

УДК 004.424
ББК 32.372
Б30

Бахши С.

Б30 Power BI: моделирование на экспертном уровне / пер. с англ. А. Ю. Гинько. – М.: ДМК Пресс, 2022. – 490 с.: ил.

ISBN 978-5-97060-906-4

В этой книге описываются техники моделирования данных с помощью Power BI. Показано, как подключаться к данным в различных источниках, преобразовывать их с помощью Power Query и DAX, объединять посредством связей и строить модели данных с учетом разнообразных, даже самых изысканных, бизнес-требований. На примерах рассмотрена оптимизация сложных моделей данных без потери функционала. К концу книги читатель будет обладать всеми необходимыми знаниями для структурирования и обработки данных, поступающих из разных источников, и создания на их основании полноценных моделей, пригодных для построения отчетов.

Книга предназначена для пользователей систем бизнес-аналитики, а также специалистов и разработчиков в области анализа данных, желающих повысить квалификацию и расширить навыки использования Power BI.

УДК 004.424
ББК 32.372

Copyright ©Packt Publishing 2021. First published in the English language under the title 'Expert Data Modeling with Power BI - (9781800205697)

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.

ISBN 978-1-80020-569-7 (англ.)
ISBN 978-5-97060-906-4 (рус.)

© Packt Publishing, 2021
© Перевод, оформление, издание,
ДМК Пресс, 2022

Дизайн обложки разработан с использованием ресурса freepik.com

Содержание

От издательства	13
Предисловие	14
Об авторе	15
О технических редакторах	16
Введение	17
Глава 1. Введение в моделирование данных в Power BI	23
Понятие слоев в Power BI Desktop	24
Слой подготовки данных (Power Query)	25
Слой модели данных	25
Вкладка Данные	26
Вкладка Модель данных	27
Слой визуализации данных	28
Вкладка Отчет	28
Поток данных в Power BI	29
Что означает моделирование данных в Power BI	30
Семантическая модель	31
Построение эффективной модели данных в Power BI	32
Схемы «звезда» (многомерное моделирование) и «снежинка»	34
Транзакционные модели против схемы «звезда»	34
Схема «снежинка»	36
Понятие денормализации	36
Варианты лицензирования в Power BI	42
Максимальный размер набора данных	43
Добавочная загрузка данных	43
Группы вычислений	44
Общие наборы данных	45
Потоки данных Power BI	45
Итеративный подход к моделированию данных	45
Сбор информации от руководства	46
Подготовка данных на основе бизнес-логики	46
Моделирование данных	47
Проверка логики	47
Демонстрация бизнес-логики в базовой визуализации	47
Думай как профессиональный разработчик моделей данных	48
Заключение	48
Глава 2. DAX и моделирование данных	50
Понимание виртуальных таблиц	50

Создание вычисляемой таблицы.....	51
Использование виртуальных таблиц в мерах, часть 1	53
Использование виртуальных таблиц в мерах, часть 2	55
Визуальное представление виртуальных таблиц	56
Создание вычисляемых таблиц в Power BI Desktop	56
Использование DAX Studio.....	57
Связи в виртуальных таблицах.....	58
Логика операций со временем и моделирование данных	68
Определение валидности дат в измерении	68
Вычисления на основе сравнения периодов.....	76
Создание измерения дат при помощи DAX.....	84
Пометка календаря как таблицы дат.....	86
Создание измерения времени при помощи DAX	90
Заключение	92
Глава 3. Подготовка данных с помощью Power Query.....	95
Введение в язык формул M, используемый в Power Query	95
Power Query – регистрозависимый инструмент	96
Запросы	97
Выражения.....	97
Значения	97
Примитивные значения	97
Структурированные значения.....	98
Типы.....	102
Примитивные типы	102
Пользовательские типы.....	103
Введение в редактор Power Query.....	103
Панель Запросы.....	105
Таблицы	105
Настраиваемые функции	105
Параметры запросов.....	105
Константы.....	105
Группы.....	105
Панель Параметры запроса.....	106
Свойства.....	108
Область данных.....	109
Строка состояния	112
Расширенный редактор	113
Возможности Power Query в области моделирования данных	114
Качество столбца	115
Распределение столбцов.....	118
Профиль столбца.....	121
Параметры запросов	122
Настраиваемые функции	128
Рекурсивные функции	133
Заключение	135

Глава 4. Получение данных из различных источников	136
Получение данных из распространенных источников данных.....	136
Папка	137
CSV/Текст/TSV	142
Excel	148
Наборы данных Power BI.....	155
Потоки данных Power BI	159
SQL Server.....	160
SQL Server Analysis Services и Azure Analysis Services.....	162
SSAS многомерная/табличная	163
AAS.....	165
Канал OData	166
Сертификаты источников данных	169
Bronze	169
Silver.....	169
Gold/Platinum.....	170
Режимы подключения к данным.....	170
Импорт	171
Применение	171
Ограничения	171
DirectQuery.....	171
Применение	172
Ограничения	172
Подключение в режиме реального времени	172
Применение	173
Ограничения	173
Режимы хранения данных.....	173
Режимы хранения наборов данных	175
Заключение	177
 Глава 5. Общие шаги по подготовке данных	 178
Изменение типов данных.....	179
Разделение столбцов по разделителю	186
Объединение столбцов.....	189
Создание настраиваемого столбца.....	190
Создание столбца из примеров	193
Создание дубликата столбца	195
Фильтрация строк.....	197
Группирование данных.....	201
Добавление запросов.....	203
Объединение запросов.....	206
Создание дубликата запроса и ссылки на запрос	208
Замена значений.....	210
Извлечение чисел из текста.....	212
Работа с датой, временем и часовыми поясами	215
Заклучение	218

Глава 6. Подготовка данных в Power Query для схемы

«звезда»	219
Выявление измерений и фактов.....	219
Количество таблиц в источнике данных	220
Связи между существующими таблицами.....	221
Наименьшая требуемая гранулярность полей с датой и временем	222
Определение измерений и фактов	223
Выявление возможных измерений	224
Выявление возможных фактов	225
Создание таблиц измерений.....	227
Geography	228
Sales Order	230
Product	233
Currency	236
Customer	237
Sales Demographic	238
Date	241
Time.....	245
Создание измерений Date и Time – Power Query против DAX.....	246
Создание таблиц фактов	247
Заключение	254

Глава 7. Эффективные методики подготовки данных

Общие рекомендации по подготовке данных.....	256
При работе с источником OData используйте частичную загрузку данных.....	256
Не забывайте о регистрозависимости Power Query	259
Помните о свертывании запросов и его влиянии на обновление данных.....	260
Понятие свертывания запросов	260
Свертывание запросов и режимы хранения DirectQuery и Dual.....	261
Свертывание запросов и источники данных	261
Индикация свертывания запросов	261
Рекомендации по выполнению свертывания запросов	263
Организуйте запросы в редакторе Power Query.....	267
Преобразование типов	268
Преобразование типов и влияние на моделирование данных.....	269
Включение преобразования типов в шаги	275
Изменение типов данных за один шаг.....	276
Оптимизация размера запросов.....	277
Избавьтесь от лишних строк и столбцов	277
Выполните агрегирование (группировку)	278
Отмените загрузку запросов.....	279
Соглашение о наименованиях	279
Заключение	280

Глава 8. Элементы моделирования данных	282
Моделирование данных в Power BI Desktop.....	282
Введение в таблицы	283
Свойства таблицы	283
Рекомендуемые таблицы.....	286
Вычисляемые таблицы.....	287
Введение в поля.....	292
Типы данных	292
Пользовательское форматирование	294
Столбцы.....	295
Вычисляемые столбцы.....	295
Группирование данных в столбцах и разделение их на ячейки	296
Свойства столбцов.....	300
Иерархии.....	304
Меры	305
Неявные меры.....	305
Явные меры.....	308
Текстовые меры	308
Использование связей.....	310
Первичные и внешние ключи.....	311
Управление составными ключами.....	311
Связь «один к одному».....	316
Связь «один ко многим»	316
Связь «многие ко многим».....	316
Распространение фильтров.....	318
Двунаправленные связи	320
Заключение	323
 Глава 9. Схема «звезда» и распространенные техники при моделировании данных	324
Работа со связями типа «многие ко многим»	324
Связи «многие ко многим» с использованием таблицы-моста.....	327
Скрытие таблицы-моста.....	333
Повышенная бдительность при использовании двунаправленных связей.....	334
Работа с неактивными связями.....	337
Доступность таблицы по нескольким путям фильтра	337
Несколько прямых связей между двумя таблицами	339
Использование конфигурационных таблиц	341
Сегментирование.....	341
Динамическое условное форматирование с участием мер	342
Минусы создания вычисляемых столбцов.....	348
Организация модели данных.....	351
Скрытие второстепенных объектов.....	351
Скрытие неиспользуемых полей и таблиц	351
Скрытие ключевых полей	353
Скрытие неявных мер.....	354

Скрытие столбцов, использующихся в иерархиях, там, где это возможно	354
Создание таблиц мер.....	354
Рассуждения	356
Использование папок.....	357
Создание папки в нескольких таблицах в одно действие	357
Помещение меры в разные папки	359
Создание подпапок	359
Уменьшение размера модели путем отказа от автоматических таблиц с датами и временем	360
Заключение	362
Глава 10. Продвинутое моделирование данных	364
Использование агрегаций	364
Реализация агрегирования для источников, не поддерживающих DirectQuery.....	365
Реализация агрегации на уровне Date	366
Использование инструмента управления агрегированием.....	376
Управление агрегированием в Power BI Desktop для источников, поддерживающих DirectQuery, и больших данных	378
Проверка агрегирования	382
Добавочное обновление.....	387
Настройка добавочного обновления в Power BI Desktop	389
Проверка добавочного обновления	394
Иерархии типа родитель–потомок	396
Определение глубины иерархии	398
Создание уровней иерархии	400
Релевые измерения	403
Использование групп вычислений.....	406
Требования	407
Терминология.....	407
Группы вычислений и логика операций со временем.....	408
Тестирование групп вычислений.....	414
Проблема с форматированием строк.....	415
Функции DAX для групп вычислений.....	417
Заключение	417
Глава 11. Безопасность на уровне строк	418
Безопасность на уровне строк при моделировании данных.....	419
Чем безопасность на уровне строк не является	419
Терминология безопасности на уровне строк.....	419
Роли	420
Правила	420
Проверка ролей.....	421
Назначение участникам ролей в службе Power BI	423
Назначение участникам ролей в Power BI Report Server.....	423

Реализация безопасности на уровне строк.....	425
Распространенные подходы в организации безопасности на уровне строк.....	426
Статическая безопасность на уровне строк.....	426
Создание ролей и определение правил	427
Проверка ролей.....	428
Публикация отчета в службе Power BI	429
Назначение участникам ролей.....	430
Проверка ролей в службе Power BI.....	432
Динамическая безопасность на уровне строк.....	432
Каждый пользователь имеет доступ только к своим данным	433
Менеджер может видеть данные подчиненных	436
Получение учетных данных пользователей из стороннего источника	442
Заключение	448

Глава 12. Дополнительные опции и возможности

моделирования данных	449
Медленно меняющиеся измерения	449
Медленно меняющиеся измерения типа 0 (SCD 0)	451
Медленно меняющиеся измерения типа 1 (SCD 1)	451
Медленно меняющиеся измерения типа 2 (SCD 2)	451
Безопасность на уровне объектов	455
Реализация безопасности на уровне объектов	455
Проверка ролей	458
Назначение участникам ролей в службе Power BI	460
Введение в потоки данных	462
Сценарии для использования потоков данных	462
Терминология потоков данных	463
Создание потока данных.....	464
Создание сущностей	467
Создание связанных сущностей из других потоков данных	471
Создание вычисляемых сущностей	474
Импорт и экспорт потоков данных.....	476
Составные модели.....	478
Новая терминология	479
Построение цепочек	479
Длина цепочки	479
Заключение	485
Предметный указатель.....	486