

ZENON BINIEK

Wyższa Szkoła Technologii Informatycznych
Warszawa

KONCEPTUALNE MODELOWANIE STRUKTUR BAZ MOLAP DLA E-BIZNESU

1. Uwarunkowania e-biznesu

Możliwość prowadzenia biznesu w przestrzeni wirtualnej spowodowała liczne zmiany w samym sposobie działania firm. W praktyce występują rozliczne uwarunkowania wpływające na prowadzenie biznesu za pomocą globalnej sieci elektronicznej. Autor wyróżnia tutaj uwarunkowania ekonomiczno-informacyjne¹ i uwarunkowania organizacyjno-techniczne². Rozwój e-biznesu spowodował dodatkowe zapotrzebowanie na nowoczesne rozwiązania informatyczne umożliwiające obsługę wirtualnych firm działających w warunkach gospodarki sieciowej. Powstał nowy typ architektury systemowej (3-warstwowa), nowe technologie interoperacyjności systemowej (SOA) i nowy typ aplikacji internetowych (RIA). Występuje także zapotrzebowanie na nowe rozwiązania w zakresie baz danych. Technologia klasycznych SQL-owych bazy danych ulega modyfikacji, tak aby sprościć wymaganiom zaawansowanych użytkowników sieci Internet. W zakresie informatycznej obsługi informacji zarządczej stoimy przed wyzwaniem stworzenia nowoczesnych narzędzi obsługi wspomagania budowy kokpitów menedżerskich. Pojawiły się nowe rodzaje baz do obsługi wielowymiarowej analizy danych, tzw. bazy MOLAP (*Multidimensional On-line Analytical Processing*).

Jak widzimy, modernizacja klasycznych SQL-owych baz danych przebiega w dwóch kierunkach. Z jednej strony powstają nowe rozwiązania technologiczne

¹ Z. Biniek: *Ekonomiczno-informacyjne uwarunkowania e-biznesu*, „Współczesna Ekonomia” 2007, nr 1.

² Z. Biniek: *Organizacyjno-funkcjonalne aspekty e-biznesu*, w: *Nowe trendy i wyzwania w zarządzaniu*, red. E. Weiss, M. Godlewska, Wizja Pres&IT, Warszawa 2007.

w zakresie zastosowania SQL do obsługi hurtowni danych (ROLAP), a z drugiej strony trwają prace nad nową generacją baz danych do obsługi wielowymiarowej analizy danych, zapamiętujących dane w formie kostek OLAP (MOLAP).

2. Projektowanie konceptualne systemów MOLAP

W pierwszej fazie budowy systemów OLAP dominowało podejście ukierunkowane na tworzenie relacyjnych baz danych obsługujących hurtownie danych. Z technicznego punktu widzenia podejście do projektowania hurtowni danych nie różniło się praktycznie od tradycyjnego podejścia do projektowania relacyjnych baz danych. Zadaniem projektanta bazy danych było między innymi zdefiniowanie semantycznego modelu danych zawierającego definicje formatów danych i relacji pomiędzy poszczególnymi elementami danych. Projektanci wyróżnili, jak wiadomo, dwa podstawowe typy modeli semantycznych do obsługi hurtowni danych: płątka śniegu lub wielogwiazdy³. Wszystkie operacje na relacyjnej bazie danych odbywały się przy użyciu języka SQL, początkowo nieprzystosowanego do przetwarzania na dużych zbiorach danych. Dopiero po pewnym czasie dokonano odpowiednich modyfikacji języka SQL, przystosowując go do obsługi hurtowni danych, między innymi w niektórych implementacjach języka opracowano programową obsługę operatora Cube⁴.

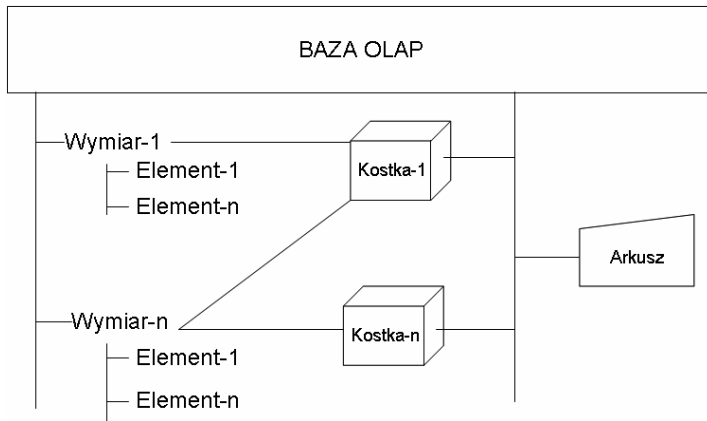
Inne podejście do tworzenia systemów OLAP polega na definiowaniu baz analitycznych obsługujących arkusze kalkulacyjne. W takim przypadku baza OLAP wspomaga wielowymiarową analizę danych. Użytkownik, korzystając jak poprzednio z arkusza kalkulacyjnego (*Spreadsheet*), jest wspomagany bazą OLAP zawierającą dane odpowiednio zagregowane, przystosowane do potrzeb zestawień analitycznych. Baza taka zasadniczo różni się od klasycznej bazy ROLAP, gdzie do modelowania danych można stosować diagramy typu E/R. W bazie OLAP nie stosujemy dwuwymiarowych tablic relacyjnych zawierających kolumny kluczowe i kolumny danych. Na rysunku 1 przedstawiono schemat funkcjonalny bazy OLAP obsługującej arkusze kalkulacyjne.

W modelu konceptualnym bazy MOLAP nie występują tradycyjne schematy relacyjnego modelu danych: gwiazdy i płątka śniegu. W takim przypadku już w trakcie budowy modelu konceptualnego definiujemy tylko te elementy danych, które są niezbędne do zdefiniowania wielowymiarowej kostki OLAP, zakładając,

³ Zob. Z. Biniek, *Informatyka w zarządzaniu*, Vizja Press&IT, Warszawa 2009, s. 287 i n.

⁴ A. Eickler, A. Kemper, *Datenbanksysteme*, Oldenbourg, München 1997, s. 465.

że zdefiniowane kostki będą przeznaczone do obsługi różnorodnych raportów analitycznych. Jak widać na rysunku 1, projektant definiuje wymiary i przynależne im elementy.

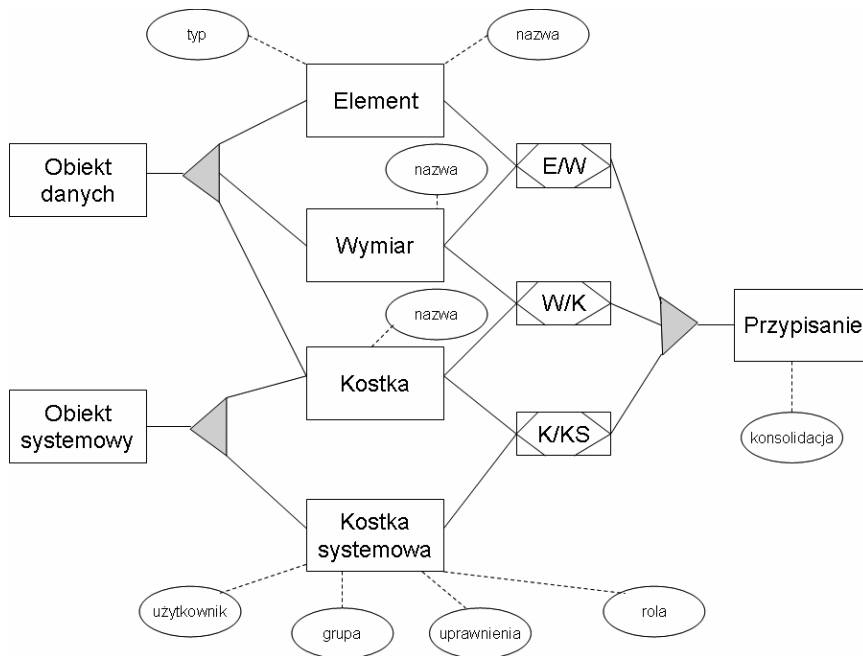


Rys. 1. Struktura funkcjonalna bazy MOLAP

Źródło: opracowanie własne.

Baza MOLAP składa się z jednej lub wielu wielowymiarowych kostek OLAP. W jednej bazie może być wiele kostek (kostka – 1, ..., kostka – n). Kostka OLAP zawiera wcześniej zdefiniowane wymiary. Wymiar służy do opisu tych składników analizy danych, które występują w danym systemie zarządzania, na przykład czas, obszar, produkty. Konkretny wymiar może występować w jednej kostce lub w wielu jednocześnie. W jednej bazie może występować wiele wymiarów w zależności od zakresu analizy danych. Wymiar składa się z elementów danych (element – 1, ..., element – n). Element zawsze jest przypisany do konkretnego wymiaru. Element jest definiowany przez nadanie mu unikatowej nazwy. Podobnie definiowane są wymiar i kostka. Przy definiowaniu elementów najpierw określamy wymiar, a następnie wymiarowi przypisujemy elementy, oznacza to, że elementy nie istnieją samodzielnie w bazie. Kostka OLAP musi zawierać co najmniej jeden wymiar. Arkuszem (*Worksheet*) posługuje się użytkownik bazy danych przy wprowadzaniu danych oraz przygotowywaniu zestawień informacyjnych z bazy danych.

Jednym z istotnych zagadnień projektowania systemów informatycznych dla e-biznesu jest modelowanie struktur baz MOLAP. W modelowaniu koncepcyjnym posługujemy się pojęciem metamodelu. Wprowadził je do literatury przedmiotu E. Sinz⁵, definiując metamodel danych typu E/R, co w znacznym stopniu uporządkowało dyskusję o semantycznym modelowaniu struktur relacyjnych baz danych. Posługiwanie się kategorią metamodelu ułatwia definiowanie konkretnych modeli danych, które zaliczone do konkretnej kategorii muszą spełniać ogólne reguły przewidziane dla danego metamodelu. Także klasyczna notacja E/R odgrywa tutaj ważną rolę⁶. Wychodząc z ogólnych przesłanek metamodelowania, autor podjął zadanie zdefiniowania metamodelu bazy MOLAP, co przedstawiono na rysunku 2. Przedstawiono na nim graficzne odzwierciedlenie



Rys. 2. Metamodel koncepcyjny bazy MOLAP (klasyczny)

Źródło: opracowanie własne.

⁵ O. Ferstl, E. Sinz, *Grundlagen der Wirtschaftsinformatik*, Oldenbourg, München 2001, s. 122.

⁶ M. Goecken, *Entwicklung von Data – Warehouse – Systemen*, DUV, Berlin 2007, s. 191 i n.

zależności między elementami bazy MOLAP stosowane w użytkowanych dotychczas pakietach softwarowych, tak zwane podejście klasyczne. Obiekty danych reprezentują fakty biznesowe zachodzące w działalności gospodarczej. W procesie modelowania konceptualnego fakt zostaje przeniesiony do bazy jako obiekt danych analitycznych. Obiekt danych może w procesie modelowania przyjmować postać elementu, wymiaru lub kostki, przy czym element danych składa się z jednorodnej wartości definiowanej w konkretnych jednostkach miary. Wymiar składa się z elementów, a kostka OLAP z wymiarów. Przypisanie elementów do wymiaru i wymiaru do kostki jest suwerenną decyzją projektanta bazy OLAP. Tym samym ostateczną postać danego obiektu danych określa projektant bazy OLAP, kierując się ogólnymi regułami modelowania zapisanymi w metamodelu.

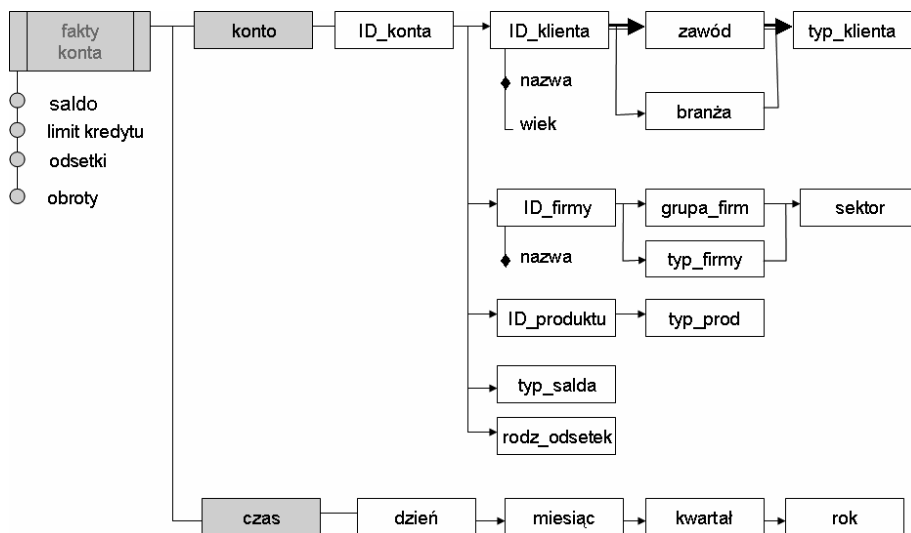
W modelowaniu zależności organizacyjno-funkcjonalnych posługujemy się obiektami systemowymi: Cube, View, Dim, Subset, Funkcja. Są to obiekty istniejące w każdej bazie MOLAP, niezależnie od konkretnych wpisów rzeczowych (element, wymiar, kostka). Ponadto projektant bazy musi zadbać o przypisanie elementów bazy. Obowiązuje ogólna zasada, że w bazie nie może wystąpić element nieprzypisany do wymiaru. Przypisanie jest rozumiane tutaj jako logiczne powiązanie elementów i wymiarów (E/W), logiczne powiązanie wymiarów i kostek (W/K) oraz logiczne powiązanie kostek danych i kostki systemowej (K/KS). W przypisaniu E/W dokonujemy konsolidacji niektórych elementów, a zatem niezbędne jest zdefiniowanie reguł konsolidacji. Konsolidacja elementów polega na powiązaniu hierarchicznym elementów, czyli określeniu przynależności do pewnej zdefiniowanej, nadrzędnej grupy elementów. Każdy element bazy może mieć zdefiniowane atrybuty, wyrażające pewne dodatkowe cechy tego elementu.

Powiązanie między jednostkami logicznymi bazy OLAP następuje przez unikatową nazwę elementu. Oprócz unikatowej nazwy element ma przypisany typ (numeryczny, tekstowy). Do modelowania powiązań między wymiarami i elementami są stosowane specjalne diagramy, tak zwane schematy faktów⁷. Na rysunku 3 przedstawiono przykładowy schemat faktów.

Fakty reprezentują jednostkowe elementy informacji (danych) w wielowymiarowej bazie analitycznej. Fakt opisuje wartości ilościowe w postaci miary (jednostki miary) oraz cechy jakościowe, które są określone w bazie przez różne wymiary, poczynając od poziomu bazowego (*terminal dimension level*). Dla tego samego wymiaru i tego samego poziomu bazowego można zdefiniować

⁷ J. Lechtenboerger, G. Vossen, *Qualitätsorientierter Schemaentwurf für Datenlager*, „Information Technology” 2003, 45, 4, s. 190.

kilka ścieżek agregacji. Elementy danego wymiaru są funkcjonalnie zależne od elementu bazowego. Każdy pojedynczy wymiar zawiera zdefiniowaną grupę instancji lub elementów. Wymiary są ukształtowane hierarchicznie i mają konkretną ścieżkę agregacji. Każdy element bazy musi należeć do co najmniej jednego konkretnego wymiaru.



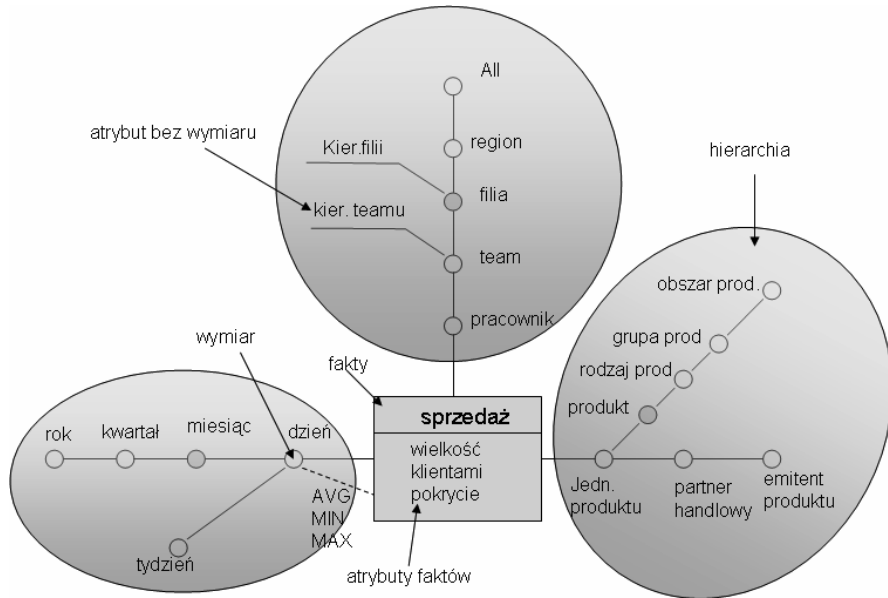
Rys. 3. Schemat wymiarów i faktów bazy OLAP

Źródło: opracowanie własne na podstawie B. Hüsemann i in., *Conceptual Data Warehousing Design*, w: *Proceedings of the International Workshop on Design and Management of Data Warehousing*, Stockholm 2000, s. 6–3.

Innym rodzajem diagramu faktów jest schemat wymiarów i faktów według Golfarelli-Rizzi⁸. W centralnie usytuowanej tabeli faktów zawarte są podstawowe atrybuty opisujące sprzedaż. Wymiary i elementy zostały ukształtowane hierarchicznie. Dodatkowo można umieszczać w diagramie operacje wykonywane na elementach, na przykład AVG, MIN, MAX.

Diagram przedstawiony na rysunku 4 umożliwia zdefiniowanie hierarchicznej struktury elementów występujących w konkretnej bazie MOLAP. Centralna tabela faktów (sprzedaż) została otoczona różnymi niezbędnymi wymiarami ukształtowanymi hierarchicznie bądź liniowo.

⁸ M. Goecken, *op.cit.*, s. 224.

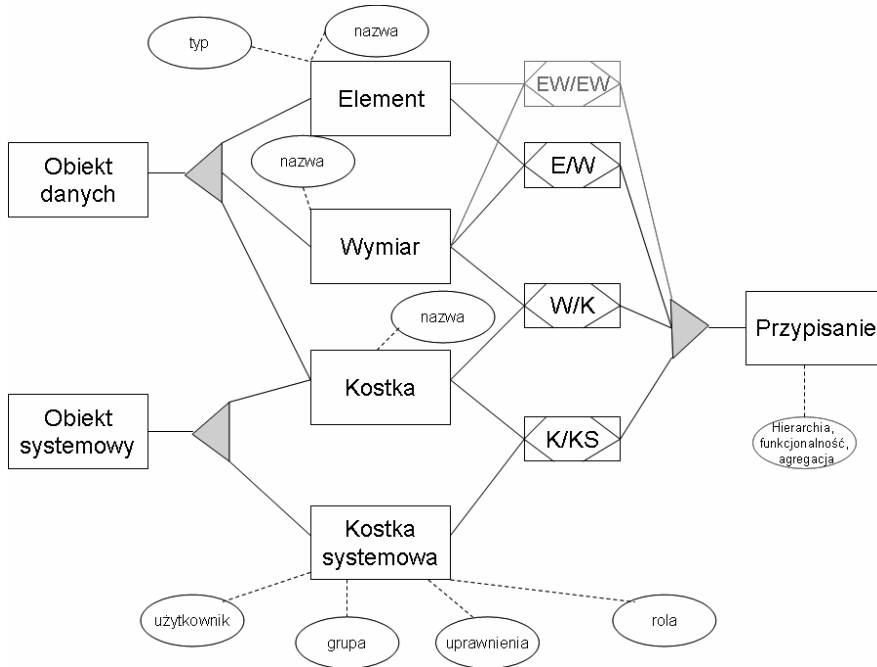


Rys. 4. Schemat faktów według Golfarelli-Rizzi

Źródło: opracowanie własne.

Przedstawiony na rysunku 2 klasyczny metamodel bazy MOLAP jest niewystarczający do pełnej reprezentacji obiektów i relacji w wielowymiarowej analizie danych. Klasyczne pakiety MOLAP mają najczęściej tę wadę, że umożliwiają definiowanie tylko i wyłącznie hierarchii elementów i wymiarów w kostkach, a brak im możliwości definiowania specjalizacji elementów. Dodatkowym wymogiem jest to, że postulat specjalizacji elementów jest aktualny w obrębie wielu wymiarów⁹. W modelowaniu konceptualnym bazy MOLAP należy wyróżnić trzy typy relacji: relacje specjalizacji, agregacji i przynależności – członkostwa (*membership*). W dotychczas użytkowanych systemach dominują relacje agregacji i członkostwa. Relacja specjalizacji w zasadzie nie występuje w obecnie użytkowanych bazach MOLAP. Przypisanie elementowi zależności od innych elementów, jednego lub wielu, wymaga zmian w metamodelu bazy MOLAP. Zmiany te zostały przedstawione na rysunku 5 (jaśniejszym kolorem).

⁹ A. Abello i in., *Understanding Analysis Dimensions in a Multidimensional Object-Oriented Model*, w: *Proceedings of the International Workshop of Design and Management of Data Warehouse*, DMDW 2001, Interlaken 2001.



Rys. 5. Rozszerzony metamodel bazy MOLAP

Źródło: opracowanie własne.

Niezbędne jest, aby pakiety do wielowymiarowej analizy danych pracujące na bazie OLAP obsługiwały wszystkie dające się wyróżnić relacje między elementami, w tym relację funkcjonalności (*functional dependency*)¹⁰. Szarym kolorem zaznaczono przypisanie typu wiele-do-wielu ($n : m$)¹¹ pomiędzy elementami różnych wymiarów w ramach tej samej kostki, na przykład pracownik (element wymiaru pracownicy) obsługuje jeden produkt lub kilka produktów (wymiar produktu). Do trzech tradycyjnie stosowanych powiązań (do obsługi relacji hierarchicznej i członkostwa) należy dodać powiązanie typu EW/EW obsługujące relacje funkcjonalności. Dwa elementom można przypisać relację 1 : 1 lub relację $n : m$, a zbiór powiązań elementów jest zbiorem skończonym par powiązań elementów. Każdy element danego zbioru relacji należy do konkretnego wymia-

¹⁰ G. Vossen, *Datenbankmodelle, Datenbanksprachen und Datenbankmanagement-Systeme*, wyd. 3, Oldenbourg Verlag, München 1999, s. 152.

¹¹ Właściwości typowych dla relacji występujących w relacyjnych bazach danych nie można wprost przenosić na bazy OLAP.

ru. Przypisanie tego typu znaczenia ułatwi posługiwanie się elementami bazy przy sporządzaniu konkretnych raportów wieloprzekrojowych, a zwłaszcza przygotowaniu raportów dla kokpitów menedżerskich.

W trakcie analizy wymagań powstaje model konceptualny wielowymiarowej bazy OLAP. Budowa modelu konceptualnego bazy OLAP przebiega w kilku krokach:

- a) zdefiniowanie miar, zależność funkcjonalna bazy, określenie zależności między wymiarami i miarami;
- b) określenie wymiarów i ich hierarchii;
- c) zdefiniowanie konsolidacji elementów.

Dotychczas omawialiśmy modelowanie w ujęciu funkcjonalno-strukturalnym. W tworzonych diagramach uwaga modelujących skupiała się na definicji obiektów oraz powiązaniach tych wyróżnionych obiektów. Istotne znaczenie mają właściwie zdefiniowane relacje między wymiarami w ramach danej kostki OLAP. Przy sporządzaniu wielowymiarowych analiz danych posługujemy się funkcjami tabelarycznymi umożliwiającymi pobieranie danych z kostek OLAP i umieszczanie tych danych w przygotowywanych raportach¹². Funkcje są obiektami systemowymi wywoływanyymi przez nazwę z przeznaczeniem głównie do manipulowania danymi. Mamy tutaj zarówno funkcje przeznaczone do filtrowania danych, jak i funkcję operowania na tablicach, macierzach. Projektant bazy w fazie tworzenia jej modelu konceptualnego może przewidzieć, jakie funkcje znajdą zastosowanie w konkretnej implementacji.

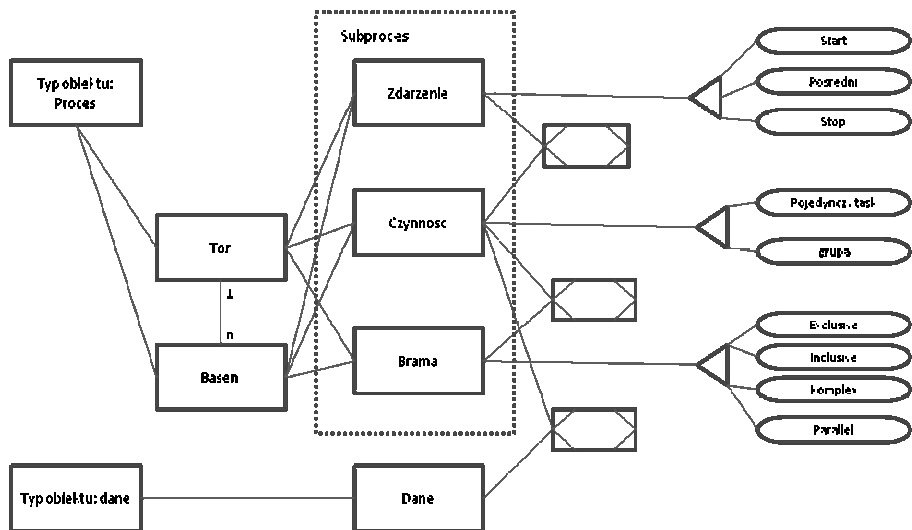
W modelowaniu konceptualnym baz MOLAP wyróżniamy ponadto podejście procesowe. Modelowanie procesów biznesowych, a zwłaszcza procesów w relacji B2B, przebiega według ustalonych i skodyfikowanych reguł modelowania. Założenia metamodelu procesowego w projektowaniu baz OLAP przedstawiono na rysunku 6.

Idea metamodelu procesowego wywodzi się z ogólnych założeń metamodelowania obiektowego zdefiniowanych przez E. Sinza¹³. Obecnie w modelowaniu procesów biznesowych dominuje notacja BPMN (*Business Proces Modeling Notation*)¹⁴. Proces biznesowy jest modelowany przez sekwencję elementów: zdarzenie – czynność – brama. Można tutaj modelować pojedyncze procesy lub

¹² B. Held, H. Erb, *Advanced Controlling mit Excel*, Franzis, Poing 2006.

¹³ K. Ferstl, E. Sinz, *op.cit.*, s. 199.

¹⁴ Innym rodzajem notacji do definiowania procesów biznesowych jest notacja UML, którą tutaj pominiemy dla zachowania przejrzystości wywodu.



Rys. 6. Metamodel procesowy w e-biznesie

Źródło: opracowanie własne.

subprocesy. Do modelowania procesów B2B konieczne jest uporządkowanie elementów w ramach jednego toru (*lane*) lub wielotorowo (*pool*). W ramach ścieżki *pool* można na przykład wyróżnić i opisać graficznie zadania serwera aplikacji w trzywarstwowej architekturze systemów informatycznych. Poszczególne czynności lub subprocessy mogą przebiegać wielotorowo, czyli równolegle, w ramach określonych reguł dokonywania zmian w postępowaniu uczestników procesów biznesowych. W modelowaniu procesów można wyróżnić kilka sposobów przypisania elementów metamodelu. Z jednej strony mamy przypisanie zdarzeń i czynności oraz czynności i bram. Nie występują przypisanie typu brama – brama lub czynność – czynność, a także zdarzenie – zdarzenie. Z drugiej strony można stosować przypisanie elementów danych do poszczególnych czynności. W poszczególnych notacjach procesów biznesowych znajdujemy różne typy zdarzeń, czynności i bram.

W podsumowaniu należy stwierdzić, że stosowanie metody metamodelowania konceptualnego struktur danych pozwala na uzyskanie różnych korzyści w projektowaniu baz danych obsługujących procesy biznesowe. Tworząc uniwersalny model na potrzeby definiowania konkretnych modeli implementacyjnych, doprowadzamy do uporządkowania semantycznego struktur danych, zapewnia-

my semantyczną kompletność konkretnych notacji, odpowiednią dokładność semantyczną, a także możliwość uzupełniania notacji o nowe elementy dostępne w danym metamodelu. Metamodelowanie w znacznym stopniu ułatwia tworzenie nowych implementacji softwarowych. Na podstawie definicji klas w metamodelu możemy tworzyć odpowiednie biblioteki modeli, z których pobieramy elementy do tworzenia konkretnych implementacji procesów biznesowych.

Literatura

- Abello A. i in., *Understandig Analysis Dimensions in a Multidimensional Object-Oriented Model*, w: *Proceedings of the International Workshop of Design and Management of Data Warehouses*, DMDW 2001, Interlaken 2001.
- Biniek Z., *Informatyka w zarządzaniu*, Vizja Press&IT, Warszawa 2009.
- Business Process Modeling Notation*, Working Draft 1.0 [BPMI.org], 2003.
- Eickler A., Kemper A., *Datenbanksysteme*, Oldenbourg Verlag, München 1997.
- Erb H., Held B., *Advanced Controlling mi Excel*, Franzis, Poing 2006.
- Ferstl K., Sinz E., *Grundlagen der Wirtschaftsinformatik*, t. 1, Oldenburg Verlag, München 2001.
- Hüsemann B., Lechtenböberger J., Vossen G., *Conceptual Data Warehouse Design*, w: *Proceedings of the International Workshop on Design and Management of Data Warehouses*, Stockholm 2000.
- Lechtenboerger J., Vossen G., *Qualitätsorientierter Schemaentwurf für Datenlager*, „Information Technology” 2003, 45, 4.
- Vossen G., *Datenbankmodelle, Datenbanksprachen und Datenbankmanagement-Systeme*, wyd. 3, Oldenbourg Verlag, München 1999.

CONCEPTUAL MODELING OF MOLAP DATABASES' STRUCTURES FOR E-BUSINESS

Summary

In this paper we address the issue of conceptual modeling of data used in multi-dimensional analysis. We view this problem from database developers point of view. We describe conceptual and logical models used in e-bussines areas. Meta Model of MOLAP e-bussines Systems will be describes and discussed. We distinguish between Meta Data Model and Meta Process Model .It will be made a critical analyses of meta data model for data base modeling.

Translated by Zenon Biniek

