

Usługi biznesowe w SQL Server 2008. Omówienie oraz przykład zastosowania w przemyśle

Artur Gramacki, Jarosław Gramacki

Streszczenie: W pracy zaprezentowano składniki systemu bazodanowego SQL Server 2008 firmy Microsoft tworzące kompletną platformę do świadczenia tzw. usług biznesowych. Pokazano przykład jej zastosowania, wykorzystując dane pobrane z rzeczywistego systemu klasy ERP. Praca jest efektem 3 miesięcznego stażu, jaki pierwszy autor artykułu odbył w firmie LUMEL S.A. w Zielonej Górze. Staż ten był częścią unijnego projektu „Zróbmy coś razem”.

Słowa kluczowe: SQL Server 2008, Business Intelligence, hurtownia danych, system klasy ERP, projekt unijny „Zróbmy coś razem”

1. WPROWADZENIE

Z pewnością prawdziwym jest stwierdzenie, że jednym z najważniejszych kluczy do sukcesu firmy jest umiejętne wykorzystywanie posiadanych informacji do podejmowania kluczowych i trafnych decyzji biznesowych. Wsparcie informatyczne w tym obszarze zapewniają systemy świadczące tzw. usługi biznesowe (ang. *Business Intelligence*, BI). System BI nie jest jakimś jednym zamkniętym programem, lecz raczej jest to zbiór programów i technologii informatycznych umożliwiających zbieranie i przetwarzanie danych w informacje a informacji w użyteczną wiedzę w celu ułatwienia użytkownikowi podejmowania lepszych decyzji biznesowych [1].

Obecnie, niezależnie od wielkości firmy, prawie każda z nich posiada jakiś system do gromadzenia danych. Najczęściej jest to baza relacyjna, którą, z racji pełnionych funkcji, zwykło się nazywać bazą OLTP (ang. *OnLine Transaction Processing*). Baza taka zwykle bardzo dobrze radzi sobie z gromadzeniem i przetwarzaniem bieżących danych, jednak niespecjalnie nadaje się do wykonywania analiz danych historycznych – zwłaszcza, gdy danych tych jest bardzo dużo. Kluczowymi parametrami baz OLTP są szybkość przetwarzania transakcji oraz możliwość efektywnej pracy w środowisku wieloużytkownikowym.

Aby sprawnie analizować duże wolumeny danych historycznych, często pochodzących z bardzo różnych źródeł, w tym również oczywiście z baz OLTP, wymagane są systemy bazodanowe o specyficznych cechach. Szybkie przetwarzanie transakcyjne oraz praca wieloużytkownikowa nie są już wymaganiami kluczowymi. Tym razem najważniejsze są możliwości analizowania bardzo dużych wolumenów danych, wykonywania na nich nierzadko

skomplikowanych analiz oraz poszukiwania wartościowych informacji biznesowych. Systemy takie noszą nazwę OLAP (ang. *OnLine Analytical Processing*) i są ściśle związane z koncepcją hurtowni danych (ang. *Data Warehouse*, HD) [2], czyli najogólniej mówiąc specjalnie zaprojektowanych baz danych, służących do przechowywania i analizowania dużych wolumenów danych historycznych. W systemach OLAP często korzysta się też z najprzeróżniejszych zaawansowanych metod i algorytmów wspomagających wykonywanie tzw. eksploracji danych (ang. *data mining*, DM).

HD zasilane są danymi z operacyjnych baz danych (bazy OLTP, ale też i inne nierelacyjne źródła danych, np. arkusze kalkulacyjne). Poważnym wyzwaniem dla twórców HD jest zapewnienie efektywnych mechanizmów ich zasilania, czyszczenia danych źródłowych zanim zostaną one zeskalowane w HD, oraz aktualizowania jej zawartości. Ogół zadań z tym związanych zwykło nazywać się procesami ETL (ang. *Extraction Transformation Loading*). Zdefiniowanie, utworzenie oraz utrzymanie procesów ETL jest często najtrudniejszym i najbardziej pracochłonnym etapem tworzenia hurtowni danych.

Wspomniane więc na początku tego rozdziału systemy klasy BI można uznać za umiejętne połączenie technik OLTP, OLAP, HD, DM oraz ETL. Należy również dodać w tym miejscu, że ważnym elementem składowym systemów BI są moduły wspierające tworzenie różnego rodzaju raportów i zestawień. W dalszej części pracy pokazano implementację BI zaproponowaną przez firmę Microsoft w ramach produktu SQL Server 2008. Omówiono też jak został on w praktyce wykorzystany przez autorów pracy.

2. SQL SERVER SQL 2008 JAKO PLATFORMA BUSINESS INTELLIGENCE

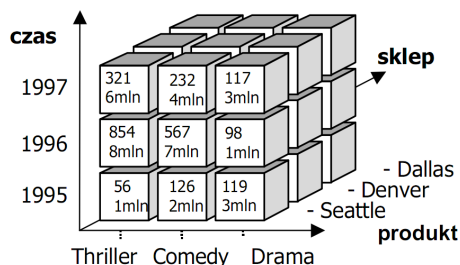
Najnowsza wersja SQL Server 2008 została wyposażona w rozwiązania pozwalające uznać ją za kompletną platformę BI. Jest ona ponadto dobrze zintegrowana z popularnym arkuszem Excel 2007 (który stał się de facto niepisanym standardem w bardzo wielu firmach¹), co powoduje, że staje się bardzo ciekawą ofertą

¹ Z obserwacji autorów wynika, że niestety w polskich firmach nie za bardzo przyjął się pakiet OpenOffice, będący niejako alternatywą dla pakietu Microsoft Office.

dla przemysłu. System SQL Server 2008 oferowany jest w dwóch podstawowych wersjach komercyjnych (tzw. *Core Editions*): tańszej *Standard* oraz droższej *Enterprise*. Dodatkowo dostępna jest darmowa wersja *Express*, jednak o dość mocno ograniczonych możliwościach w stosunku do wersji komercyjnych [3]. Gdy planowane rozwiązanie BI nie będzie operowało na ekstremalnie dużych ilościach danych, prawdopodobnie wersja *Standard* będzie wystarczająca. Obie wersje posiadają podobną funkcjonalność, z tym że wersja *Enterprise* daje większe wsparcie dla przetwarzania dużych wolumenów danych. SQL Server 2008 składa się z wielu różnych modułów. Z punktu widzenia platformy BI jego najważniejsze elementy to:

- *Database Engine Services* (transakcyjna baza danych OLTP),
- *Analysis Services* (SSAS, analityczna baza danych OLAP),
- *Reporting Services* (SSRS, raportowanie),
- *Integration Services* (SSIS, wsparcie dla procesów ETL),
- *Business Intelligence Development Studio* (BIDS, podstawowe narzędzie projektanta i programisty systemów BI),
- *SQL Server Management Studio* (administracja systemem jako całością).

Podstawowymi obiektami baz analitycznych są tzw. kostki wielowymiarowe (zwane też kostkami OLAP-owymi) [2] a podstawowy model danych nosi nazwę modelu wielowymiarowego. Ideę kostki wielowymiarowej pokazuje rysunek 1 (zaczepiony z [5]).



Rys. 1. Przykładowa kostka analityczna OLAP

Na rysunku pokazano kostkę 3-wymiarową ale uogólnienie na N wymiarów jest oczywiste. Kostka jest zbiorem komórek, które przechowują określone dane, zwane tu faktami. Na kostkach analitycznych można wykonywać typowe operacje, takie jak:

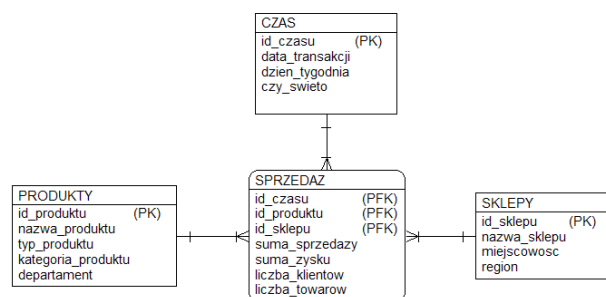
- Zwijanie, rozwijanie (ang. *roll-up*, *roll-down*)²,
- wycinanie (ang. *sliping and dicing*),
- obracanie (ang. *rotating*),
- filtrowanie (ang. *filtering*),
- wyznaczanie rankingu (ang. *ranking*).

Budowa kostek analitycznych umożliwia łatwe uzyskiwanie odpowiedzi na pytania biznesowe typu: „Jaka była sprzedaż w ciągu ostatnich 12 miesięcy, z podziałem na poszczególne miasta oraz kategorie produktów?”.

Kostka analityczna jako taka jest tworem abstrakcyjnym. W systemach analitycznych jest ona implementowana w jednym z dwóch modeli fizycznych noszących nazwy MOLAP (ang. *Multidimensional OLAP*) oraz ROLAP (ang. *Relational OLAP*).

W modelu MOLAP dane oraz wyliczane na ich podstawie agregaty (inaczej: podsumowania) są składowane, indeksowane i przetwarzane w specjalnych tablicach wielowymiarowych, a nie w tradycyjnych tablicach relacyjnych. Dane zeskładowane w hurtowni danych (zwykle w postaci tradycyjnych tabel relacyjnych) muszą więc być skopiowane do bazy analitycznej i zeskładowane w jej specyficznych logiczno-fizycznych strukturach. Mamy więc jakby 2. oddzielne bazy danych, co rodzi automatycznie określone konsekwencje (np. konieczność ich synchronizacji).

W modelu ROLAP dane przechowywane są w tradycyjnych tabelach relacyjnych w jednej z dwóch podstawowych struktur: tzw. gwiazdy (ang. *star schema*) oraz tzw. płatka śniegu (ang. *snowflake schema*). Pierwsza z nich, jako zdecydowanie częściej używana w praktyce, pokazana została na rysunku 2 (zaczepionym z [5]). Implementuje on kostkę pokazaną na rysunku 1. Tabela *sprzedaż* nosi nazwę tabeli faktów (ang. *facts*, inna stosowana nazwa to tabela miar, ang. *measures*) a pozostałe tabele zwykle się nazywają tabelami wymiarów (ang. *dimensions*). Specjalne znaczenie ma tabela wymiaru czasu (tabela *czas*), która pozwala na wykonywanie agregacji danych podług praktycznie dowolnych kryteriów czasowych.



Rys. 2. Kostka OLAP w strukturze gwiazdy

Oba modele mają swoje zalety i wady. Więcej informacji na ten temat można znaleźć np. w [4]. W SQL Server 2008 domyślnym modelem składowania jest MOLAP. Dodatkowo dostępny jest tam też model HOLAP (ang. *Hybrid OLAP*), który jest czymś pośrednim pomiędzy modelami podstawowymi.

3. PRZYKŁAD WYKORZYSTANIA

Jako przykład praktycznego wykorzystania omawianego wyżej rozwiązania klasy BI pokazano fragment pracy wykonanej przez pierwszego autora artykułu w ramach stażu, jaki odbył on w firmie LUMEL S.A. w Zielonej Górze³. Jednym ze zrealizowanych zadań

² Inne nazwy tych operacji to drażnienie w górę, drażnienie dół (ang. *drill-up*, *drill-down*).

³ Tytuł projektu: "Zróbmy coś razem" - podnoszenie kwalifikacji pracowników.

Projekt współpracy WEIT, UZ i firm woj. lubuskiego (akronim ZCR).

Numer umowy: UDA-POKL.08.02.01-08-011/08-00

Program Operacyjny Kapitał Ludzki, Priorytet: VIII -Regionalne kadry gospodarki, Działanie: 8.2 - Transfer wiedzy, Poddziałanie: 8.2.1 - Wsparcie dla współpracy sfery nauki, Beneficjent: Uniwersytet Zielonogórski, ul. Licealna 9, 65-417 Zielona Góra

było rozpoznanie możliwości wykonania hurtowni danych bazujących na działającym w firmie systemie klasy ERP (ang. *Enterprise Resource Planning*) oraz wykonanie systemu demonstrującego możliwości jakie niosą ze sobą systemy klasy BI⁴.

Wspomniany system ERP to rozwiązanie o nazwie MFG/PRO (wersji 9) firmy QAD. Jest on zbudowany w oparciu o relacyjną bazę danych Progress w wersji 8 i zainstalowany na serwerze IBM z systemem operacyjnym AIX. Wersja bazy danych oraz systemu MFG/PRO są dość stare, gdyż powstały w latach 90. ubiegłego wieku. Mimo to system ten spełnia względnie dobrze swoje zadania w obszarze planowania i prowadzenia produkcji oraz związanych z tym zagadnień finansowych. Nie obejmuje on swoim zasięgiem obszarów kadrowo-płacowych, gdyż są one obsługiwane przez inny system informatyczny istniejący w firmie. Struktura relacyjna systemu MFG/PRO zaprojektowana jest bardzo spójnie i konsekwentnie. Składa się na nią ponad 650 tabel relacyjnych w bazie podstawowej oraz około 270 w bazie dodatkowej. Baza dodatkowa pełni rolę elementu lokalizującego system, dopasowując go do obowiązujących w Polsce przepisów prawnych.

W trakcie tworzenia przykładowego rozwiązania BI napotkano na szereg trudności natury technicznej, głównie spowodowanej tym, że baza Progress oraz wspomniany system ERP są to już stosunkowo stare rozwiązania. System ERP projektowany był w czasach, gdy takie pojęcia jak hurtownie danych czy BI nie były jeszcze znane.

Przykładowo dużym utrudnieniem był brak możliwości dostępu do danych za pomocą jakiegos w miarę współczesnego interfejsu, jak choćby ODBC. Użytkownicy końcowi pracują z systemem w dość archaicznym trybie terminalowym, uruchamiając aplikację kliencką bezpośrednio na serwerze AIX, łącząc się z nim za pomocą protokołu Telnet. Przykładowy ekran roboczy pokazano na rysunku 3. Jedyną sensowną możliwością łatwego pobierania danych do przyszłej hurtowni danych okazało się więc wykorzystanie oprogramowania klienckiego dla bazy Progress ukrytego w pewnym sensie w jednej z aplikacji klienckich (rodzaj systemu raportującego) dla MFG/PRO. „Odkrycie” to otworzyło drogę do skomunikowania MFG/PRO z bazą danych SQL Server 2008, co w efekcie pozwoliło zrealizować założone zadanie projektowe.

ppptmt.p a*		1.4.1 Kartoteka indeksów (główna)		07/09/09
Nr indeksu: 1301-130-004		Opis: KOŁO ZĘBATE KR1 1018		
JM: SI				
Dane indeksy				
Kategoria: M013	Typ indk: H	Rysunek:		
Wprow: 08/04/97	Status:	Wer:		
Grupa projek:	Grupa:	Mag. rysunku:		
Grupa promocij:	Promoc.:	Przerwa ceny:		
Dane zapas indeksu				
Klasa ABC:	Sr. okr.: 90			
L/S-Part/Nrser:	Cykl inwent: 120			
Zakł.: ZA	Okres ważn:			
Magazyn: M05	Wydać z poj. partii: Nie			
Typ magazynu:	Krytyczny: Nie			
Auto nr partii: Nie	Status PZ:			
Grupa part:	Status przyj. ZR:			
Nr artykułu:	Part/Nrser:			
		Aktywn: Nie		
		Aktywn: Nie		
F1-Start 2-Pom 3-Wst 4-Kon 5-Usuń 6-Menu 7-Wygr 8-Czy				

Rys. 3. Przykładowy ekran z aplikacji MFG/PRO

Kolejnym problemem, na jaki napotkano były dość ograniczone możliwości bazy Progress, jeżeli chodzi

o eksport danych. W ostateczności udało się jednak stworzyć taki mechanizm, który umożliwiał praktycznie automatyczny eksport wybranych danych z bazy MFG/PRO do plików tekstowych w formie akceptowanej przez bazę SQL Server 2008. Wymagało to napisania szeregu skryptów w języku Progress 4GL oraz języku skryptowym bazy SQL Server 2008. W efekcie powstał więc prosty system, który można zaliczyć do wspomnianej już we wprowadzeniu klasy procesów ETL. Wykonuje on automatycznie następujące czynności:

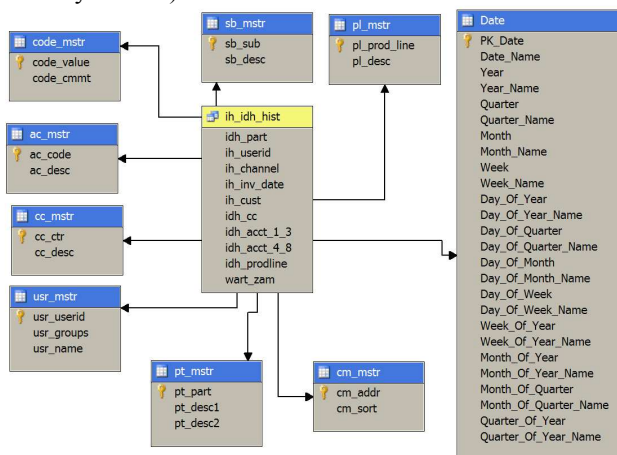
- eksport wymaganych danych z systemu MFG/PRO do plików tekstowych,
- import danych do docelowej hurtowni zlokalizowanej na SQL Server 2008,
- dogenerowanie ewentualnych brakujących danych (mowa jest o tym poniżej) oraz zaimportowanie ich do docelowej hurtowni danych,
- aktywowanie wymaganych kluczy obcych między tabelami (aby przyspieszyć import danych są one wcześniej deaktywowane),
- odświeżanie zawartości analitycznej bazy MOLAP zaimplementowanej w module SSAS.

Kolejnym poważnym problemem była słabo dostosowana do potrzeb hurtowni danych struktura relacyjna bazy MFG/PRO. W wielu miejscach można odczuć brak tabel o charakterze katalogowym, bez których trudno jest sensownie grupować dane. Przykładowo nie ma łatwej metody na pogrupowanie danych dotyczących faktur ze względu na ich rodzaj (np. sprzedaż krajowa, sprzedaż eksportowa, produkcja odlewnicza, usługi krajowe itd.). Zestawienia tego typu oczywiście da się wykonać, jednak trzeba wówczas pisać karkołomne skrypty i/lub zapytania SQL. Gdy pojawi się przykładowo kolejna faktura dotycząca nowej usługi, nie ma możliwości, aby automatycznie pojawiła się ona w istniejących już raportach. A taka funkcjonalność jest kluczowa w systemach klasy BI.

Ostatnim wreszcie z poważniejszych problemów było to, że struktura relacyjna bazy MFG/PRO nie zawiera mechanizmów obsługi kluczy obcych. Wynika to z ograniczeń (dość starej już) bazy danych Progress. Stąd też obsługa integralności referencyjnej jest w systemie MFG/PRO zrealizowana programowo. Brak kluczy obcych nie jest specjalnie odczuwalny dla użytkowników, jednak niesie pewne przykre konsekwencje. W trakcie importowania danych do struktur relacyjnych bazy SQL Server 2008 zauważono, że w niektórych tabelach nadrzędnych (tych z hipotetycznymi kluczami obcymi) brakuje danych, co skutecznie uniemożliwiało założenie kluczy obcych w docelowej hurtowni danych, która takie klucze obsługuje i w zasadzie wymaga ich istnienia. Po rozmowach z pracownikami firmy LUMEL S.A. okazało się, że niektóre dane historyczne są co jakiś czas kasowane, gdyż, jako nieużywane od dłuższego czasu, są zbędne. Zezwalają na to odpowiednie przepisy, choć kłóci się to z dobrą praktyką pracy z bazami relacyjnymi. W efekcie autorzy zdecydowali się na automatyczne „dogenerowywanie” (odtworzenie) brakujących danych. Zadanie to realizują specjalnie w tym celu napisane skrypty po stronie bazy SQL Server 2008.

⁴ Pierwszy autor pracy składa serdeczne podziękowania Panu Markowi Rajszczakowi, kierownikowi działu IT firmy LUMEL S.A., za okazaną wielką życzliwość oraz udzieloną pomoc w czasie trwania stażu.

Aby zademonstrować korzyści, jakie niosą ze sobą systemy klasy BI utworzono stosunkowo prostą hurtownię danych zawierającą bieżące oraz historyczne dane na temat zarejestrowanych faktur. Powstała prosta struktura w schemacie gwiazdy, którą pokazano na rysunku 4. Tabela faktów *ih_idh_hist* zawiera dane o wszystkich fakturach, jakie zostały zarejestrowane w systemie MFG/PRO od początku jego działania w firmie. Pozostałe tabele są tabelami wymiarów zawierającymi takie informacje, jak np. kody z księgi głównej, tzw. główną kartotekę indeksów, kartotekę użytkowników i inne. Będą one używane do agregowania danych podług różnych kryteriów. Nazwy tych tabel są takie same, jak odpowiednie tabele w systemie MFG/PRO. Podkreślimy w tym miejscu, że hurtownia, której model pokazano na rysunku 4 zawiera dane skopiowane ze źródłowego systemu MFG/PRO i dane te są zeskalowane w bazie SQL Server 2008, w modelu składowania MOLAP obsługiwanym przez moduł SSAS. W module SSAS wygenerowano również tabelę wymiarów *Date*, która pełni ważną rolę w procesie agregowania danych w ujęciu czasowym (jest ona odpowiednikiem tabeli *czas* z rysunku 2).



Rys. 4. Kostka OLAP dla fragmentu systemu MFG/PRO

W dalszej kolejności przygotowano szablon przyszłego systemu pokazujący, jak można w praktyce korzystać z danych zapisanych w utworzonej hurtowni danych. Wykorzystując możliwości jakie daje pakiet Office 2007 połączono się z poziomu programu Excel z utworzoną hurtownią danych (dokładniej mówiąc połączono się z analityczną kostką OLAP zeskalowaną w module SSAS) i następnie wykonano praktycznie gotową do użycia przez końcowego użytkownika aplikację, mogącą pełnić rolę prostego ale w pełni funkcjonalnego narzędzia do „biznesowej analizy danych” dotyczących analizy faktur. Przykładowy zrzut fragmentu ekranu pokazano na rysunku 5⁵. Jest to znana już we wcześniejszych wersjach pakietu Excel tzw. tabela przestawna (ang. *Pivot Table*). Tym co jest tutaj nowe i godne podkreślenia to to, że Excel 2007 potrafi połączyć się bezpośrednio z kostką analityczną OLAP zlokalizowaną na serwerze analitycznym SSAS, pobrać z niej dane oraz odświeżyć jej zawartość, gdy dane źródłowe zostaną zmienione. Agregowanie danych w wierszach odbywa się podług czasowej tabeli

wymiarów a agregowanie w kolumnach podług dwóch przykładowych tabeli wymiarów (w ogólności może być ich więcej).

	A	B	C	D	E	F	G	H
1	Fakture LUMEL S.A.							
2	SPE	SPE Suma	SPK			SPK Suma	Suma	
3	1004		1002	1004	1005			
4	Calendar 2007	798532	798532		4924463	165674	5090137	5888670
5	Calendar 2008	802586	802586	1916	4837910	115452	4955278	5757864
6	Calendar 2009	621435	621435		3329959	11659	3341618	3963053
7	Quarter 1, 2009	154873	154873		791124	6659	797783	952656
8	January 2009	14996	14996		243301	968	244269	259265
9	February 2009	51113	51113		305284	1719	307003	358116
10	March 2009	88764	88764		242539	3972	246511	335275
11	Quarter 2, 2009	298141	298141		880406	1191	881598	1179739
12	Quarter 3, 2009	130032	130032		794932	288	795220	925252
13	Quarter 4, 2009	38388	38388		863496	3521	867017	905405
14	Suma	2222553	2222553	1916	13092333	292784	13387033	15609586
15	LEGENDA							
16	1002	Analizatory medyczne	SPE	Kanał sprzedaży eksportowej				
17	1004	Mier. tablicowe analogowe	SPK	Kanał sprzedaży krajowej				
18	1005	Mier. samochodowe						

Rys. 5. Tabela przestawna w Excel 2007 utworzona na bazie kostki OLAP widocznej na rysunku 4

4. PODSUMOWANIE

W pracy w wielkim skrócie przedstawiono możliwości, jakie daje SQL Server 2008 firmy Microsoft w obszarze tworzenia systemów klasy BI. Wykorzystano je w firmie LUMEL S.A. do wykonania pilotażowego prostego systemu demonstrującego jego najważniejsze cechy. Autorzy artykułu liczą, że wykonana praca będzie miała „ciąg dalszy” i uda się stworzyć finalny system BI, który da wymierne korzyści dla firmy LUMEL S.A., zakończy się wdrożeniem i będzie pozytywnym przykładem współpracy przemysłu i nauki.

LITERATURA

- [1] Mendrala D., Szeląga M., *Serwer SQL 2008. Usługi biznesowe*, Wydawnictwo Helion, 2009.
- [2] Królikowski Z., *Hurtownie danych. Logiczne i fizyczne struktury danych*, Wydawnictwo Politechniki Poznańskiej, Poznań, 2007.
- [3] www.microsoft.com/sqlserver/2008/en/us/editions-compare.aspx
- [4] www.sql-server-performance.com/articles/biz/ssas_storage_modes_p1.aspx
- [5] www.ploug.org.pl/szkola/szkola_4/materialy/02_MW_logiczne.pdf



dr inż. Artur Gramacki
Uniwersytet Zielonogórski
Wydział Elektrotechniki, Informatyki i Telekomunikacji
Instytut Informatyki i Elektroniki

ul. Podgórna 50
65-246 Zielona Góra
e-mail: a.gramacki@iie.uz.zgora.pl



dr inż. Jarosław Gramacki
Uniwersytet Zielonogórski
Wydział Elektrotechniki, Informatyki i Telekomunikacji
Instytut Informatyki i Elektroniki

ul. Podgórna 50
65-246 Zielona Góra
e-mail: a.gramacki@iie.uz.zgora.pl

⁵ Oryginalne dane o sprzedaży (stanowiące poufne informacje firmy LUMEL S.A.) zostały zamienione na losowe wartości.