

УДК 004.42DAX
ББК 32.97
P89

Руссо М., Феррари А.

P89 Подробное руководство по DAX: бизнес-аналитика с Microsoft Power BI, SQL Server Analysis Services и Excel / пер. с англ. А. Ю. Гинько. – М.: ДМК Пресс, 2021. – 776 с.: ил.

ISBN 978-5-97060-859-3

Расширенная и дополненная с учетом современных требований и техник, эта книга представляет собой наиболее полное руководство по языку DAX, применяемому в области бизнес-аналитики, моделирования данных и анализа. Эксперты Microsoft BI Марко Руссо и Альберто Феррари излагают как основы, так и отдельные нюансы работы с DAX: от простых табличных функций до продвинутых техник программирования и оптимизации моделей. Вы узнаете, что происходит под капотом движка DAX при запуске выражений; полученные знания пригодятся при написании быстрого и надежного кода.

В книге используются примеры, которые можно запустить в бесплатной версии Power BI Desktop и разобраться во всех тонкостях синтаксиса создания переменных (VAR) в Power BI, Excel или Analysis Services.

Издание предназначено для опытных пользователей и профессионалов в сфере бизнес-аналитики, использующих в своей работе DAX и аналитические инструменты от Microsoft.

УДК 004.42DAX
ББК 32.97

Authorized Translation from the English language edition, entitled DEFINITIVE GUIDE TO DAX, THE: BUSINESS INTELLIGENCE FOR MICROSOFT POWER BI, SQL SERVER ANALYSIS SERVICES, AND EXCEL, 2nd Edition by MARCO RUSSO; ALBERTO FERRARI, published by Pearson Education, Inc, publishing as Microsoft Press. Russian-language edition copyright © 2021 by DMK Press. All rights reserved.

No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage retrieval system, without permission from Pearson Education, Inc.

Electronic RUSSIAN language edition published by DMK PRESS PUBLISHING LTD. Copyright © 2021.

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 978-1-5093-0697-8 (англ.)

ISBN 978-5-97060-859-3 (рус.)

Copyright © 2020 by Alberto Ferrari
and Marco Russo
© Оформление, издание, перевод,
ДМК Пресс, 2021

Содержание

Рецензия	14
Об авторах	15
От команды разработчиков.....	16
Благодарности	17
От издательства.....	19
Предисловие ко второму изданию	20
Предисловие к первому изданию	21
Глава 1 Что такое DAX?	27
Введение в модель данных.....	27
Введение в направление связи	29
DAX для пользователей Excel.....	31
Ячейки против таблиц	32
Excel и DAX: два функциональных языка	34
Итерационные функции в DAX.....	34
DAX требует изучения теории.....	35
DAX для разработчиков SQL.....	35
Работа со связями	35
DAX как функциональный язык	36
DAX как язык программирования и язык запросов	37
Подзапросы и условия в DAX и SQL.....	37
DAX для разработчиков MDX	38
Многомерность против табличности.....	39
DAX как язык программирования и язык запросов	39
Иерархии	40
Вычисления на конечном уровне.....	41
DAX для пользователей Power BI.....	41
Глава 2 Знакомство с DAX	43
Введение в вычисления DAX.....	43
Типы данных DAX	45
Операторы DAX	48
Конструкторы таблиц.....	49
Условные операторы	50
Введение в вычисляемые столбцы и меры.....	51
Вычисляемые столбцы.....	51
Меры.....	52
Введение в переменные	56
Обработка ошибок в выражениях DAX.....	57
Ошибки преобразования	57
Ошибки арифметических операций	58

Перехват ошибок.....	61
Генерирование ошибок	64
Форматирование кода на DAX.....	65
Введение в агрегаторы и итераторы.....	68
Использование распространенных функций DAX	71
Функции агрегирования.....	71
Логические функции	73
Информационные функции	74
Математические функции	75
Тригонометрические функции.....	76
Текстовые функции	76
Функции преобразования	77
Функции для работы с датой и временем.....	78
Функции отношений.....	79
Заключение	81

Глава 3 **Использование основных табличных функций** 83

Введение в табличные функции	83
Введение в синтаксис EVALUATE	86
Введение в функцию FILTER	87
Введение в функции ALL и ALLEXCEPT.....	90
Введение в функции VALUES, DISTINCT и пустые строки.....	94
Использование таблиц в качестве скалярных значений	100
Введение в функцию ALLSELECTED	102
Заключение	104

Глава 4 **Введение в контексты вычисления**..... 105

Введение в контексты вычисления	106
Знакомство с контекстом фильтра	106
Знакомство с контекстом строки	112
Тест на понимание контекстов вычисления.....	114
Использование функции SUM в вычисляемых столбцах	114
Использование ссылок на столбцы в мерах	115
Использование контекста строки с итераторами.....	116
Вложенные контексты строки в разных таблицах.....	117
Вложенные контексты строки в одной таблице	119
Использование функции EARLIER.....	123
Функции FILTER, ALL и взаимодействие между контекстами	125
Работа с несколькими таблицами.....	128
Контексты строки и связи	129
Контекст фильтра и связи	132
Использование функций DISTINCT и SUMMARIZE	
в контекстах фильтра.....	136
Заключение	140

Глава 5 **Функции CALCULATE и CALCULATETABLE** 142

Введение в функции CALCULATE и CALCULATETABLE.....	142
Создание контекста фильтра	143

Знакомство с функцией CALCULATE	147
Использование функции CALCULATE для расчета процентов	152
Введение в функцию KEEPFILTERS	163
Фильтрация по одному столбцу	167
Фильтрация по сложным условиям	168
Порядок вычислений в функции CALCULATE	172
Преобразование контекста	177
Повторение темы контекста строки и контекста фильтра	177
Введение в преобразование контекста	179
Преобразование контекста в вычисляемых столбцах	183
Преобразование контекста в мерах	186
Циклические зависимости	190
Модификаторы функции CALCULATE	194
Модификатор USERELATIONSHIP	195
Модификатор CROSSFILTER	198
Модификатор KEEPFILTERS	199
Использование модификатора ALL в функции CALCULATE	200
Использование ALL и ALLSELECTED без параметров	202
Правила вычисления в функции CALCULATE	203
Глава 6 Переменные	206
Введение в синтаксис переменных VAR	206
Переменные – это константы	208
Области видимости переменных	209
Использование табличных переменных	212
Отложенное вычисление переменных	214
Распространенные шаблоны использования переменных	215
Заключение	217
Глава 7 Работа с итераторами и функцией CALCULATE	219
Использование итерационных функций	219
Кратность итератора	220
Использование преобразования контекста в итераторах	223
Использование функции CONCATENATEX	226
Итераторы, возвращающие таблицы	228
Решение распространенных сценариев при помощи итераторов	232
Расчет среднего и скользящего среднего	232
Использование функции RANKX	235
Изменение гранулярности вычисления	243
Заключение	247
Глава 8 Логика операций со временем	249
Введение в логику операций со временем	249
Автоматические дата и время в Power BI	250
Автоматические столбцы с датами в Power Pivot для Excel	251

Шаблон таблицы дат в Power Pivot для Excel	251
Создание таблицы дат	253
Использование функций CALENDAR и CALENDARAUTO	254
Работа со множественными датами.....	257
Поддержка множественных связей с таблицей дат.....	257
Поддержка нескольких таблиц дат	259
Знакомство с базовыми вычислениями в работе со временем ...	260
Пометка календарей как таблиц дат	265
Знакомство с базовыми функциями логики операций со временем	266
Нарастающие итоги с начала года, квартала, месяца	268
Сравнение временных интервалов	270
Сочетание функций логики операций со временем	273
Расчет разницы по сравнению с предыдущим периодом	275
Расчет скользящей годовой суммы	276
Выбор порядка вложенности функций логики операций со временем.....	278
Знакомство с полуаддитивными вычислениями	280
Использование функций LASTDATE и LASTNONBLANK	282
Работа с остатками на начало и конец периода.....	288
Усовершенствованные методы работы с датой и временем	292
Вычисления нарастающим итогом.....	293
Функция DATEADD	296
Функции FIRSTDATE, LASTDATE, FIRSTNONBLANK и LASTNONBLANK.....	303
Использование детализации с функциями логики операций со временем.....	305
Работа с пользовательскими календарями.....	306
Работа с неделями.....	307
Пользовательские вычисления нарастающим итогом	309
Заключение	312

Глава 9 Группы вычислений.....313

Знакомство с группами вычислений	313
Создание групп вычислений.....	316
Знакомство с группами вычислений	322
Применение элемента вычисления.....	325
Очередность применения групп вычислений	334
Включение и исключение мер из элементов вычисления.....	339
Косвенная рекурсия	341
Два основных правила	346
Заключение	347

Глава 10 Работа с контекстом фильтра348

Использование функций HASONEVALUE и SELECTEDVALUE	349
Использование функций ISFILTERED и ISCROSSFILTERED	354
Понимание разницы между функциями VALUES и FILTERS.....	357

Понимание разницы между ALLEXCEPT и ALL/VALUES	359
Использование функции ALL для предотвращения преобразования контекста	364
Использование функции ISEMPTY	366
Привязка данных и функция TREATAS.....	368
Фильтры произвольной формы	372
Заключение	379
Глава 11 Работа с иерархиями.....	381
Вычисление процентов внутри иерархии	381
Работа с иерархиями типа родитель/потомок	386
Заклучение	398
Глава 12 Работа с таблицами.....	399
Функция CALCULATETABLE.....	399
Манипулирование таблицами	402
Функция ADDCOLUMNS.....	402
Функция SUMMARIZE	405
Функция CROSSJOIN.....	409
Функция UNION.....	411
Функция INTERSECT.....	415
Функция EXCEPT	417
Использование таблиц в качестве фильтров	418
Применение условных конструкций OR	419
Ограничение расчетов постоянными покупателями с первого года.....	422
Вычисление новых покупателей.....	423
Повторное использование табличных выражений при помощи функции DETAILROWS	425
Создание вычисляемых таблиц.....	427
Функция SELECTCOLUMNS	427
Создание статических таблиц при помощи функции ROW	429
Создание статических таблиц при помощи функции DATATABLE	430
Функция GENERATESERIES.....	431
Заклучение	432
Глава 13 Создание запросов.....	433
Знакомство с DAX Studio.....	433
Инструкция EVALUATE	434
Введение в синтаксис EVALUATE	434
Использование VAR внутри DEFINE	435
Использование MEASURE внутри DEFINE	437
Реализация распространенных шаблонов запросов в DAX.....	438
Использование функции ROW для проверки мер.....	439
Функция SUMMARIZE	440
Функция SUMMARIZECOLUMNS.....	442

Функция TOPN	448
Функции GENERATE и GENERATEALL	454
Функция ISONORAFTER	457
Функция ADDMISSINGITEMS	460
Функция TOPNSKIP	461
Функция GROUPBY	461
Функции NATURALINNERJOIN и NATURALLEFTOUTERJOIN	464
Функция SUBSTITUTEWITHINDEX	466
Функция SAMPLE	468
Автоматическая проверка существования данных	
в запросах DAX	469
Заключение	476

Глава 14 Продвинутое концепции языка DAX

Знакомство с расширенными таблицами	478
Функция RELATED	483
Использование функции RELATED в вычисляемых столбцах	484
Разница между фильтрами по таблице и фильтрами по столбцу	486
Использование табличных фильтров в мерах	489
Введение в активные связи	492
Разница между расширением таблиц и фильтрацией	495
Преобразование контекста в расширенных таблицах	497
Функция ALLSELECTED и неявные контексты фильтра	498
Знакомство с неявными контекстами фильтра	499
ALLSELECTED возвращает строки из итераций	503
Применение функции ALLSELECTED без параметров	506
Функции группы ALL*	506
Функция ALL	508
Функция ALLEXCEPT	509
Функция ALLNOBLANKROW	509
Функция ALLSELECTED	509
Функция ALLCROSSFILTERED	509
Использование привязки данных	510
Заклучение	512

Глава 15 Углубленное изучение связей

Реализация вычисляемых физических связей	514
Создание связей по нескольким столбцам	514
Реализация связей на основе диапазонов	517
Циклические зависимости в вычисляемых физических связях	520
Реализация виртуальных связей	523
Распространение фильтров в DAX	524
Распространение фильтра с использованием функции TREATAS	526

Распространение фильтра с использованием функции INTERSECT	527
Распространение фильтра с использованием функции FILTER.....	528
Динамическая сегментация с использованием виртуальных связей.....	529
Реализация физических связей в DAX.....	533
Использование двунаправленной кросс-фильтрации	536
Связи типа «один ко многим»	538
Связи типа «один к одному»	539
Связи типа «многие ко многим»	540
Реализация связи «многие ко многим» через таблицу-мост ..	540
Реализация связи «многие ко многим» через общее измерение	546
Реализация связи «многие ко многим» через слабые связи ..	551
Выбор правильного типа для связи	553
Управление гранулярностью	555
Возникновение неоднозначностей в связях.....	559
Появление неоднозначностей в активных связях.....	561
Устранение неоднозначностей в неактивных связях	563
Заключение	565

Глава 16 Вычисления повышенной сложности в DAX.....567

Подсчет количества рабочих дней между двумя датами.....	567
Данные о продажах и бюджетировании в одном отчете	575
Расчет сопоставимых продаж по магазинам	578
Нумерация последовательности событий	585
Вычисление продаж по предыдущему году до определенной даты.....	588
Заключение	593

Глава 17 Движки DAX

Знакомство с архитектурой движков DAX.....	594
Введение в движок формул	596
Введение в движок хранилища данных	596
Движок хранилища данных VertiPaq.....	597
Движок хранилища данных DirectQuery.....	598
Процедура обновления данных.....	599
Принципы работы движка хранилища данных VertiPaq	600
Введение в столбчатые базы данных	600
Сжатие данных движком VertiPaq.....	603
Сегментация и секционирование	613
Использование представлений динамического управления	614
Использование связей в движке VertiPaq	617
Материализация	620
Агрегирование.....	623

Выбор аппаратного обеспечения для VertiPaq	625
Возможность выбора аппаратного обеспечения	626
Приоритеты при выборе аппаратного обеспечения	626
Модель центрального процессора	627
Быстродействие памяти	628
Количество ядер процессора	628
Объем памяти	629
Дисковый ввод/вывод и постраничная подкачка	630
Заключение	630

Глава 18 Оптимизация движка VertiPaq 632

Сбор информации о модели данных	632
Денормализация	637
Кратность столбцов	645
Работа с датой и временем	646
Вычисляемые столбцы	649
Оптимизация сложных фильтров при помощи булевых вычисляемых столбцов	652
Обработка вычисляемых столбцов	653
Выбор столбцов для хранения	654
Оптимизация хранения столбцов	657
Оптимизация при помощи разделения столбцов	657
Оптимизация столбцов с высокой кратностью	658
Отключение иерархий атрибутов	659
Оптимизация атрибутов детализации	659
Управление агрегированием VertiPaq	660
Заключение	663

Глава 19 Анализ планов выполнения запросов DAX 664

Перехват запросов DAX	664
Введение в планы выполнения запросов	667
Создание плана выполнения запроса	668
Логический план выполнения запроса	669
Физический план выполнения запроса	670
Запросы движка хранилища данных	671
Сбор информации для оптимизации	672
Использование DAX Studio	673
Использование SQL Server Profiler	676
Чтение запросов движка хранилища VertiPaq	680
Введение в синтаксис xmsQL	681
Время сканирования	689
Внутренние события DISTINCTCOUNT	691
Параллелизм и кеш данных	692
Кеш движка VertiPaq	694
Функция обратного вызова CallbackDataID	696
Чтение запросов движка хранилища DirectQuery	702
Анализ составных моделей данных	703

Использование агрегатов в модели данных.....	704
Чтение планов выполнения запросов	706
Заключение	713
Глава 20 Оптимизация в DAX.....	715
Выбор стратегии оптимизации.....	716
Выделение выражения DAX для оптимизации.....	716
Создание проверочного запроса.....	719
Анализ времени выполнения запроса и информации из плана	723
Поиск узких мест в движке формул и движке хранилища данных	726
Внесение изменений и повторные запуски тестовых запросов	727
Оптимизация узких мест в выражениях DAX.....	727
Оптимизация условий фильтрации	728
Оптимизация преобразования контекста	732
Оптимизация условных выражений IF.....	739
Снижение влияния функции CallbackDataID на производительность	751
Оптимизация вложенных итераторов.....	754
Отказ от использования табличных фильтров с функцией DISTINCTCOUNT.....	761
Уход от множественных вычислений путем использования переменных.....	766
Заклучение	771
Предметный указатель	772