

Kompletny przewodnik po DAX

Analiza biznesowa przy użyciu
Microsoft Power BI,
SQL Server Analysis Services i Excel

2 wydanie
**UZUPEŁNIONE
I ROZSZERZONE**

Marco Russo i Alberto Ferrari

Przekład:
Marek Włodarz

APN Promise
Warszawa 2019

Wstęp do drugiego wydania

Kiedy postanowiliśmy, że czas już, aby odświeżyć tę książkę, sądziliśmy, że będzie to łatwe zadanie: ostatecznie nie tak wiele zmieniło się w języku DAX i teoretyczne podstawy tej książki nadal są bardzo dobre. Wydawało się nam, że wystarczy się skupić głównie na zmianie rzutów ekranu z Excela na Power BI, dodanie kilku poprawek tu i tam i wszystko będzie zrobione. Jak bardzo się myliliśmy!

Gdy tylko zaczęliśmy uaktualniać pierwszy i drugi rozdział, szybko odkryliśmy, że w rzeczywistości musimy przepisać niemal wszystko. Potrzeba ta narastała z każdą kolejną stroną. Dlatego w istocie nie jest to drugie wydanie; to zupełnie nowa książka.

Powodem nie jest to, że język zmienił się aż tak drastycznie; narzędzia też nie. Przyczyna leży w tym, że w ciągu minionych paru lat to my – jako nauczyciele i autorzy – bardzo się zmieniliśmy, miejmy nadzieję na lepsze. Uczyliśmy języka DAX tysiące użytkowników i deweloperów na całym świecie; ciężko pracowaliśmy z naszymi studentami, zawsze starając się znaleźć najlepszą metodę wyjaśniania trudnych zagadnień. W rezultacie znaleźliśmy różne sposoby opisywania języka, który tak lubimy.

W tym wydaniu zwiększyliśmy liczbę przykładów, pokazując praktyczne zastosowania funkcjonalności po przedstawieniu teoretycznych fundamentów DAX. Staraliśmy się używać prostszego stylu, ale bez rezygnowania z precyzji. Stałe kłóciłyśmy się z redaktorem o zwiększenie liczby stron, co było konieczne dla omówienia wszystkich tematów, którymi chcieliśmy się podzielić. Nie zmieniliśmy jednak podstawowej myśli przewodniej tej książki: nie zakładamy żadnej wcześniejszej wiedzy na temat DAX, mimo że książka nie jest adresowana do osób, które jedynie okazjonalnie używają tego języka. Naszymi odbiorcami powinni być ci, którzy chcą się go nauczyć, aby zdobyć pogłębione zrozumienie i umiejętności wykorzystywania całej mocy i złożoności języka DAX.

Tak, jeśli ktoś chce wykorzystać prawdziwą moc DAX, musi przygotować się na długą drogę, czytając tę książkę od deski do deski, a później jeszcze raz, szukając wielu detali, które nie były oczywiste na pierwszy rzut oka.

Spis treści

<i>Przedmowa</i>	xv
<i>Wstęp do drugiego wydania</i>	xvii
<i>Wstęp</i>	xix
<i>Podziękowania</i>	xxv
<i>O autorach</i>	xxvii
1 Czym jest DAX?	1
Istota modelu danych	2
Zrozumienie kierunku relacji	4
DAX dla użytkowników Excela	6
Komórki kontra tabele	6
Excel i DAX: dwa języki funkcyjne	9
Korzystanie z iteratorów	9
DAX wymaga nieco teorii	10
DAX dla programistów SQL	11
Obsługiwanie relacji	11
DAX jest językiem funkcyjnym	12
DAX jako język programowania i zapytań	13
Podzapytania i warunki w DAX i SQL	13
DAX dla programistów MDX	14
Model wielowymiarowy kontra tabelaryczny	14
DAX jako język programowania i zapytań	15
Hierarchie	15
Obliczenia na poziomie liści	17
DAX dla użytkowników Power BI	17
2 Wprowadzenie do DAX	19
Istota obliczeń DAX	19
Typy danych DAX	21
Operatory języka DAX	26
Konstruktory tabel	27
Wyrażenia warunkowe	28
Istota kolumn obliczanych i miar	29
Kolumny obliczane	29
Miary	31

Zmienne	35
Obsługa błędów w wyrażeniach DAX	36
Błędy konwersji	36
Błędy w operacjach arytmetycznych	37
Przechwytywanie błędów	41
Generowanie błędów	44
Formatowanie kodu DAX	45
Wprowadzenie do iteratorów i funkcji agregujących	49
Korzystanie z podstawowych funkcji DAX	52
Funkcje agregujące	52
Funkcje logiczne	54
Funkcje informacyjne	56
Funkcje matematyczne	57
Funkcje trygonometryczne	58
Funkcje tekstowe	58
Funkcje konwersji	59
Funkcje daty i czasu	60
Funkcje relacyjne	61
Podsumowanie	64
3 Podstawowe funkcje tablicowe	65
Wprowadzenie do funkcji tablicowych	65
Wprowadzenie do składni polecenia EVALUATE	68
Istota funkcji FILTER	70
Funkcje ALL i ALLEXCEPT	73
Funkcje VALUES, DISTINCT i puste wiersze	78
Używanie tabel jako wartości skalarnych	83
Funkcja ALLSELECTED	86
Podsumowanie	88
4 Istota kontekstów wykonania	89
Wprowadzenie do kontekstów wykonania	90
Istota kontekstów filtru	91
Kontekst wiersza	96
Sprawdzenie zrozumienia kontekstów wykonania	99
Użycie funkcji SUM w kolumnie obliczanej	99
Użycie kolumn w mierze	100
Tworzenie kontekstu wiersza poprzez iteratory	101
Zagnieżdżone konteksty wiersza w różnych tabelach	102
Zagnieżdżone konteksty wiersza w tej samej tabeli	104
Korzystanie z funkcji EARLIER	109
Iteratory FILTER, ALL i interakcje kontekstów	111
Praca z wieloma tabelami	114

Kontekst wiersza a relacje	115
Kontekst filtru a relacje	119
Stosowanie funkcji DISTINCT i SUMMARIZE w kontekstach filtru	123
Podsumowanie	127
5 Funkcje CALCULATE i CALCULATETABLE	129
Funkcja CALCULATE	129
Tworzenie kontekstu filtru	130
Przedstawiamy funkcję CALCULATE	134
Wykorzystanie CALCULATE do obliczania udziału procentowego	139
Funkcja KEEPFILTERS	151
Filtrowanie pojedynczej kolumny	155
Złożone warunki filtrowania	157
Kolejność przetwarzania w funkcji CALCULATE	161
Przejście kontekstu	166
Przypomnienie wiadomości o kontekście wiersza i kontekście filtru	166
Wprowadzenie do przejścia kontekstu	169
Przejście kontekstu w kolumnach obliczanych	172
Przejście kontekstu dla miar	175
Zależności cykliczne	179
Modyfikatory funkcji CALCULATE	183
Modyfikator USERRELATIONSHIP	184
Modyfikator CROSSFILTER	187
Modyfikator KEEPFILTERS	188
Istota funkcji ALL w CALCULATE	189
ALL i ALLSELECTED bez żadnych parametrów	191
Reguły dotyczące CALCULATE	192
6 Zmienne	195
Wprowadzenie do składni VAR	195
Zmienne są stałymi	197
Zakres zmiennej	199
Korzystanie ze zmiennych tablicowych	202
Istota leniwego wartościowania	203
Typowe wzorce wykorzystania zmiennych	205
Wnioski	207
7 Korzystanie z iteratorów i CALCULATE	209
Korzystanie z iteratorów	209
Istota kardynalności iteratora	210
Wykorzystywanie przejścia kontekstu w iteratorach	213
Korzystanie z CONCATENATEX	217
Iteratory zwracające tabele	220

Rozwiązywanie typowych scenariuszy przy użyciu iteratorów	223
Obliczanie średnich i średnich kroczących	223
Korzystanie z funkcji RANKX	227
Zmianie granularności obliczeń	235
Podsumowanie	240
8 Funkcje analizy czasowej	241
Wprowadzenie do analizy czasowej	241
Automatyczne funkcje Daty/czasu w Power BI	242
Automatyczne kolumny dat w Power Pivot for Excel	243
Szablon tabeli kalendarzowej w Power Pivot for Excel	243
Budowanie tabeli kalendarzowej	245
Korzystanie z funkcji CALENDAR i CALENDARAUTO	246
Praca z wieloma datami	249
Obsługa wielu relacji do tabeli Date	250
Obsługiwanie wielu tabel kalendarzowych	252
Wprowadzenie do analizy czasowej	254
Ustawienie Mark as Date Table	258
Przedstawiamy podstawowe funkcje analizy czasowej	260
Od początku okresu (roku, kwartału, miesiąca)	261
Obliczanie wartości dla wcześniejszych okresów	264
Łączenie funkcji analizy czasowej	267
Obliczanie różnic względem wcześniejszych okresów	270
Obliczanie rocznej sumy kroczącej	271
Właściwa kolejność wywoływania zagnieżdżonych funkcji analizy czasowej	274
Miary częściowo agregowalne	275
Funkcje LASTDATE i LASTNONBLANK	278
Obliczanie sald początkowych i końcowych	283
Zaawansowana analiza czasowa	288
Przedziały „do dzisiaj”	289
Funkcja DATEADD	293
Funkcje FIRSTDATE, LASTDATE, FIRSTNONBLANK i LASTNONBLANK	300
Drażnienie danych w analizie czasowej	302
Niestandardowe kalendarze	303
Praca z tygodniami	304
Niestandardowe obliczenia od początku roku, kwartału i miesiąca	308
Podsumowanie	309
9 Grupy obliczeń	311
Wprowadzenie do grup obliczeń	311
Tworzenie grup obliczeń	314
Istota grup obliczeń	321
Stosowanie elementów obliczanych	324

Pierwszeństwo grup obliczeń	333
Dołączanie i wykluczanie miar z elementów obliczanych	339
Istota rekurencji pobocznej	341
Najlepsze praktyki	346
Podsumowanie	347
10 Posługiwanie się kontekstem filtru	349
Korzystanie z funkcji HASONESVALUE i SELECTEDVALUE	350
Funkcje ISFILTERED i ISCROSSFILTERED	356
Różnice pomiędzy funkcjami VALUES i FILTERS.	359
Różnice pomiędzy funkcją ALLEXCEPT a ALL/VALUES.	361
Wykorzystanie ALL w celu uniknięcia przejścia kontekstu.	366
Korzystanie z funkcji ISEMPY	368
Wprowadzenie do rodowodu danych i TREATAS	370
Arbitralnie ukształtowane filtry	374
Podsumowanie	382
11 Hierarchie	383
Obliczanie procentowych udziałów w hierarchiach	383
Obsługa hierarchii rodzic-dziecko	388
Podsumowanie	401
12 Praca z tabelami	403
Funkcja CALCULATETABLE	403
Przekształcanie tabel	406
Funkcja ADDCOLUMNS	406
Funkcja SUMMARIZE	410
Funkcja CROSSJOIN	413
Funkcja UNION	416
Funkcja INTERSECT	420
Funkcja EXCEPT	422
Wykorzystywanie tabel jako filtrów	424
Implementowanie alternatywy	424
Zawężanie obliczeń do klientów z pierwszego roku	428
Znajdowanie nowych klientów	429
Ponowne wykorzystanie wyrażeń tablicowych dzięki DETAILROWS	432
Tworzenie tabel obliczanych	434
Funkcja SELECTCOLUMNS	434
Tworzenie tabel statycznych przy użyciu funkcji ROW	436
Tworzenie tabel statycznych przy użyciu DATATABLE	437
Funkcja GENERATESERIES	438
Podsumowanie	439

13 Budowanie zapytań	441
Wprowadzenie do DAX Studio	441
Istota funkcji EVALUATE	442
Składnia EVALUATE	442
Używanie VAR w sekcji DEFINE	444
Korzystanie z MEASURE w sekcji DEFINE	446
Implementowanie typowych wzorców zapytań	447
Wykorzystanie funkcji ROW do testowania miar	447
Korzystanie z funkcji SUMMARIZE	448
Korzystanie z funkcji SUMMARIZECOLUMNS	450
Korzystanie z funkcji TOPN	457
Korzystanie z funkcji GENERATE i GENERATEALL	464
Korzystanie z funkcji ISONORAFTER	468
Korzystanie z funkcji ADDMISSINGITEMS	470
Korzystanie z funkcji TOPNSKIP	471
Funkcja GROUPBY	471
Korzystanie z funkcji NATURALINNERJOIN i NATURALLEFTOUTERJOIN	474
Korzystanie z funkcji SUBSTITUTEWITHINDEX	477
Korzystanie z funkcji SAMPLE	479
Istota zachowania <i>auto-exists</i> w zapytaniach DAX	480
Podsumowanie	487
14 Zaawansowane koncepcje języka DAX	489
Wprowadzenie do tabel rozszerzonych	489
Istota funkcji RELATED	494
Używanie funkcji RELATED w kolumnach obliczanych	496
Istota różnicy pomiędzy filtrami tabel a filtrami kolumn	497
Używanie filtrów tablicowych w miarach	501
Relacje aktywne	504
Rozróżnienie pomiędzy rozszerzaniem tabel a filtrowaniem	507
Przejsięcie kontekstu w tabelach rozszerzonych	509
Istota funkcji ALLSELECTED i przesłanianych kontekstów filtru	510
Pojęcie przesłanianych kontekstów filtru	511
ALLSELECTED zwraca iterowane wiersze	515
ALLSELECTED bez parametrów	518
Rodzina funkcji ALL*	518
ALL	520
ALLEXCEPT	521
ALLNOBLANKROW	521
ALLSELECTED	521
ALLCROSSFILTERED	521
Istota rodowodu danych	522
Podsumowanie	525

15	Zaawansowana obsługa relacji	527
	Stosowanie obliczanych relacji fizycznych	527
	Tworzenie relacji wielokolumnowych	528
	Implementowanie relacji opartych na zakresach	530
	Zależności cykliczne w fizycznych relacjach obliczanych	533
	Implementowanie relacji wirtualnych	537
	Przenoszenie filtrów w DAX	537
	Transferowanie filtru przy użyciu TREATAS	540
	Transferowanie filtru przy użyciu INTERSECT	541
	Transferowanie filtru przy użyciu FILTER	542
	Implementacja dynamicznej segmentacji przy użyciu relacji wirtualnych	543
	Istota mechanizmu relacji fizycznych	546
	Używanie dwukierunkowego filtrowania	550
	Istota relacji jeden-do-wielu	552
	Istota relacji jeden-do-jednego	552
	Istota relacji wiele-do-wielu	553
	Implementowanie relacji wiele-do-wielu przy użyciu tabeli mostka	553
	Implementowanie relacji wiele-do-wielu przy użyciu wspólnego wymiaru	559
	Implementowanie wiele-do-wielu przy użyciu słabych relacji mmr	564
	Wybieranie właściwego typu relacji	567
	Zarządzanie granularnościami	568
	Zarządzanie niejednoznacznościami w relacjach	573
	Niejednoznaczność w aktywnych relacjach	575
	Rozwiązywanie niejednoznaczności w nieaktywnych relacjach	577
	Podsumowanie	579
16	Zaawansowane obliczenia w języku DAX	581
	Obliczanie liczby dni roboczych pomiędzy dwiema datami	581
	Pokazywanie łącznie danych budżetu i sprzedaży	590
	Obliczanie sprzedaży w tym samym sklepie	593
	Numerowanie sekwencji zdarzeń	600
	Obliczanie wartości sprzedaży z poprzedniego roku do wskazanej daty	603
	Podsumowanie	609
17	Silniki DAX	611
	Architektura silników DAX	612
	Istota silnika zapytań	613
	Istota silnika magazynowego	614
	Wprowadzenie do silnika magazynowego VertiPaq	615
	Wprowadzenie do silnika magazynowego DirectQuery	616
	Istota odświeżania danych	616
	Działanie silnika magazynowego VertiPaq	617
	Wprowadzenie do kolumnowych baz danych	618

Istota kompresji VertiPaq	621
Istota ponownego kodowania	627
Znajdowanie najlepszego uporządkowania	628
Segmentacja i partycjonowanie	631
Korzystanie z dynamicznych widoków zarządzania	632
Wykorzystywanie relacji w VertiPaq	635
Wprowadzenie do materializacji	638
Wprowadzenie do agregacji	641
Wybieranie sprzętu dla bazy danych VertiPaq	643
Czy możemy wybrać sprzęt?	644
Ustalanie priorytetów sprzętowych	644
Model procesora	645
Szybkość pamięci	646
Wielkość pamięci	646
Liczba rdzeni	647
Dyskowe operacje I/O i stronicowanie	647
Najlepsze praktyki przy wybieraniu sprzętu	648
Podsumowanie	648
18 Optymalizowanie VertiPaq	649
Gromadzenie informacji o modelu danych	650
Denormalizacja	655
Kardynalność kolumn	662
Obsługa daty i czasu	663
Kolumny obliczane	667
Optymalizowanie złożonych filtrów przy użyciu logicznych kolumn obliczanych	670
Przetwarzanie kolumn obliczanych	671
Wybieranie właściwych kolumn do przechowania	672
Optymalizowanie przechowywania kolumn	675
Optymalizacja przez podział kolumny	675
Optymalizowanie kolumn o wysokiej kardynalności	677
Wyłączanie hierarchii atrybutów	677
Optymalizowanie atrybutów drążenia w głąb	678
Zarządzanie agregacjami VertiPaq	678
Podsumowanie	682
19 Analizowanie planów zapytań	683
Przechwytywanie zapytań DAX	683
Wprowadzenie do planów zapytań DAX	687
Gromadzenie planów zapytań	688
Logiczny plan zapytania	689
Fizyczny plan zapytania	689

Zapytanie do silnika magazynowego	691
Przechwytywanie informacji profilowania	692
Korzystanie z DAX Studio	693
Korzystanie z SQL Server Profiler	696
Czytanie zapytań do silnika magazynowego VertiPaq	700
Wprowadzenie do składni xSQL	701
Czas skanowania	710
Wewnętrzne mechanizmy funkcji DISTINCTCOUNT	712
Istota równoległości i buforów danych	713
Pamięć podręczna VertiPaq	715
Istota elementu CallbackDataID	718
Czytanie zapytań do silnika DirectQuery	724
Analizowanie modeli kompozytowych	726
Używanie agregacji w modelu danych	727
Czytanie planów zapytań	729
Podsumowanie	737
20 Optymalizowanie kodu DAX	739
Definiowanie strategii optymalizacji	740
Identyfikacja pojedynczego wyrażenia DAX, które wymaga optymalizacji ..	740
Utworzenie zapytania reprodukującego problem	744
Analiza czasów wykonania i informacji zawartych w planie zapytania	748
Identyfikacja wąskich gardeł w silniku magazynowym lub silniku formuł ..	751
Implementowanie zmian i ponowne uruchamianie zapytania testowego ..	752
Optymalizowanie wąskich gardeł silnika magazynowego	752
Optymalizowanie warunków filtrowania	753
Optymalizowanie przejść kontekstu	757
Optymalizowanie warunków IF	764
Redukowanie wpływu CallbackDataID	777
Optymalizowanie zagnieżdżonych iteratorów	781
Unikanie filtrów tablicowych dla funkcji DISTINCTCOUNT	788
Unikanie wielokrotnego obliczania przy użyciu zmiennych	793
Podsumowanie	798
<i>Indeks</i>	799