

УДК 004.424
ББК 32.372
М52

Мерчант Б.

М52 Power BI: передовые методы оптимизации / пер. с англ. А. Ю. Гинько. – М.: ДМК Пресс, 2023. – 282 с.: ил.

ISBN 978-5-93700-168-9

Эта книга научит вас поддерживать решения Power BI любой степени сложности с минимальными усилиями. Вы узнаете, как проводить оптимизацию на всех слоях Power BI – начиная с рабочей области отчета и заканчивая моделированием данных, их преобразованием, хранением и архитектурой. Выясните, что необходимо сделать, чтобы при масштабировании проекта не страдало его быстродействие. Научитесь определять распространенные ошибки на этапе проектирования данных, приводящие к снижению эффективности решения и расходованию лишней памяти.

Издание предназначено для аналитиков данных, разработчиков в области бизнес-аналитики и специалистов по работе с Power BI. Оно пригодится тем, кто хочет создавать решения на базе Power BI, способные масштабироваться в отношении объема данных и количества пользователей без потери эффективности. Для изучения материала потребуется базовое знание Power BI и всех его компонентов.

УДК 004.424
ББК 32.372

First published in the English language under the title ‘Microsoft Power BI Performance Best Practices – (9781801076449).

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 978-1-80107-644-9 (англ.)
ISBN 978-5-93700-168-9 (рус.)

© 2022 Packt Publishing
© Перевод, оформление, издание,
ДМК Пресс, 2023

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

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.....

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

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
ВІ-системы самообслуживания	135
ВІ-системы на основе отдела или команды	136
Корпоративные или управляемые ИТ-отделами ВІ-системы	136
Заключение	138

Часть III. ИЗВЛЕЧЕНИЕ, ПРЕОБРАЗОВАНИЕ И ВИЗУАЛИЗАЦИЯ ДАННЫХ

140

Глава 8. Загрузка, преобразование и обновление данных

141

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

166

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

237

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