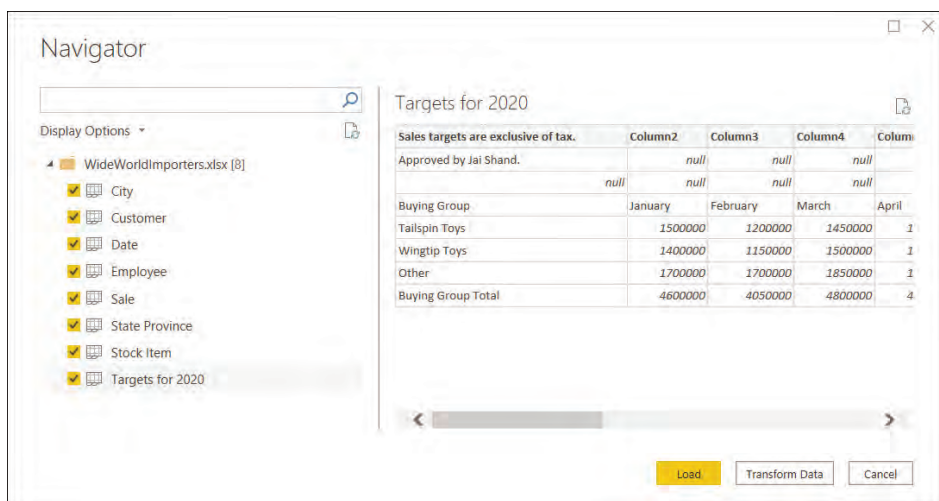


Рис. 1.1. Окно навигации подключения в Power BI

4. Нажмите на кнопку **Преобразовать данные** (Transform Data). В результате откроется окно **Редактор Power Query** (Power Query Editor), показанное на рис. 1.2.

Если в окне навигатора нажать на кнопку **Загрузить** (Load), окно редактора Power Query не появится, и выбранные вами листы Excel откроются как есть.

Обратите внимание, что в окне навигатора есть предварительный просмотр объектов, которые вы выбираете. Например, на рис. 1.1 мы видим содержимое листа *Targets for 2020*. Его структура предполагает выполнение

предварительной обработки перед загрузкой, поскольку в первых строках листа мы видим несколько странных значений.

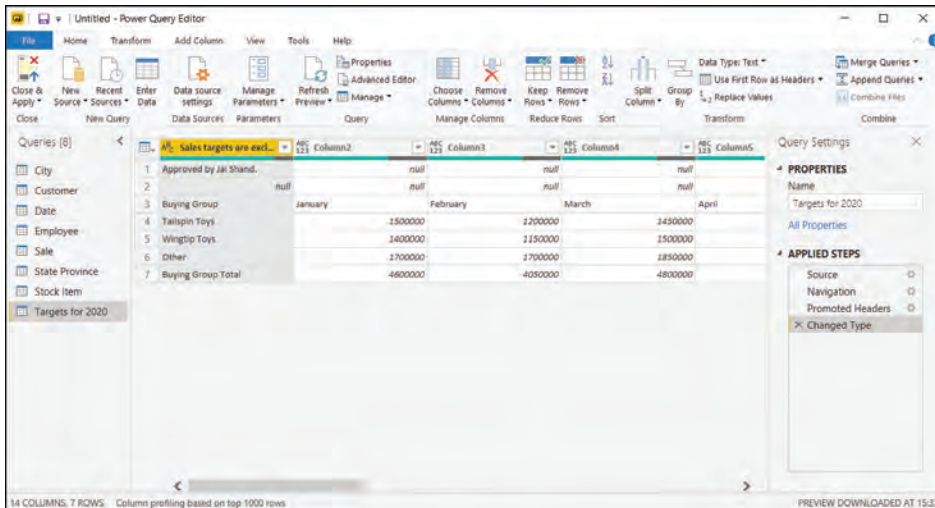


Рис. 1.2. Редактор Power Query

Окно **Навигатор** открывается не только при импорте данных из Excel. Такое же окно вы увидите, например, при подключении к сложным источникам – скажем, к базам данных.

ПРИМЕЧАНИЕ. Информация в предварительном просмотре

В целях ускорения работы с данными в Power Query предусмотрен механизм кеширования предварительного просмотра. Таким образом, если ваши данные часто меняются, вы можете видеть в редакторе Power Query не самую актуальную информацию. Для обновления предварительного просмотра на вкладке **Главная** (Home) нажмите на кнопку **Обновить предварительный просмотр** (Refresh Preview). Для актуализации данных во всех запросах раскройте эту выпадающую кнопку и выберите пункт **Обновить все** (Refresh All).

Преобразованием данных мы займемся позже в этой главе, а пока давайте попробуем подключиться к другому источнику, а именно к папке. Находясь в редакторе Power Query, выполните следующие действия.

1. На вкладке **Главная** (Home) нажмите на выпадающую кнопку **Создать источник** (New source) и выберите пункт **Дополнительно** (More).
2. В открывшемся окне **Получить данные** (Get data) выберите пункт **Папка** (Folder) и нажмите на кнопку **Подключить** (Connect).

3. Нажмите на кнопку **Обзор** (Browse), перейдите в папку *C:\DA-100\Targets* и дважды нажмите на кнопку **ОК**. Должен открыться список файлов, показанный на рис. 1.3.

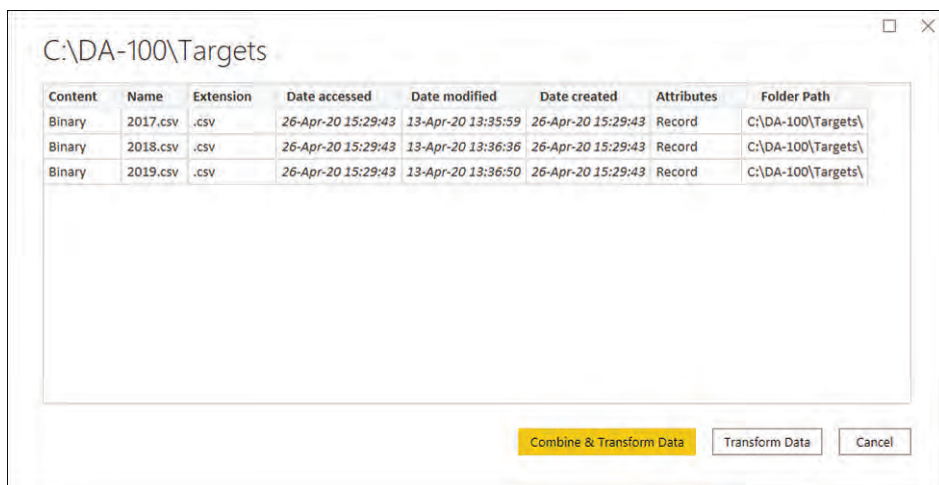


Рис. 1.3. Список файлов из папки *C:\DA-100\Targets*

4. Нажмите на кнопку **Объединить и преобразовать данные** (Combine & Transform Data).
5. В открывшемся окне **Объединить файлы** (Combine files) нажмите на кнопку **ОК**, не производя никаких изменений.

На данном этапе вы подключились к двум источникам одновременно – к файлу Excel и к папке, содержащей несколько файлов CSV.

И хотя мы не указывали тип файлов при подключении явно, Power Query автоматически определил его и применил подходящие преобразования. Помимо файлов Excel и CSV, Power BI умеет подключаться и ко многим другим типам файлов, включая JSON, XML, PDF и базы данных Access.

ВАЖНО. Единство форматов

Очень важно, чтобы форматы открываемых файлов в папке совпадали – в противном случае вы можете столкнуться с серьезными проблемами. Power Query применяет одинаковые преобразования ко всем файлам в папке, а решение о применяемых трансформациях принимается на основе файла, выбранного в поле **Пример** (Sample file) на этапе объединения файлов.

Редактор Power Query

Если вы в точности следовали всем приведенным выше инструкциям, ваше окно редактора Power Query в данный момент должно выглядеть так, как на рис. 1.4.

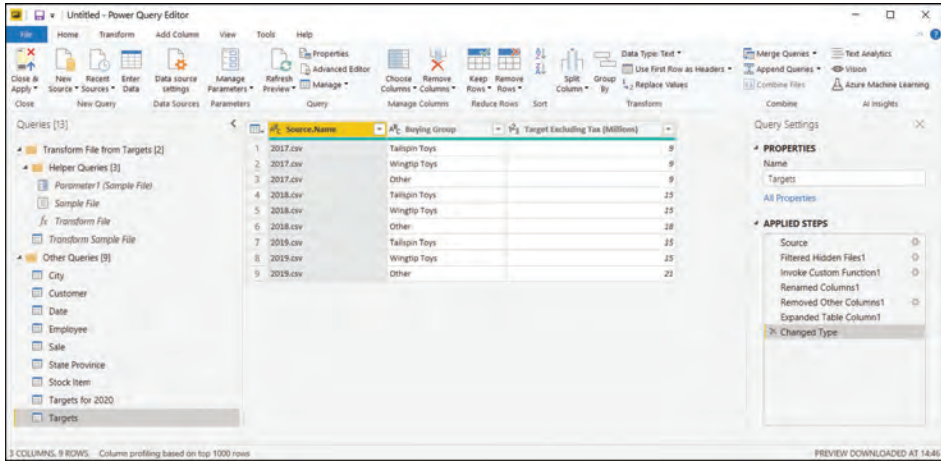


Рис. 1.4. Редактор Power Query после подключения к файлу Excel и папке

Как видите, по вашему требованию автоматически объединить все файлы из папки был создан запрос с именем *Targets* и несколько вспомогательных запросов, названия которых написаны курсивом, – это означает, что они не будут загружены. Варианты загрузки данных мы обсудим позже в этой главе, при этом мы будем использовать созданный здесь сценарий.

ПРИМЕЧАНИЕ. Сопроводительные файлы

Вы можете увидеть все произведенные до сих пор действия, открыв файл *1.1.1 Connect to data sources.pbix*, находящийся в папке Power BI files в архиве с сопроводительными файлами.

Зависимости запросов

Вы можете отследить все *зависимости* (query dependencies) созданных запросов, нажав на кнопку **Зависимости запроса** (Query dependencies) на вкладке **Просмотр** (View). В открывшемся окне будет отображена диаграмма с двумя источниками данных и всеми созданными запросами, как показано на рис. 1.5.



Рис. 1.5. Окно зависимостей запроса

Для просмотра зависимостей конкретного запроса выделите его, и Power BI подсветит все запросы, зависящие от него, а также все запросы и источники, от которых зависит выбранный запрос.

По умолчанию для отображения диаграммы зависимостей выбран макет с направлением сверху вниз, но вы можете изменить этот выбор, нажав на выпадающую кнопку **Структура** (Layout) в нижней правой части окна.

Изменение настроек источника данных

После подключения к источнику данных вам может понадобиться изменить некоторые его настройки. К примеру, если вы перенесете файл *WideWorldImporters.xlsx* в другую папку, вам придется обновить путь к нему в Power BI для продолжения работы с ним.

Изменить настройки источника данных можно, нажав на иконку с шестеренкой справа от шага **Источник** (Source) в разделе **Примененные шаги** (Applied steps) на панели **Параметры запроса** (Query Settings) в редакторе Power Query. В открывшемся окне вы можете отредактировать путь к файлу и его тип. Недостатком такого подхода является то, что вам необходимо будет производить изменения в каждом запросе, ссылающемся на файл, что при большом количестве запросов может быть довольно обременительно и рискованно с точки зрения возникновения ошибок.

Еще один способ внести изменения в источник данных – нажать на кнопку **Настройки источника данных** (Data source settings) на вкладке **Главная** (Home). Это приведет к открытию одноименного диалогового окна, показанного на рис. 1.6.

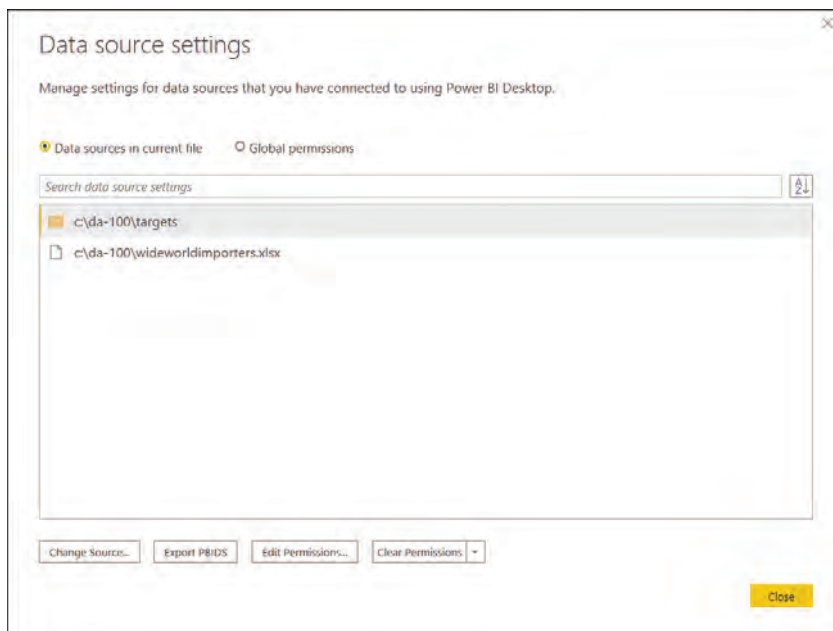


Рис. 1.6. Окно изменения настроек источника данных

В этом окне вы можете изменить настройки источника данных для нескольких запросов одновременно, нажав на кнопку **Изменить источник** (Change Source). Управлять *разрешениями* (permissions) для каждого источника данных можно, нажимая на кнопки **Править разрешения** (Edit Permissions) и **Очистить разрешения** (Clear Permissions). Разрешения включают в себя *учетные данные* (credentials), используемые для подключения к источнику, и настройку *уровня конфиденциальности* (privacy level). Об уровне конфиденциальности стоит помнить при объединении нескольких источников данных в одном запросе. Подробнее об этом мы поговорим далее в этой главе.

Выбор между использованием общего набора данных и созданием локального

До сих пор в этой главе мы создавали свои собственные наборы данных, также называемые *локальными* (local dataset). Но, если набор данных, подготовленный и опубликованный в службе Power BI ранее вами или кем-то другим, уже существует, он именуется *общим* (shared dataset), и вы можете подключиться к нему. Использование общих наборов данных дает следующие преимущества:

- данные в разных отчетах будут одинаковыми;
- во время подключения к общему набору данных не происходит ненужного копирования данных;
- вы можете создать копию существующего отчета и изменить ее, что занимает меньше времени, чем создавать отчет с нуля.

НА ПРАКТИКЕ. Использование общих наборов данных

Часто бывает, что сотрудникам разных отделов необходимо использовать одни и те же данные для просмотра при помощи разных элементов визуализации. В этом случае есть смысл создать один набор данных и разные отчеты, которые будут к нему подключаться.

Чтобы иметь возможность подключаться к опубликованному набору данных, вам необходимо обладать *разрешением на сборку* (Build permission) или быть участником рабочей области, в которой располагается набор данных. Подробнее о разрешениях мы будем говорить в главе 5.

К общему набору данных вы можете подключаться как из Power BI Desktop, так и из службы Power BI:

- в Power BI Desktop нажмите на вкладке **Главная** (Home) на кнопку **Наборы данных Power BI** (Power BI datasets);
- в службе Power BI, находясь в рабочей области, выберите пункт **Отчет** (Report) в меню **Создать** (New).

В том и другом случае вы увидите список общих наборов данных, к которым можете подключиться, как показано на рис. 1.7. Кроме того, в службе Power BI вы можете выбрать пункт **Сохранить копию** (Save a copy) напротив отчета в рабочей области для создания дубликата отчета без необходимости делать копию набора данных. Это похоже на обычное подключение к набору данных в Power BI Desktop, поскольку вы создадите отчет без соответствующей ему модели данных.

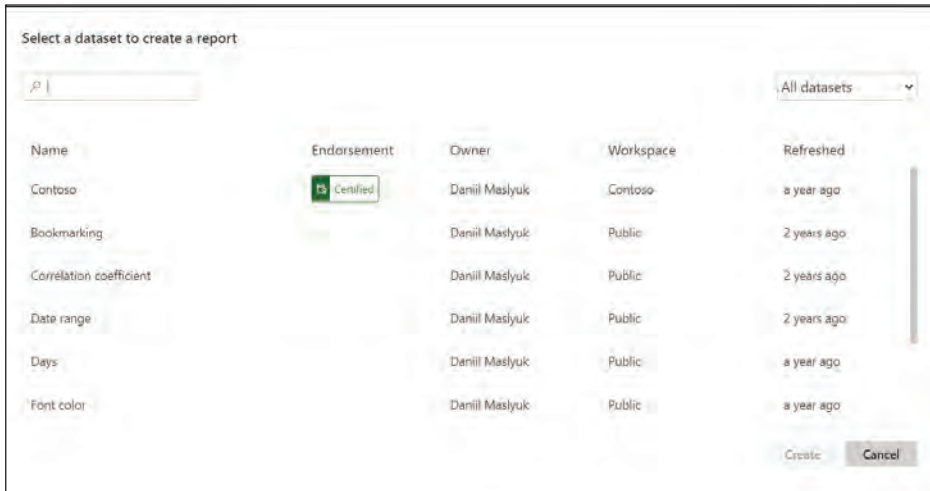


Рис. 1.7. Список доступных наборов данных

После выполнения подключения к общему набору данных в Power BI Desktop некоторые кнопки интерфейса могут стать недоступными по причине ограничений на режим подключения. К примеру, в этом случае вам будет недоступен редактор Power Query и вкладка **Данные** (Data). А в нижнем правом углу вы увидите название набора данных и рабочей области, к которой подключены, как показано на рис. 1.8.

Connected live to the Power BI dataset: Contoso in Contoso

Рис. 1.8. Power BI Desktop подключен к набору данных Power BI

Несмотря на то что кнопка преобразования данных не активна, вы можете раскрыть ее и выбрать пункт настройки источника данных для изменения набора данных, к которому вы подключены.

Обратите внимание, что вы по-прежнему можете создавать меры, и они будут сохраняться в вашем файле PBIX, но не в самом общем наборе данных. Это не позволит другим пользователям, подключенным к этому набору данных, видеть созданные вами меры. Такие меры называются *локальными* (local) или *мерами уровня отчета* (report-level measure). О создании мер мы подробно поговорим в главе 2.

Выбор режима хранения

Наиболее распространенным способом получения данных в Power BI является импорт в модель данных. Импортируя данные, вы фактически создаете их копию, которая остается статической вплоть до момента обновления. При этом в случае с данными из файлов и папок, к которым мы подключались ранее в этой главе, вам доступен только импорт в Power BI, тогда как при соединении с базами данных есть два возможных варианта.

Первый из них подразумевает импорт данных в Power BI и их кеширование. Этот способ предполагает максимальную гибкость в отношении моделирования данных, поскольку вам будут доступны все без исключения средства моделирования Power BI.

Второй подход, получивший название *DirectQuery*, связан с подключением к данным, хранящимся непосредственно в источнике. В этом случае данные не кешируются в Power BI. Вместо этого при каждом взаимодействии с элементом визуализации в Power BI создается и отправляется запрос в базу данных. Далеко не любые источники данных поддерживают режим хранения *DirectQuery*.

Особый случай режима хранения *DirectQuery* под названием *Live Connection* доступен для *Analysis Services* (как в *табличном* (Tabular), так и в *многомерном* (Multidimensional) режиме) и службы Power BI. Этот тип подключения гарантирует, что все вычисления будут выполняться в соответствующей модели данных.

Режим импорта данных

При *импорте данных* (import data) вы создаете их полную копию в Power BI. А поскольку в основе Power BI лежит движок столбчатых баз данных, работающий с памятью, такой способ подключения ведет к повышенному расходу оперативной памяти и дискового пространства, – ведь данные фактически хранятся в файлах. На этапе разработки проекта данные расходуют память и дисковое пространство на вашей рабочей машине, а после публикации отчета ресурсы будут расходоваться на сервере. Вывод простой – вам не удастся загрузить в Power BI данных больше, чем позволяет ваше оборудование. Это может стать проблемой при работе с действительно большими объемами данных.

У вас есть возможность выполнить преобразование данных перед их импортом в Power BI, и в этом случае вы будете ограничены только функционалом этого инструмента. Если загрузить ограниченный набор таблиц из базы данных и применить фильтр к некоторым таблицам, в Power BI будут импортированы только отфильтрованные данные.

После загрузки в Power BI данные сжимаются при помощи движка и хранятся в сжатом виде. Степень компрессии при этом зависит от множества факторов, включая типы данных столбцов, их значения и *кратность* (cardinality). В большинстве случаев итоговые данные после загрузки в Power BI будут занимать намного меньше места, чем в источнике.

Одним из преимуществ этого типа подключения является возможность использования всего функционала Power BI без ограничений, включая обработку данных в редакторе Power Query и применение функций на языке DAX при моделировании данных.

Кроме того, вы можете объединять в одной модели данные, импортированные из разных источников. Например, ничто не мешает вам собрать в одной таблице информацию из базы данных и файла Excel.

Еще одним преимуществом этого типа подключения можно назвать скорость вычислений. Поскольку Power BI хранит данные в памяти в сжатом виде, при доступе к данным практически не возникает задержки. Также стоит отметить, что движок Power BI очень хорошо оптимизирован для вычислений, что положительно сказывается на скорости расчетов.

Импортированные данные можно увидеть в Power BI Desktop на левой вкладке **Данные** (Data), а связи между загруженными таблицами – на вкладке **Модель** (Model). Кнопки **Отчет** (Report), **Данные** (Data) и **Модель** (Model), расположенные на левой панели, показаны на рис. 1.9.

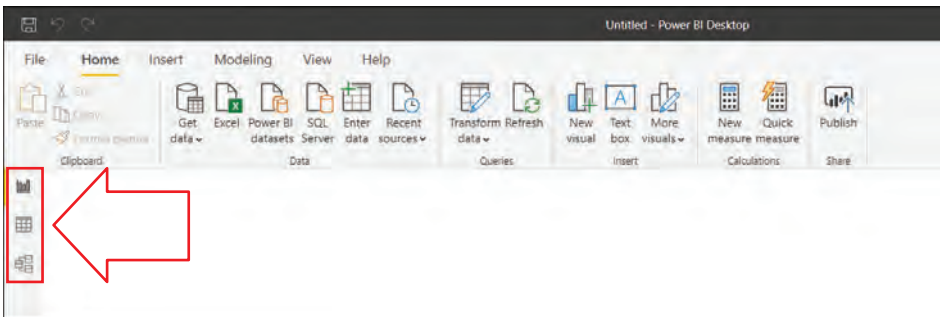


Рис. 1.9. Вид Power BI Desktop при импорте данных

Режим DirectQuery

При использовании режима подключения *DirectQuery* данные в Power BI не кешируются. Вместо этого все они остаются в источнике, за исключением метаданных, которые попадают в кеш Power BI. Эти метаданные включают в себя названия таблиц и столбцов, типы данных и связи между таблицами.

При работе с большинством источников данных, поддерживающих режим *DirectQuery*, во время подключения вы выбираете *сущности* (entity), такие как таблицы или представления. Каждая сущность в модели данных превращается в таблицу. Все происходит примерно так же, как при работе с окном навигатора во время импорта данных из рабочей книги Excel.

При использовании в модели данных только режима подключения *DirectQuery* размер файла Power BI будет несоизмеримо меньшим по сравнению со сценарием с импортом данных.

Главным преимуществом этого типа подключения к данным является то, что вы не ограничены ресурсами компьютера для разработки и сервера, на

котором будете публиковать отчеты. Все данные остаются в источнике, и вычисления производятся там же.

При выборе режима DirectQuery таблицы не будут видны на вкладке **Данные** (Data) в Power BI Desktop. Более того, если все таблицы в модели данных будут загружены в таком режиме, кнопка **Данные** (Data) на левой панели будет просто отсутствовать за ненадобностью, при этом вы сможете продолжать пользоваться вкладкой **Модель** (Model). Фрагмент интерфейса Power BI Desktop при работе с режимом DirectQuery показан на рис. 1.10.

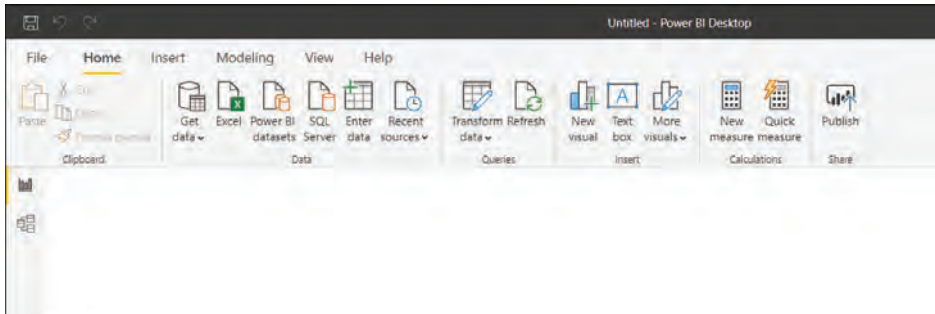


Рис. 1.10. Интерфейс Power BI Desktop при выборе режима хранения DirectQuery

Режим Live Connection

Для наборов данных службы Power BI и моделей данных Analysis Services доступен особый режим подключения DirectQuery, называемый *Live Connection*. Он отличается от обычного DirectQuery тем, что:

- вы не можете применять операции преобразования к данным;
- вы не можете устанавливать физические связи;
- процесс моделирования данных ограничен созданием мер.

Вы можете рассмотреть режим Live Connection вместо импорта по причине расширенных возможностей моделирования данных и улучшенной безопасности в источнике данных. В отличие от DirectQuery при работе с некоторыми базами данных, режим Live Connection подразумевает строгое фиксирование имен пользователей, просматривающих отчет. Это позволяет настроить систему безопасности динамически. Кроме того, SQL Server Analysis Services может быть настроен на любую частоту обновления данных, тогда как в службе Power BI обновления регулируются расписанием, ограниченным восемью сеансами в день без покупки дорогостоящей версии Power BI Premium.

Составные модели

Составная модель (composite model) представляет собой модель данных, сочетающую в себе импортированные данные и информацию, доступ к которой осуществляется в режиме DirectQuery, либо данные, полученные в режиме DirectQuery из разных источников. Например, вы можете брать

актуальную информацию о продажах непосредственно из базы данных посредством DirectQuery и импортировать данные о планах продаж из Excel. Скомбинировать оба источника данных в одной модели можно, создав составную модель.

ВАЖНО. Потенциальные проблемы с безопасностью в составных моделях

Создание и использование составных моделей данных может быть сопряжено с серьезными рисками в отношении безопасности. К примеру, данные из файла Excel могут быть посланы в запросе в базу данных, и администратор баз данных получит к ним доступ.

В составной модели данных свойство *режим хранения* (storage mode) для каждой таблицы определяет тип подключения к ней. Посмотреть значение этого свойства можно во всплывающей подсказке при наведении мыши на таблицу в панели **Поля** (Fields) на вкладке **Отчет** (Report) или **Данные** (Data). Также вы можете посмотреть и изменить это свойство на вкладке **Модель** (Model) в секции **Дополнительно** (Advanced) на панели **Свойства** (Properties) для выбранной таблицы. Здесь режим хранения может быть установлен в одно из трех значений:

- Импорт (Import);
- DirectQuery;
- Двойной (Dual).

Двойной режим (Dual) предполагает, что таблица может как кешироваться, так и извлекаться в режиме DirectQuery, когда это необходимо, – в зависимости от режимов хранения других таблиц, участвующих в запросе. Этот режим может оказаться полезным, когда у вас есть таблица, связанная одновременно с импортированными таблицами и таблицами с режимом подключения DirectQuery. Рассмотрим для примера модель данных, показанную в табл. 1.1.

Таблица 1.1. Пример модели данных

Имя таблицы	Источник данных	Режим хранения
Sales	База данных	DirectQuery
Date	База данных	Двойной
Targets	Файл Excel	Импорт

В представленной модели данных таблица *Date* одновременно связана и с таблицей *Sales*, и с *Targets*. Когда вы запрашиваете данные в запросе из таблиц *Date* и *Sales*, они извлекаются напрямую из базы данных в режиме DirectQuery, а когда обращаетесь к таблицам *Date* и *Targets*, никакие запро-

сы в базу данных не посылаются, что позволяет повысить производительность отчетов.

ВАЖНО. Изменение режима хранения

При смене режима хранения с DirectQuery или Двойного на Импорт важно помнить, что обратного пути нет. Если вам необходимо установить для таблицы режим Двойной, вы должны сначала создать таблицу в режиме DirectQuery.

Выбор подходящего типа запроса

Для достижения максимальной эффективности при работе с данными их необходимо импортировать. Но есть сценарии, в которых режим хранения DirectQuery будет даже более предпочтительным по сравнению с импортом. Вот некоторые из них:

- если размер модели данных превышает объем памяти для ее хранения, режим DirectQuery может стать хорошим вариантом. При этом не стоит забывать, что производительность отчетов в этом случае будет зависеть от ресурсов сервера, на котором расположен источник данных;
- если оперативные данные часто подвергаются изменениям, а в отчетах должна всегда отражаться актуальная информация, режим DirectQuery будет незаменим. Опять же, при условии, что источник сможет физически возвращать рассчитанные данные с относительно небольшими задержками. В противном случае нет смысла пытаться постоянно обращаться к актуальным данным;
- если политика компании предполагает хранение данных в конкретном источнике, без режима DirectQuery будет не обойтись.

Последствия использования режима DirectQuery

Использование режима хранения DirectQuery может приводить к описанным ниже последствиям в отношении функциональности.

Непостоянство производительности

Применяя режим хранения DirectQuery, вы обрекаете систему отчетности на зависимость от ресурсов аппаратного обеспечения, на котором располагается источник данных. Если сервер способен возвращать информацию менее чем за пять секунд, решение использовать такой метод подключения можно считать оправданным, хотя пользователи, привыкшие к скорости работы родного движка Power BI, даже в этом случае могут быть недовольны. Если же отчеты формируются дольше, систему может быть

просто проблематично использовать. Также стоит учитывать, сможет ли источник данных справиться с дополнительной нагрузкой, связанной с частыми запросами. При использовании режима DirectQuery любой визуальный элемент, к которому обращается пользователь, может посылать запросы в источник, причем происходит это одновременно для всех пользователей, работающих с системой.

Не все типы запросов допустимы

В режиме хранения DirectQuery можно использовать не все типы запросов к источнику. Когда пользователь взаимодействует с визуальным элементом в отчете, использующем режим DirectQuery, все запросы на извлечение данных объединяются и посылаются на сервер. В связи с этим не допускается применять нативные для источника запросы, включающие, например, *общие табличные выражения* (common table expressions) и *храняемые процедуры* (stored procedures).

Ограниченная функциональность в отношении преобразования данных

По сравнению с импортом данных, в свободе выбора преобразований исходной информации в случае использования режима DirectQuery вы будете ограничены из соображений производительности. Преобразования данных выполняются при каждом обращении к элементу визуализации, а не один раз в момент обновления данных, как в случае с импортом. И в вашем арсенале будут только те преобразования, которые могут быть эффективно переведены на язык источника. Если вы попытаетесь использовать преобразование, не поддерживаемое источником, вы получите ошибку, показанную на рис. 1.11, с советом либо отказаться от этой операции, либо импортировать данные.

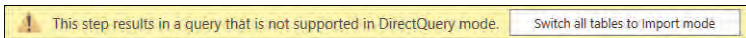


Рис. 1.11. Ошибка об отсутствии поддержки операции в режиме DirectQuery

Ограничения в моделировании данных

В области моделирования данных при использовании режима хранения DirectQuery вы также будете испытывать некоторые ограничения. К моделированию данных относится создание мер, вычисляемых столбцов, иерархий и связей, переименование и скрытие колонок, форматирование мер и столбцов, а также определение метода сортировки и типа агрегации для столбцов. Применение режима DirectQuery может быть связано со следующими ограничениями.

1. В режиме DirectQuery не создаются встроенные таблицы дат для столбцов, хранящих временные характеристики, как в случае с импортом. Эти таблицы используются в вычислениях, связанных с логикой операций со временем, и, если в источнике есть календарь, он может быть применен для этих целей.

2. При работе с вычисляемыми столбцами ограничения режима DirectQuery могут проявляться двояко. Во-первых, в них может использоваться только текущая строка таблицы или связанная строка, извлеченная по связи «один ко многим», что исключает возможность применения агрегационных функций. Во-вторых, вычисляемые столбцы в этом случае могут использовать лишь ограниченный набор функций, возвращающих скалярные значения, а именно только те из них, которые могут быть легко переведены на родной для источника язык. К примеру, вы можете создать столбец *Month Name* в таблице *Sales* с использованием функции *RELATED*, но при этом не сможете подсчитать количество строк в таблице *Sales* для каждой строки в таблице *Date* в вычисляемом столбце, поскольку это потребовало бы использования агрегационной функции вроде *COUNTROWS*. Обычно технология автодополнения функций от Microsoft (*IntelliSense*) дает на выбор только поддерживаемые функции.
3. Вычисляемые таблицы не поддерживаются в режиме DirectQuery.
4. Также в режиме DirectQuery нельзя использовать функции иерархий типа родитель–потомок вроде *PATH*. Если вам нужно создать иерархию по сотрудникам или план счетов, рассмотрите вариант размещения ее в источнике данных.
5. Недопустимо в режиме DirectQuery строить *кластеры* (cluster), опирающиеся на функции языка DAX.

В табл. 1.2 мы свели воедино все схожие и различающиеся характеристики трех режимов подключения к данным при работе в Power BI.

Таблица 1.2. Режимы подключения к данным

	Импорт	DirectQuery	Live Connection
Максимальный размер модели данных	Зависит от лицензии: Power BI Pro: 1 Гб в расчете на один набор данных; Power BI Premium: зависит от ресурсов	Ограничен аппаратным обеспечением сервера, на котором хранится источник данных	Служба Power BI: такие же ограничения, как в режиме импорта. Другие источники: ограничен аппаратным обеспечением сервера, на котором хранится источник данных
Количество источников данных	Без ограничений	Без ограничений	Только один
Обновление данных	Зависит от лицензии: Power BI Pro: до 8 раз в день с 30-мин интервалами; Power BI Premium: неограниченно	В отчетах отображается актуальная информация из источника данных	В отчетах отображается актуальная информация из источника данных

	Импорт	DirectQuery	Live Connection
Производительность	Лучший вариант	Зависит от источника данных. Будет почти всегда медленнее по сравнению с импортом и Live Connection	Лучший вариант
Преобразование данных	Без ограничений	Ограничено инструкциями, которые могут быть переведены на язык источника	Нет
Моделирование данных	Без ограничений	Сильно ограничено	Analysis Services и служба Power BI: меры могут создаваться без ограничений
Безопасность	Безопасность на уровне строк может быть применена на основе информации о пользователе	Безопасность на уровне строк, определенная в источнике данных, доступна не для всех источников. В Power BI Desktop безопасность на уровне строк реализуется	Могут быть применены правила безопасности источника данных на основе информации о пользователе



СОВЕТ ПЕРЕД ЭКЗАМЕНОМ

Вы должны уметь определять ситуации, в которых уместно применить режим DirectQuery или Live Connection на основании бизнес-требований клиента.

Поиск проблем с производительностью запросов

При подключении к данным в Power BI проблемы с производительностью могут возникать по разным причинам. В Power BI Desktop присутствуют специальные инструменты для обнаружения этих проблем.

Просмотр машинных запросов

При получении информации в Power BI из определенных источников, таких как базы данных, инструмент Power Query будет стараться наиболее эффективно переводить операции преобразования данных на родной язык для источника – например, на SQL. Такие действия, выполняемые Power Query, получили название *свертывание запросов* (query folding). В большинстве случаев это приводит к повышению эффективности получения данных из внешних источников. Допустим, если вы подключаетесь к базе данных и

извлекаете набор столбцов из таблицы, Power Query может отфильтровать нужные колонки еще в источнике данных, чтобы не переносить их все и не выбирать нужные на месте.

В некоторых случаях мы можем посмотреть на сам запрос, который Power Query отправляет в источник для извлечения данных. Для этого нужно щелкнуть правой кнопкой мыши по запросу в редакторе Power Query и выбрать пункт **Просмотреть машинный запрос** (View Native Query). Окно, которое появится на экране, показано на рис. 1.12.

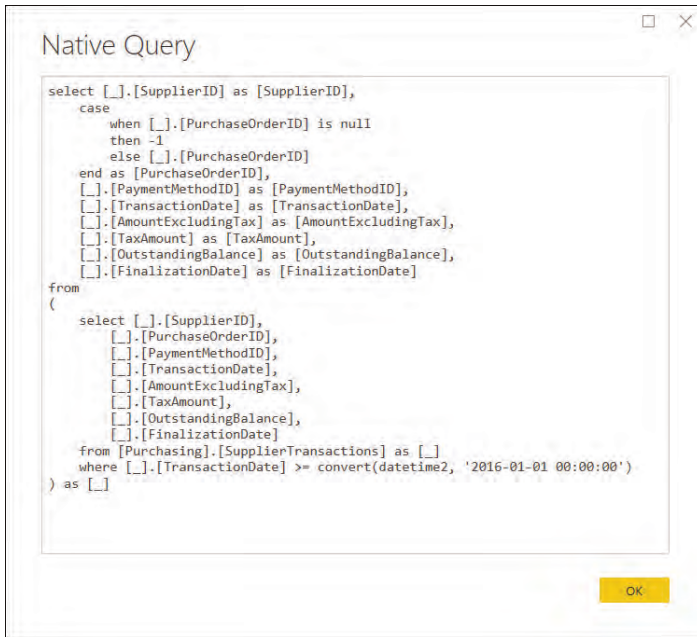


Рис. 1.12. Окно просмотра машинного запроса

В запросе, показанном на рис. 1.12, мы подключаемся к базе данных SQL Server, применяем фильтр, выбираем нужные столбцы и заменяем значения. Поскольку эти операции можно легко перевести на язык запросов SQL, Power Query решил произвести эти преобразования на стороне источника данных вместо того, чтобы забирать все данные и обрабатывать их локально. Это бы негативно сказалось на производительности.

Стоит отметить, что вы не можете редактировать машинный запрос, он открывается только на просмотр. Если хотите, чтобы Power BI отправлял определенный запрос на сервер, вы должны представить выражение на языке запросов SQL при подключении к источнику.

Если пункт **Просмотреть машинный запрос** не активен, это означает, что этот источник данных не поддерживает операцию свертывания запросов или конкретный шаг не может быть эффективно переведен на машинный язык. Например, если мы запустим операцию очистки по текстовому полю, свертывания запроса не произойдет ввиду отсутствия аналогичной операции в языке запросов SQL.