

Kompletny przewodnik po DAX

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

Marco Russo i Alberto Ferrari

Przekład:
Marek Włodarz

APN Promise
Warszawa 2016

Spis treści

<i>Wstęp</i>	xi
1 Czym jest DAX?	1
Istota modelu danych	2
Zrozumienie kierunku relacji	3
DAX dla użytkowników programu Excel	6
Komórki kontra tabele	6
Excel i DAX: dwa języki funkcyjne	8
Korzystanie z iteratorów	9
DAX wymaga nieco teorii	9
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 projektantów MDX	14
Model wielowymiarowy kontra tabelaryczny	14
DAX jako język programowania i zapytań	15
Hierarchie	15
Obliczenia na poziomie liści	17
2 Wprowadzenie do DAX	19
Istota obliczeń DAX	19
Typy danych DAX	20
Operatory języka DAX	24
Kolumny obliczane i miary	25
Kolumny obliczane	25
Miary	26
Zmienne	29
Obsługa błędów w wyrażeniach DAX	30
Błędy konwersji	30
Operacje arytmetyczne	31
Przechwytywanie błędów	34

Formatowanie kodu DAX	36
Typowe funkcje DAX	40
Funkcje agregujące.....	40
Funkcje logiczne.....	42
Funkcje informacyjne.....	44
Funkcje matematyczne	45
Funkcje trygonometryczne	46
Funkcje tekstowe	46
Funkcje konwersji	47
Funkcje daty i czasu	48
Funkcje relacyjne.....	48
3 Korzystanie z podstawowych funkcji tablicowych.....	51
Wprowadzenie to funkcji tablicowych	51
Składnia polecenia <i>EVALUATE</i>	54
Korzystanie z wyrażeń tablicowych	57
Istota funkcji <i>FILTER</i>	58
Funkcje <i>ALL</i> , <i>ALLEXCEPT</i> oraz <i>ALLNOBLANKROW</i>	61
Funkcje <i>VALUES</i> oraz <i>DISTINCT</i>	65
Używanie <i>VALUES</i> jako wartości skalarnej	67
4 Istota kontekstów wykonania.....	69
Wprowadzenie do kontekstów wykonania	70
Kontekst wiersza	74
Testowanie zrozumienia kontekstów wykonania.....	75
Użycie funkcji <i>SUM</i> w kolumnie obliczanej	75
Użycie kolumn w mierze	77
Tworzenie kontekstu wiersza poprzez iteratory	78
Korzystanie z funkcji <i>EARLIER</i>	79
Iteratory <i>FILTER</i> , <i>ALL</i> i interakcje kontekstów.....	83
Praca z wieloma tabelami	86
Kontekst wiersza a relacje	88
Kontekst filtru a relacje	91
Funkcja <i>VALUES</i>	94
Funkcje <i>ISFILTERED</i> i <i>ISCROSSFILTERED</i>	95
Podsumowanie kontekstów wykonania	98
Tworzenie tabeli parametrów.....	100
5 Funkcje <i>CALCULATE</i> i <i>CALCULATE</i>TABLE	103
Funkcja <i>CALCULATE</i>	104
Istota kontekstu filtru.....	106
Przedstawiamy funkcję <i>CALCULATE</i>	108
Przykłady użycia funkcji <i>CALCULATE</i>	111

Filtrowanie pojedynczej kolumny.	112
Złożone warunki filtrowania	117
Korzystanie z <i>CALCULATETABLE</i>	120
Istota przejścia kontekstu	122
Przejście kontekstu dla miar	125
Ile wierszy jest widocznych po przejściu kontekstu?.	128
Kolejność wykonywania działań przy przejściu kontekstu.	129
Zmienne i konteksty wykonania.	130
Zależności cykliczne	131
Reguły dotyczące <i>CALCULATE</i>	135
Funkcja <i>ALLSELECTED</i>	136
Funkcja <i>USERRELATIONSHIP</i>	138
6 Przykłady kodu DAX	141
Obliczanie proporcji i udziałów procentowych	141
Obliczanie sum bieżących (skumulowanych).	145
Korzystanie z klasyfikacji ABC (Pareto)	148
Obliczanie sprzedaży dziennej i na dzień roboczy	156
Obliczanie różnic w dniach roboczych	163
Obliczanie statycznych średnich ruchomych	165
7 Funkcje analizy czasowej.	169
Wprowadzenie do analizy czasowej	169
Budowanie tabeli kalendarzowej	170
Korzystanie z funkcji <i>CALENDAR</i> i <i>CALENDARAUTO</i>	172
Praca z wieloma datami.	175
Obsługa wielu relacji do tabeli <i>Date</i>	176
Obsługiwanie wielu tabel kalendarzowych.	177
Wprowadzenie do analizy czasowej	179
Ustawienie <i>Mark as Date Table</i>	181
Agregowanie i porównywanie danych względem czasu	184
Od początku okresu (roku, kwartału, miesiąca).	184
Obliczanie wartości dla wcześniejszych okresów.	187
Obliczanie różnic względem wcześniejszych okresów	190
Obliczanie rocznej sumy ruchomej	191
Bilans zamknięcia względem czasu	195
Miary częściowo agregowalne	195
Funkcje <i>OPENINGBALANCE</i> i <i>CLOSINGBALANCE</i>	201
Zaawansowana analiza czasowa	205
Przedziały „do dzisiaj”	206
Funkcja <i>DATEADD</i>	208
Funkcje <i>FIRSTDATE</i> i <i>LASTDATE</i>	214
Funkcje <i>FIRSTNONBLANK</i> i <i>LASTNONBLANK</i>	217

Drażnienie danych w analizie czasowej	218
Niestandardowe kalendarze	218
Praca z tygodniami	219
Niestandardowe obliczenia od początku roku, kwartału i miesiąca	222
Obliczenia dla nieciągłych zakresów dat	224
Niestandardowe porównania pomiędzy okresami	228
8 Funkcje statystyczne	231
Funkcja <i>RANKX</i>	231
Typowe problemy związane z <i>RANKX</i>	234
Funkcja <i>RANK.EQ</i>	238
Obliczanie średnich i ruchomych średnich	238
Obliczanie wariancji i odchylenia standardowego	240
Obliczanie mediany i percentyli	242
Obliczanie odsetek	244
Alternatywna implementacja funkcji <i>PRODUCT</i> i <i>GEOMEAN</i>	246
Korzystanie z wewnętrznej stopy zwrotu (<i>XIRR</i>)	246
Korzystanie z bieżącej wartości netto (<i>XNPV</i>)	247
Korzystanie z funkcji statystycznych programu Excel	249
Próbkowanie przy użyciu funkcji <i>SAMPLE</i>	250
9 Zaawansowane funkcje tablicowe	253
Funkcja <i>EVALUATE</i>	253
Korzystanie ze zmiennych w funkcji <i>EVALUATE</i>	256
Stosowanie funkcji filtrujących	257
Funkcja <i>CALCULATETABLE</i>	257
Funkcja <i>TOPN</i>	259
Istota funkcji rzutujących	261
Funkcja <i>ADDCOLUMNS</i>	261
Funkcja <i>SELECTCOLUMNS</i>	265
Funkcja <i>ROW</i>	268
Powiązanie a relacje	269
Funkcje grupujące/złączające	271
Funkcja <i>SUMMARIZE</i>	272
Funkcja <i>SUMMARIZECOLUMNS</i>	277
Funkcja <i>GROUPBY</i>	283
Funkcja <i>ADDMISSINGITEMS</i>	284
Funkcja <i>NATURALINNERJOIN</i>	287
Funkcja <i>NATURALLEFTOUTERJOIN</i>	288
Funkcje zbiorów	289
Funkcja <i>CROSSJOIN</i>	289
Funkcja <i>UNION</i>	291
Funkcja <i>INTERSECT</i>	295

Funkcja <i>EXCEPT</i>	296
Funkcje <i>GENERATE</i> i <i>GENERATEALL</i>	298
Stosowanie funkcji narzędziowych	300
Funkcja <i>CONTAINS</i>	300
Funkcja <i>LOOKUPVALUE</i>	302
Funkcja <i>SUBSTITUTEWITHINDEX</i>	305
Funkcja <i>ISONORAFTER</i>	306
10 Zaawansowane konteksty wykonania	307
Działanie funkcji <i>ALLSELECTED</i>	307
Istota funkcji <i>KEEPFILTERS</i>	316
Istota funkcjonalności <i>AutoExists</i>	326
Pojęcie tabel rozszerzonych	330
Różnice pomiędzy rozszerzaniem tabeli a filtrowaniem	338
Redefiniowanie kontekstu filtru	339
Przecięcia kontekstów filtru	342
Zastępowanie kontekstu filtru	344
Arbitralnie kształtowane filtry	345
Działanie funkcji <i>ALL</i>	350
Istota powiązania z elementem nadrzędnym	352
Korzystanie z zaawansowanych wyrażeń filtrowania	355
Dalsze poznawanie kontekstów wykonania	361
11 Hierarchie	363
Obliczanie procentowych udziałów w hierarchiach	363
Obsługa hierarchii drzewiastych	371
Operatory jednoargumentowe	384
Implementowanie operatorów jednoargumentowych przy użyciu <i>DAX</i>	386
12 Zaawansowana obsługa relacji	395
Stosowanie obliczanych fizycznych relacji	395
Tworzenie relacji wielokolumnowych	395
Obliczanie statycznego grupowania	397
Korzystanie z relacji wirtualnych	399
Dynamiczne grupowanie	399
Relacje wiele-do-wielu	402
Korzystanie z relacji o różnym stopniu szczegółowości	408
Różnice pomiędzy relacjami fizycznymi a wirtualnymi	411
Wyszukiwanie braku relacji	412
Obliczanie liczby niesprzedanych produktów	412
Znajdowanie nowych i powracających klientów	414
Przykłady złożonych relacji	416
Wykonywanie konwersji walut	416

	Wyszukiwanie zbiorów częstych	422
13	Silnik bazodanowy VertiPaq	429
	Istota przetwarzania baz danych	430
	Wprowadzenie do kolumnowych baz danych	431
	Istota kompresji VertiPaq	434
	Kodowanie wartości	435
	Istota kodowania słownikowego	436
	Algorytm Run Length Encoding (RLE)	437
	Istota ponownego kodowania	440
	Znajdowanie najlepszego uporządkowania	441
	Hierarchie i relacje	442
	Segmentacja i partycjonowanie	443
	Korzystanie z dynamicznych widoków zarządzania	445
	Korzystanie z DISCOVER_OBJECT_MEMORY_USAGE	446
	Korzystanie z DISCOVER_STORAGE_TABLES	447
	Korzystanie z DISCOVER_STORAGE_TABLE_COLUMNS	447
	Korzystanie z DISCOVER_STORAGE_TABLE_COLUMN_SEGMENTS	448
	Materializacja	449
	Wybieranie odpowiedniego sprzętu dla bazy danych VertiPaq	454
	Czy możemy wybrać sprzęt?	454
	Ustalanie priorytetów sprzętowych	455
	Model procesora	455
	Szybkość pamięci	456
	Liczba rdzeni	457
	Wielkość pamięci	457
	Dyskowe operacje I/O i stronicowanie	458
	Podsumowanie	458
14	Optymalizowanie modelu danych	459
	Gromadzenie informacji o modelu danych	460
	Denormalizacja	471
	Kardynalność kolumn	478
	Obsługa daty i czasu	480
	Kolumny obliczane	484
	Optymalizowanie złożonych filtrów przy użyciu logicznych kolumn obliczanych	487
	Wybieranie właściwych kolumn do przechowania	488
	Optymalizowanie przechowywania kolumn	491
	Optymalizacja przez podział kolumny	491
	Optymalizowanie kolumn o wysokiej kardynalności	492
	Optymalizowanie atrybutów drążenia danych	493

15	Analizowanie planów wykonania	495
	Wprowadzenie do silnika zapytań DAX	495
	Istota silnika zapytań	496
	Istota silnika magazynowego (VertiPaq)	497
	Wprowadzenie do planów zapytań DAX	498
	Logiczny plan zapytania	499
	Fizyczny plan zapytania	500
	Zapytanie do silnika magazynowego	501
	Przechwytywanie informacji profilowania	502
	Korzystanie z SQL Server Profiler	502
	Korzystanie z DAX Studio	507
	Odczytywanie zapytań do silnika magazynowego	510
	Wprowadzenie do składni xSQL	510
	Czas skanowania	516
	Wewnętrzne mechanizmy funkcji <i>DISTINCTCOUNT</i>	518
	Istota równoległości i buforów danych	519
	Pamięć podręczna VertiPaq	521
	Istota elementu <i>CallbackDataID</i>	523
	Czytanie planów zapytań	529
16	Optymalizowanie kodu DAX	537
	Definiowanie strategii optymalizacji	538
	Identyfikacja pojedynczego wyrażenia DAX, które wymaga optymalizacji	538
	Utworzenie zapytania reprodukującego problem	542
	Analiza czasów wykonania i informacji zawartych w planie zapytania	544
	Identyfikacja wąskich gardeł w silniku magazynowym lub silniku formuł	547
	Optymalizowanie wąskich gardeł silnika magazynowego	548
	Wybór pomiędzy <i>ADDCOLUMNS</i> a <i>SUMMARIZE</i>	549
	Redukowanie wpływu <i>CallbackDataID</i>	554
	Optymalizowanie warunków filtrowania	557
	Optymalizowanie warunków <i>IF</i>	558
	Optymalizacja kardynalności	560
	Optymalizowanie zagnieżdżonych iteratorów	563
	Optymalizowanie wąskich gardeł w silniku formuł	569
	Tworzenie zapytania repro w języku MDX	574
	Redukowanie materializacji	575
	Optymalizowanie złożonych wąskich gardeł	580
	<i>Indeks</i>	587
	<i>O autorach</i>	610