

Как и многие другие авторы, я посвящаю эту книгу первым делом моей жене и пятилетнему сыну. Особенно сыну, который оказался большим молодцом, позволив мне сосредоточиться на сверхурочной работе по выходным вместо совместного отдыха и игр. До заключительных глав книги я не осознавал в полной мере, насколько важной будет для меня поддержка семьи, а долгие месяцы с COVID заставили нас преодолевать новые личные и профессиональные преграды. Несмотря на изоляцию и привыкание к новой стране, они продолжали поддерживать меня и отмечать окончание каждой главы как маленький праздник. И за это я выражаю им свою самую глубокую благодарность.

Также я хотел бы сказать спасибо всем коллегам во времена работы в Microsoft в отделе Power BI. Я многому научился у парней из команды Power BI CAT, архитектурных проектировщиков, менеджеров по разработке ПО и экспертов в области отчетности. Список тех, кому я благодарен, слишком велик, чтобы его здесь приводить. К тому же есть риск кого-то забыть, чего мне не хотелось бы. Надеюсь, их глубокие знания в совокупности с моим богатым опытом помогут вам вывести свои решения Power BI на новый уровень.

Содержание

От издательства	12
Предисловие	13
Об авторе	14
О редакторах	15
Введение	16
Часть I. АРХИТЕКТУРА, УЗКИЕ МЕСТА И ЦЕЛЕВЫЕ ПОКАЗАТЕЛИ ПРОИЗВОДИТЕЛЬНОСТИ	21
Глава 1. Постановка целей и определение проблемных областей	22
Определение уровня производительности.....	23
Показатели производительности отчетов.....	23
Установка реалистичных целевых показателей производительности.....	24
Области с возможными замедлениями.....	25
Подключение к источникам данных.....	26
Режим Import.....	26
Режим DirectQuery.....	27
Режим Live connection.....	27
Шлюз Power BI.....	27
Сетевая задержка.....	28
Служба Power BI.....	29
Решения, влияющие на производительность.....	30
Заключение.....	30
Глава 2. Обзор архитектуры и конфигурации Power BI	32
Средства подключения к источникам и режимы хранения данных.....	32
Выбор между режимами Import и DirectQuery.....	33

Когда лучше подойдет режим DirectQuery?.....	36
Составные модели.....	37
Режим LiveConnect.....	38
Извлечение локальных данных с помощью шлюза.....	39
Как работает шлюз.....	40
Предпосылки для оптимальной работы шлюза.....	40
Технические характеристики шлюза.....	42
Настройка ведения логов в шлюзе.....	43
Анализ и моделирование логов шлюза.....	45
Анализ логов шлюза.....	47
Масштабирование шлюза.....	48
Горизонтальное масштабирование с увеличением количества шлюзов.....	48
Общая инструкция по архитектуре.....	50
Планирование расписания обновлений.....	50
Снижение сетевой задержки.....	50
Заключение.....	51
Глава 3. Оптимизация DirectQuery.....	53
Моделирование данных для режима DirectQuery.....	54
Оптимизация связей для DirectQuery.....	57
Настройки быстрого действия режима DirectQuery.....	60
Настройки Power BI Desktop.....	60
Оптимизация внешних источников данных.....	62
Заключение.....	64

Часть II. АНАЛИЗ, УЛУЧШЕНИЕ И УПРАВЛЕНИЕ ПРОИЗВОДИТЕЛЬНОСТЬЮ.....65

Глава 4. Анализ логов и метрик.....	66
Метрики использования в Power BI.....	66
Доработка отчета о метриках использования.....	69
Фильтрация метрик использования.....	69
Доступ к сырым данным посредством создания редактируемой копии метрик использования.....	70
Доступ к сырым данным посредством создания собственного отчета о метриках использования.....	73
Доступ к сырым данным с помощью анализа метрик использования в Excel.....	74
Анализ детализированной информации о производительности.....	74
Анализ метрик отчета о производительности.....	76
Получение показателей производительности из нескольких рабочих областей.....	79
Логи Power BI и трассировка.....	80
Журнал действий и единый журнал аудита.....	80
Трассировка Analysis Services с помощью конечных точек XMLA.....	81

Интеграция с Azure Log Analytics	81
Отслеживание показателей в Azure Analysis Services и Power BI Embedded.....	82
Метрики Azure для AAS	82
Диагностика в Azure для Analysis Services	83
Метрики Azure и диагностика для PBIE	84
Заключение	84
Материалы к прочтению.....	85
Глава 5. Анализатор производительности	86
Технические требования.....	86
Обзор Анализатора производительности	87
Действия и метрики в Анализаторе производительности.....	88
Определение действий пользователя.....	89
Определение и устранение проблем с производительностью	92
Единообразие тестов.....	93
Возможности и ограничения Анализатора производительности.....	97
Интерпретация и выводы о данных от Анализатора производительности	98
Медленные запросы.....	98
Медленные визуальные элементы	100
Эффект от добавления новых визуальных элементов.....	102
Экспорт и анализ данных о производительности	103
Заключение	107
Глава 6. Внешние инструменты.....	109
Технические требования.....	110
Power BI Helper.....	110
Поиск столбцов, занимающих много места	110
Поиск неиспользуемых столбцов	111
Поиск двунаправленных и неактивных связей	112
Поиск зависимостей в мерах	112
Tabular Editor	113
Использование утилиты Best Practice Analyzer	113
DAX Studio и VertiPaq Analyzer	118
Анализ размера модели данных при помощи VertiPaq Analyzer	118
Настройка производительности модели данных и запросов DAX.....	120
Перехват и повторный запуск запросов	120
Получение информации о времени выполнения запросов.....	122
Изменение и настройка запросов.....	123
Заключение	126
Глава 7. Общие принципы управления производительностью	128
Налаживание воспроизводимого и упреждающего процесса повышения производительности.....	129
Цикл управления производительностью.....	130
Установка/обновление контрольных целевых показателей.....	130

Мониторинг и хранение истории	132
Обнаружение проблем и расстановка приоритетов	132
Диагностирование и исправление	132
Принятие превентивных мер	132
Обмен опытом и знаниями	133
Помощь конечным пользователям	133
Инструкция для разработчиков	134
Совместный подход к повышению производительности	134
Применение цикла управления производительностью в разных сценариях.....	135
ВИ-системы самообслуживания	135
ВИ-системы на основе отдела или команды	136
Корпоративные или управляемые ИТ-отделами ВИ-системы	136
Заключение	138

Часть III. ИЗВЛЕЧЕНИЕ, ПРЕОБРАЗОВАНИЕ И ВИЗУАЛИЗАЦИЯ ДАННЫХ

Глава 8. Загрузка, преобразование и обновление данных.....

Технические требования.....	142
Основные принципы преобразования данных.....	142
Обновление данных, параллелизм и использование ресурсов	142
Улучшение среды разработки.....	145
Свертывание запросов, объединение и агрегация	149
Использование добавочного обновления	152
Использование диагностики запросов	154
Сбор диагностической информации в Power Query	156
Анализ логов Power Query	157
Оптимизация потоков данных	160
Заключение	165

Глава 9. Разработка отчетов и дашбордов.....

Технические требования.....	166
Оптимизация интерактивных отчетов.....	167
Управление визуальными элементами и запросами.....	167
Установите выбор по умолчанию в срезах/фильтрах для первой загрузки.....	168
Избегайте вывода подробных таблиц со множеством столбцов в базовом отчете.....	169
Объединяйте индивидуальные карточки в многострочные или в таблицы	170
Используйте фильтр Ведущие N для ограничения данных в отчете.....	172
Переместите редко используемые срезы на панель фильтров	173
Исключите ненужные взаимодействия пользователя с отчетом	173
Используйте всплывающие подсказки для снижения объема и сложности запросов	174

Проверяйте на производительность пользовательские визуальные элементы и отдавайте предпочтение сертифицированным элементам.....	175
Используйте технику сокращения числа запросов для сложных отчетов	176
Оптимизация дашбордов	176
Оптимизация отчетов с разбивкой на страницы.....	177
Заключение	179

Часть IV. МОДЕЛИ ДАННЫХ, ВЫЧИСЛЕНИЯ И РАБОТА С ОБЪЕМНЫМИ НАБОРАМИ..... 181

Глава 10. Моделирование данных и безопасность на уровне строк	182
Технические требования.....	183
Построение эффективных моделей данных.....	183
Теория Кимбалла и реализация схемы «звезда».....	183
Разработка схемы «звезда»	184
Работа со связями типа «многие ко многим»	187
Уменьшение размера набора данных.....	190
Ловушки при использовании безопасности на уровне строк.....	194
Заключение	199

Глава 11. Улучшаем DAX	201
Технические требования.....	201
Ловушки DAX и способы оптимизации	202
Процесс отладки выражений DAX	202
Руководство по оптимизации в DAX	203
Используйте переменные вместо повторения определений мер	203
Используйте функцию DIVIDE вместо оператора деления	205
Избегайте преобразования пустых значений в ноль или какого-то текста при вычислении числовых мер.....	206
Используйте функцию SELECTEDVALUE вместо VALUES	209
Используйте функции IFERROR и ISERROR уместно	210
Используйте функцию SUMMARIZE только с текстовыми столбцами....	210
Избегайте использования функции FILTER при передаче фильтрующих условий.....	210
Используйте функцию COUNTROWS вместо COUNT	211
Используйте функцию ISBLANK вместо BLANK	211
Оптимизируйте виртуальные связи при помощи функции TREATAS....	211
Заключение	213

Глава 12. Шаблоны работы с большими данными	215
Технические требования.....	216
Масштабирование при помощи Power BI Premium и Azure Analysis Services.....	216

Использование Power BI Premium для масштабирования данных	216
Использование Azure Analysis Services для масштабирования данных и пользователей	218
Использование горизонтального масштабирования запросов для увеличения количества пользователей	218
Использование секционирования с AAS и Premium	220
Масштабирование с использованием составных моделей и агрегатов	223
Составные модели данных	223
Использование агрегатов	226
Масштабирование с Azure Synapse и Azure Data Lake	230
Современная архитектура хранилища данных.....	232
Azure Data Lake Storage	233
Azure Synapse Analytics	233
Заключение	234
Материалы для чтения	236

Часть V. ОПТИМИЗАЦИЯ ЕМКОСТЕЙ PREMIUM И EMBEDDED

Глава 13. Оптимизация емкостей Premium и Embedded.....	238
Возможности Premium, использование ресурсов и автомасштабирование ...	239
Поведение емкостей Premium и использование ресурсов	240
Как оценивается нагрузка на емкость?	243
Перегрузка емкости и автомасштабирование	245
Управление пиковыми нагрузками при помощи автомасштабирования.....	246
Планирование емкости, мониторинг и оптимизация	248
Определение исходного размера емкости.....	249
Проверка емкости с помощью нагрузочного тестирования	250
Мониторинг использования ресурсов емкости и перегрузки	253
Исследование перегрузки	258
Заключение	266
Глава 14. Встраивание в приложения	268
Повышение производительности внедрения.....	269
Измерение производительности внедрения	273
Заключение	275
Послесловие	276
Предметный указатель.....	277

От издательства

Отзывы и пожелания

Мы всегда рады отзывам наших читателей. Расскажите нам, что вы думаете об этой книге – что понравилось или, может быть, не понравилось. Отзывы важны для нас, чтобы выпускать книги, которые будут для вас максимально полезны.

Вы можете написать отзыв на нашем сайте www.dmkpress.com, зайдя на страницу книги и оставив комментарий в разделе «Отзывы и рецензии». Также можно послать письмо главному редактору по адресу dmkpress@gmail.com; при этом укажите название книги в теме письма.

Если вы являетесь экспертом в какой-либо области и заинтересованы в написании новой книги, заполните форму на нашем сайте по адресу http://dmkpress.com/authors/publish_book/ или напишите в издательство по адресу dmkpress@gmail.com.

Список опечаток

Хотя мы приняли все возможные меры для того, чтобы обеспечить высокое качество наших текстов, ошибки все равно случаются. Если вы найдете ошибку в одной из наших книг, мы будем очень благодарны, если вы сообщите о ней главному редактору по адресу dmkpress@gmail.com. Сделав это, вы избавите других читателей от недопонимания и поможете нам улучшить последующие издания этой книги.

Нарушение авторских прав

Пиратство в интернете по-прежнему остается насущной проблемой. Издательство «ДМК Пресс» очень серьезно относится к вопросам защиты авторских прав и лицензирования. Если вы столкнетесь в интернете с незаконной публикацией какой-либо из наших книг, пожалуйста, пришлите нам ссылку на интернет-ресурс, чтобы мы могли применить санкции.

Ссылку на подозрительные материалы можно прислать по адресу электронной почты dmkpress@gmail.com.

Мы высоко ценим любую помощь по защите наших авторов, благодаря которой мы можем предоставлять вам качественные материалы.

Предисловие

Спросите любого, кто когда-либо присутствовал на конференции, посвященной базам данных, писал посты или вел блоги по этой теме, какой вопрос является наиболее актуальным во все времена, и вы наверняка получите один и тот же ответ – повышение эффективности. И если лекции по проектированию баз данных традиционно набирают достаточное количество посетителей, то на семинары, посвященные оптимизации БД, бывает, просто не пробиться. В чем здесь дело? Мне кажется, все очень просто – так же просто, как и основная цель оптимизации, состоящая в том, чтобы медленное сделать быстрым. Этому главным образом посвящена ежедневная профессиональная деятельность администраторов баз данных, разработчиков отчетов и бизнес-аналитиков. Скорость естественным образом преобразуется в удобство использования инструмента и быстроту принятия решений, что положительно сказывается на моральном духе коллектива и критически важных показателях организации. Да и сами разработчики, способные повысить скорость выполнения запросов и формирования отчетов, обычно не остаются в стороне и получают повышения и прибавку в зарплате.

Power BI в этом отношении ничем не отличается от любого другого инструмента бизнес-аналитики или базы данных. Одной из самых популярных причин недовольства пользователей является скорость формирования отчетов. В обычных условиях Power BI славится своей высокой производительностью даже при работе с довольно большими объемами данных. Но достаточно допустить небольшую ошибку при написании сложного вычисления или проектировании модели данных, и вы не оберетесь проблем. Будучи специалистом в области Power BI, вы должны уметь оптимально с точки зрения производительности проектировать модели данных и решать возникающие проблемы с отчетами.

Все это значительно повышает значимость книги, которую написал Бхавик. Несмотря на большую популярность темы оптимизации, я лично не видел до этого ни одной книги из этой области в Power BI. В этой книге собраны вместе советы, подсказки и приемы, которые раньше были беспорядочно разбросаны по официальной документации, блогам, курсам и статьям, и положены на огромный опыт автора в составе отдела разработки Power BI во взаимодействии с крупнейшими заказчиками. Вместо того чтобы сосредоточиться на одном аспекте оптимизации, например выражениях DAX, автор рассмотрел тему повышения эффективности Power BI действительно многогранно и всесторонне. В результате мы получили бесценный ресурс, способный стать краеугольным камнем на пути совершенствования навыков в деле оптимизации проектов на базе Power BI. Строго следуйте всем советам из этой книги и воплощайте их в жизнь!

*Кристофер Уэбб,
главный администратор команды Power BI CAT,
13-кратный обладатель статуса MVP
и автор множества книг в области SSAS и Power BI*

Об авторе

Бхавик Мерчант (Bhavik Merchant) обладает 18-летним опытом работы в области бизнес-аналитики и занимает пост руководителя отдела продуктовой аналитики в Salesforce. До этого работал в Microsoft сначала в роли архитектора облачных решений, а затем в статусе продуктового менеджера в проектной группе Power BI. В отделе Power BI Бхавик возглавлял программу клиентских исследований, отвечая за стратегию и техническую структуру предоставления клиентам информации о производительности системы. До Microsoft много лет работал консультантом BI-систем в отделе корпоративных клиентов. Проводил технические и теоретические тренинги в области повышения эффективности Power BI для партнеров Microsoft по всему миру.

О редакторах

Суреш Датла (Suresh Datla) работает в IT-индустрии более 20 лет и обладает большим опытом в области бизнеса и технологий. Он является разработчиком, консультантом, популяризатором и тренером по Power BI. С момента появления на рынке Azure и Power Platform тесно работает с этими системами, а также является частью проекта Microsoft по разработке и внедрению вертикальных решений. Суреш неоднократно выступал на мероприятиях от Microsoft по темам Power Platform, Power BI, Power BI Premium, безопасности и эффективности. Каждый месяц организует форум по Power Platform в Южной Калифорнии и свято верит в то, что своим успехом эта платформа всецело обязана квалифицированному сообществу. Суреш является директором компании Synergis Consulting, возглавляя группы архитектуры данных, разработчиков и инженеров.

Вишванат Музумдар (Vishwanath Muzumdar) имеет более чем 8-летний опыт работы в сфере информационных технологий и бизнес-аналитики. Специализируется на создании визуальных отчетов для клиентов. Своей целью видит применение управленческих и аналитических навыков в сфере инструментов отчетности Microsoft Power BI для помощи компании в достижении финансовых успехов.

Введение

Начать выстраивать аналитические решения с помощью Power BI очень не просто. После этого проект может жить собственной жизнью, набирая популярность и повышая объем используемых данных. Однако если не запланировать такой рост изначально, вы наверняка в какой-то момент столкнетесь с проблемами. Эта книга поможет вам провести мероприятия по оптимизации всех без исключения слоев Power BI, начиная с рабочей области отчета и заканчивая моделированием данных, их преобразованием, хранением и архитектурой.

Разработчики и архитекторы, работающие с Power BI, смогут применить полученные из этой книги знания на практике на всех стадиях жизненного цикла своих решений. Книга, которую вы держите в руках, – это не просто сборник советов и приемов по оптимизации своих проектов, но и полное структурированное руководство для обнаружения узких мест и их устранения.

Изучив все приведенные практики и примеры, вы научитесь определять распространенные ошибки на этапе проектирования данных, приводящие к снижению эффективности решения и расходованию лишней памяти. Мы рассмотрим все настройки, которые могут негативно сказываться на производительности. Вместе мы пройдем по всем слоям типичного решения Power BI и узнаем, что необходимо сделать, чтобы при масштабировании проекта не страдало его быстродействие. Начнем мы со слоя данных и постепенно поднимемся до уровня дизайна отчетов. Попутно мы рассмотрим варианты лицензирования Power BI Premium, включая процесс планирования загрузки и нагрузочное тестирование, и поговорим о службах Azure, позволяющих обеспечить дополнительное масштабирование.

Прочитав книгу, вы сможете поддерживать решения Power BI любой степени сложности с минимальными усилиями. Вдобавок вы научитесь использовать сторонние программные продукты для обнаружения проблем с производительностью.

Для кого эта книга

Книга, которую вы начинаете читать, предназначена для аналитиков данных, разработчиков в области бизнес-аналитики и специалистов по работе с Power BI. Она будет полезна тем, кто хочет создавать решения на базе Power BI, способные масштабироваться в отношении объема данных и количества пользователей без потери эффективности. Также книга поможет идентифицировать и устранить узкие места, влияющие на производительность решения. Для понимания всех концепций, описанных в этой книге, вам потребуются базовое знание Power BI и всех его компонентов.

СТРУКТУРА КНИГИ

Глава 1 «Постановка целей и определение проблемных областей». В этой главе мы рассмотрим решение на базе Power BI в виде потока данных от различных источников к потребителям в обобщенном виде. Мы посмотрим, как могут храниться и перемещаться данные в Power BI на пути к конечным потребителям. В большинстве случаев решения относительно архитектуры проекта, принятые на ранней стадии, бывает трудно и дорого отменить или изменить. Именно поэтому необходимо на самом старте проекта правильно оценивать нагрузку и планировать разработку, исходя из нее.

Глава 2 «Обзор архитектуры и конфигурации Power BI». Из этой главы вы узнаете, как можно улучшить производительность и снизить время ожидания получения информации. Здесь мы также расскажем о режимах хранения данных в Power BI и их перемещении в модель данных, поскольку решения, принятые на этом этапе, могут влиять на объемы и актуальность исходных данных. Кроме того, мы рассмотрим разные варианты развертывания шлюзов Power BI, обычно используемых для подключения к внешним источникам данных. Важность этой темы обусловлена тем, что пользователям зачастую требуется оперировать одновременно актуальными и историческими данными, объем которых ничем не ограничен.

Глава 3 «Оптимизация DirectQuery». В третьей главе книги мы познакомимся с режимом хранения DirectQuery, полагающимся на внешние источники данных. Этот режим, как правило, используется в организациях при наличии больших объемов данных. Источники DirectQuery зачастую не предназначены для аналитических запросов, что негативно сказывается на быстродействии отчетов и операций обновления данных. Мы рассмотрим методы оптимизации как в отношении Power BI, так и применительно к внешним источникам, что позволит повысить эффективность запросов.

Глава 4 «Анализ логов и метрик». В этой главе мы поговорим о том, что быстродействие отчетов может быть улучшено только в случае ее объективной оценки. Таким образом, здесь мы узнаем, где можно взять данные о производительности и как по ним определить наличие узких мест в системе. Будут рассмотрены встроенные и внешние инструменты для мониторинга показателей эффективности, а также даны полезные рекомендации по проведению такого анализа.

Глава 5 «Анализатор производительности». Здесь мы поговорим об одном из самых простых способов отслеживания временных задержек при формировании отчетов. Мы воспользуемся инструментом **Анализатор производительности**, служащим для проведения подробного анализа действий пользователей с детализацией до визуальных элементов. Мы выполним расширенный обзор всех возможностей, опишем все метрики и продемонстрируем процесс анализа на примере.

Глава 6 «Внешние инструменты». Данная глава будет посвящена сторонним инструментам, способным помочь при анализе производительности решений. Мы рассмотрим типичные сценарии использования таких инструментов с подключением к Power BI, сбором ключевых показателей эффективности и их подробным анализом.

Глава 7 «Общие принципы управления производительностью». В этой главе мы расскажем о том, что метрики и инструменты, описанные в предыдущих главах, по сути, являются строительными блоками общей системы управления показателями эффективности. При этом успех более вероятен при внедрении структурированного и воспроизводимого подхода к построению образа мышления на основе показателей эффективности на всех стадиях жизненного цикла решения в Power BI. Здесь мы приведем советы по процессу управления данными, которые помогут избежать проблем с масштабированием для новой информации и предотвратить ухудшение качества данных для имеющейся. Мы также обсудим типичные роли в рамках аналитического проекта любого уровня – от самостоятельного до управляемого из единого центра – и расскажем об их функциях в деле повышения эффективности решения.

Глава 8 «Загрузка, преобразование и обновление данных». Здесь мы поговорим о важнейшей роли периодического обновления данных для любой аналитической системы, и в Power BI это применимо к наборам данных в режиме Import. Операции по обновлению данных в этом режиме являются одними из наиболее затратных в отношении нагрузки на центральный процессор и память, и они могут повлечь серьезные задержки и даже отказы, особенно при работе с объемными наборами данных. В результате пользователи могут остаться без обновлений, процесс разработки замедлится, а ресурсы будут постоянно подвергаться огромным нагрузкам. Чтобы избежать этих проблем, требуется при преобразовании данных уделить особое внимание вопросам производительности.

Глава 9 «Разработка отчетов и дашбордов». В этой главе мы поговорим о вершине айсберга любого решения Power BI, представляющей собой отчеты и дашборды, с которыми по большей части взаимодействует пользователь. Независимо от своего визуального представления, этот слой Power BI по своей сути является приложением JavaScript, запущенным в браузере. Здесь мы коснемся ключевых приемов, позволяющих оптимизировать вывод визуального слоя, включая срезы и фильтрацию. Также мы поговорим о страничных отчетах, которые ведут себя отлично от интерактивных и обладают своими особенностями в отношении оптимизации.

Глава 10 «Моделирование данных и безопасность на уровне строк». Здесь мы подробно поговорим о наборах данных Power BI, в которых хранится исходная информация после преобразования и откуда извлекается для анализа. Таким образом, это наиболее критичная область любого решения на базе Power BI, лежащая в его основе. В то же время Power BI обладает достаточным арсеналом возможностей, которые можно применить в процессе моделирования данных. Некоторые решения способны облегчить процесс разработки ценой снижения производительности запросов и/или увеличения объема обрабатываемых данных. В этой главе мы дадим полезные советы по моделированию данных, снижению объемов задействованной информации и ускорению работы связей. В заключение коснемся темы оптимизации безопасности на уровне строк.

Глава 11 «Улучшаем DAX». В данной главе мы посмотрим, как формулы DAX позволяют разработчику расширить функционал модели данных. При

этом одного и того же результата можно добиться с применением разных формул, и не все они будут одинаково эффективными в конкретных условиях. Мы перечислим основные проблемы, связанные с формулами DAX, и научимся писать код более эффективно.

Глава 12 «Шаблоны работы с большими данными». Здесь мы узнаем, как постоянный рост объема данных в компании способен привести к серьезным проблемам. Даже с применением инновационных технологий сжатия данных, используемых в Power BI, бывает затруднительно за приемлемое время выполнять загрузку нужных нам наборов данных в режиме Import. И эта проблема усугубляется необходимостью параллельно поддерживать работу в системе сотен и тысяч пользователей. В этой главе мы рассмотрим способы борьбы с подобными проблемами, включая использование лицензирования Power BI Premium, технологий Azure, а также составных моделей и агрегатов.

Глава 13 «Оптимизация емкостей Premium и Embedded». В этой главе мы обсудим предоставляемые в рамках лицензии Power BI Premium выделенные емкости с менее строгими ограничениями, а также другие возможности, включая постраничные отчеты и службы искусственного интеллекта. Кроме того, мы поговорим о втором поколении Premium (Gen2) и узнаем, как при наличии такой подписки система справляется с повышенной нагрузкой и как работает автоматическое масштабирование. Попутно мы коснемся настроек, которые могут позволить повысить производительность системы. Мы научимся правильно планировать объемы данных и проводить нагрузочные тесты. Также мы посмотрим, как можно использовать приложение Capacity Metrics для поиска и решения проблем с нагрузкой на емкость.

Глава 14 «Встраивание в приложения». В заключительной главе книги мы посмотрим, как можно встраивать содержимое Power BI в пользовательские веб-приложения для интегрирования информации с данными из других источников. В этом случае приложение размещается на стороннем сервере при помощи вызовов API, что накладывает дополнительные ограничения. Мы поговорим о том, как можно организовать этот процесс наиболее эффективно.

КАК ИЗВЛЕЧЬ МАКСИМУМ ИЗ КНИГИ

К некоторым главам этой книги прилагаются файлы с примерами, которые можно открыть с помощью Power BI Desktop. Это поможет вам лучше понять описываемые концепции и приемы. В основном в примерах показаны ситуации до и после внесенных изменений. Вам не обязательно проверять все представленные теории на примерах, но они действительно могут помочь в их освоении.

Программное обеспечение, использованное при написании книги: ОС Windows, Power BI Desktop, DAX Studio 2.17.3, Tabular Editor 3, Power BI Helper 12.0.

Мы советуем вам всегда работать с самой свежей версией Power BI Desktop, следуя их ежемесячным обновлениям.

Если вы читаете электронную версию книги, мы советуем вам вводить код самостоятельно или копировать его из репозитория книги на GitHub (ссылка

будет дана в следующем разделе). Это поможет вам избежать ошибок при копировании скриптов из книги.

СОПРОВОДИТЕЛЬНЫЕ ФАЙЛЫ

Файлы с примерами можно загрузить из репозитория книги на GitHub по адресу <https://github.com/PacktPublishing/Microsoft-Power-BI-Performance-Best-Practices>. Все возможные обновления будут появляться там же.

Также вы можете загрузить текущую версию файлов с сайта издательства по адресу www.dmkpress.com на странице с описанием данной книги.

ЦВЕТНЫЕ ИЗОБРАЖЕНИЯ

По следующей ссылке вы можете скачать в виде PDF все рисунки и диаграммы, использованные в книге: https://static.packt-cdn.com/downloads/9781801076449_ColorImages.pdf.

УСЛОВНЫЕ ОБОЗНАЧЕНИЯ

На протяжении книги мы будем использовать следующие условные обозначения и шрифты.

Код в тексте: так в тексте книги мы будем обозначать код, имена таблиц баз данных, имена папок, файлов, расширения файлов, пути, ссылки, пользовательский ввод. Пример: «Просто скопируйте приведенный ниже код, выполните его, после чего перезапустите TabularEditor, чтобы применились новые правила».

Блоки кода будут выделены следующим образом:

```
System.Net.WebClient w = new System.Net.WebClient();
string path = System.Environment.GetFolderPath(System.Environment.SpecialFolder.
LocalApplicationData);
string url = "https://raw.githubusercontent.com/microsoft/Analysis-Services/master/
BestPracticeRules/BPARules.json";
string downloadLoc = path+@"\TabularEditor\BPARules.json";
w.DownloadFile(url, downloadLoc);
```

Новые термины, важные слова и текст, который вы видите на экране, будут выделены жирным шрифтом, например: «Раздел **Workloads** содержит настройки, связанные с производительностью».



Важные примечания будут выводиться так.



Советы будут выводиться так.



Примечания будут выводиться так.

Архитектура, узкие места и целевые показатели производительности

В этой вводной части мы дадим высокоуровневый обзор архитектуры Power BI и определим области, в которых на производительность можно влиять посредством проектных решений. После прочтения данной части вы будете понимать, как устанавливать реалистичные целевые показатели производительности.

Содержание этой части:

- глава 1 «Постановка целей и определение проблемных областей»;
- глава 2 «Обзор архитектуры и конфигурации Power BI»;
- глава 3 «Оптимизация DirectQuery».

Глава 1

Постановка целей и определение проблемных областей

При анализе производительности аналитических решений многие считают важнейшим показателем быстрдействие системы формирования отчетности. По большей части это так и есть, поскольку практически все пользователи – от операторов до управляющих – взаимодействуют с системой именно посредством отчетов как главной визуальной составляющей. Однако скоро вы узнаете, что существуют и другие не менее важные области для применения оптимизации, если смотреть на ситуацию в целом. К примеру, ускорение подсистемы формирования отчетов может не дать ожидаемых результатов, если исходный набор данных, лежащий в основе отчета, долго обновляется или выдает ошибки из-за достигнутых ограничений выделенных ресурсов. В итоге актуальные свежие данные могут просто не поспевать за великолепными и быстро формирующимися отчетами.

Автор книги, которую вы держите в руках, испытал на себе последствия снижения быстрдействия системы отчетов. Как-то раз в одной крупной коммунальной компании предприняли попытку миграции с одной системы формирования отчетности на другую, от стороннего поставщика. Несмотря на превосходство новой системы в техническом и функциональном планах, разработчики попытались напрямую перенести в нее функционал старых отчетов. В результате это решение привело к существенному снижению быстрдействия отчетов. Были потрачены миллионы на лицензирование и консультации с новым поставщиком, но большинство пользователей просто отказывались переходить на новую систему из-за ее медлительности. Этот случай мы привели в качестве демонстрации возможных последствий того, что может произойти, если изначально правильно не заложить фактор производительности в аналитическое решение.

В этой главе вы начнете свое путешествие в мир оптимизации решений на базе Microsoft Power BI. В качестве введения в полный спектр управления производительностью мы рассмотрим решения Power BI в образе потока данных от различных источников в консолидированном виде и их представления аналитикам данных и прочим пользователям, работающим с инфор-

мацией. Мы посмотрим, как данные могут храниться в Power BI и какой путь преодолевать по дороге к конечному пользователю. Многие архитектурные и проектные решения, принятые на ранней стадии становления проекта, бывает очень проблематично и дорого отменять или изменять впоследствии. В связи с этим возрастает важность всесторонней предварительной оценки возможных последствий принимаемых решений и использования подхода на основе данных к выбираемой стратегии на самом старте.

При оценке эффективности системы разработчики зачастую недооценивают или вовсе упускают из виду процесс установки *целевых показателей производительности* (performance targets). А как без этого определить, какой результата вы в итоге добились? Давайте начнем с теоретической части описания целей, после чего перейдем к техническим аспектам реализации.

Темы, которые будут рассмотрены в этой главе:

- определение уровня производительности;
- области с возможными замедлениями;
- решения, влияющие на производительность.

ОПРЕДЕЛЕНИЕ УРОВНЯ ПРОИЗВОДИТЕЛЬНОСТИ

С появлением сверхбыстрых компьютеров и средств распределенного вычисления пользователи и заказчики аналитических решений вполне обоснованно стали ожидать от них впечатляющего быстродействия, что бывает критически важно для принятия серьезных бизнес-решений. Поставщики инструментов бизнес-аналитики откликнулись на эти ожидания упоминаниями во всех своих рекламных проспектах потрясающей скорости работы предлагаемых ими решений. В результате сегодня с трудом можно найти пользователя, который воспринял бы действительно быстро формирующиеся отчеты или вовремя обновляющиеся данные как нечто необычное и поражающее воображение – скорее, как само собой разумеющееся. И наоборот, любая задержка в процессе формирования отчетов приводит к резкой негативной реакции с его стороны и жалобам во все возможные инстанции. Если такие проблемы начинают носить масштабный характер, критике начинает подвергаться как сама платформа, такая как Power BI, так и техническая команда, занимающаяся разработкой конкретного решения. В худшем случае пользователи могут отказаться работать в предлагаемом программном комплексе, в результате чего руководство может принять решение о смене платформы. Решение состоит в том, чтобы на самых ранних стадиях проекта задуматься о быстродействии строящегося проекта, поскольку исправить возникшие проблемы с производительностью при выходе на рабочие мощности с привлечением тысяч пользователей может оказаться очень сложно, если не невозможно.

Показатели производительности отчетов

Сегодня большинство решений в области бизнес-аналитики существуют в виде веб-интерфейса. При этом работа с отчетами обычно не ограни-

чивается одним лишь их формированием – пользователи с ними активно взаимодействуют. Применительно к Power BI это означает открытие отчета и дальнейшую интерактивную работу с фильтрами, срезами и визуальными элементами, детализацию до нужных уровней и переключение между страницами как непосредственно, так и с помощью закладок. Каждое взаимодействие пользователя с отчетом имеет определенное намерение, и нельзя разрывать эту связь. В нашей отрасли бытует высказывание о том, что *аналитика должна производиться со скоростью мысли*. Этот опыт и связанные с ним ожидания очень напоминают навигацию по обычному веб-сайту или взаимодействие с программным обеспечением в интернете.

Таким образом, при определении показателей производительности для аналитической системы можно воспользоваться многими наработками, принятыми в веб-студиях и используемыми на протяжении последних двух-трех десятилетий, – это не так сложно. В своей статье от 2004 года профессор Фиона На (Nah, F.) определила показатель *приемлемого времени ожидания* (tolerable wait time – TWT) для веб-пользователей. Этот показатель описывает время, которое пользователь сайта готов ждать, перед тем как закрыть веб-страницу. В своей статье профессор привела многочисленные исследования прошлых лет, ориентированные на определение пороговых временных отметок, по достижении которых у пользователя истекает терпение и появляется негатив. Из этих исследований можно сделать вывод о том, что в хорошем отчете Power BI должна полностью загружаться страница или появляться результат интерактивного взаимодействия в идеале в течение четырех секунд, а в большинстве случаев – не позже, чем через 12 секунд. При этом измерение всегда стоит производить с точки зрения пользователя, т. е. с момента вызова им отчета (например, нажатия на ссылку на веб-портале Power BI) и до завершения полной отрисовки отчета на экране.

Установка реалистичных целевых показателей производительности

Теперь, когда у нас есть руководство по установке целевых показателей на основе исследований, нам нужно применить его на практике. Распространенной ошибкой является установка единого целевого показателя для всех отчетов и ожидание, что он будет выполняться при каждом взаимодействии пользователя с любым из них. Недостаток такого подхода заключается в том, что даже хорошо спроектированная и оптимизированная система может оказаться слишком сложной для удовлетворения оптимистично установленной цели. К примеру, при наличии очень большого набора данных (десятки гигабайт) и сложного вложенного выражения DAX, результат которого отображается в табличном визуальном элементе (**Table**) с несколькими уровнями гранулярности, вам будет никак не уложиться в рамки, приемлемые для небольшого датасета с простыми агрегациями в виде суммирования и отображением в виде карточек (**Card**).

Таким образом, по причине изменчивости сложности решений и других факторов, не зависящих от разработчика (к примеру, мощности компьютера

Конец ознакомительного фрагмента.

Приобрести книгу можно

в интернет-магазине

«Электронный универс»

e-Univers.ru