

Marcin GORAWSKI, Dariusz MROZEK  
Politechnika Śląska, Instytut Informatyki

## ADMINISTROWANIE SYSTEMAMI HURTOWNI DANYCH

**Streszczenie.** W artykule przedstawiono istotne aspekty procesu administrowania systemami hurtowni danych. Dokonano analizy i porównania istniejących modeli administrowania oraz proponowanego przez nas modelu kooperacyjnego. Szczególną uwagę poświęcono roli metadanych oraz sposobom ich przechowywania i wymiany w omawianych modelach. Przedstawiono również, bazującą na proponowanym przez autorów modelu administrowania, aplikację umożliwiającą scentralizowane zarządzanie środowiskiem hurtowni danych.

**Słowa kluczowe:** hurtownia danych, administracja, metadane, repozytorium

## ADMINISTERING IN THE DATA WAREHOUSE SYSTEMS

**Summary.** In the article, we would like to introduce crucial aspects of the management process in the data warehouse environment. We investigate existing management models and the cooperative management model we created. Special place we dedicate to role of metadata concept, the way we can store it and exchange in mentioned models. Additionally, we would like to introduce management console application that is based on the cooperative management model.

**Keywords:** data warehouse, administering, metadata, repository

### 1. Wstęp

Administrowanie systemami hurtowni danych jest zagadnieniem ważnym dla cyklu rozwoju systemów i eksploatacji systemów wspomagania decyzji (DSS – ang. *Decision Support System*). Przetwarzanie informacji w systemach DSS pozwala na wykorzystanie informacji zgromadzonej w hurtowni do wykrywania trendów zachowań biznesowych korporacji [3]. Umożliwia to z kolei efektywne podejmowanie decyzji strategicznych lub krótkotermino-

wych w odniesieniu do wielu aspektów, tj.: sprzedaż, marketing, ocena ryzyka, obsługa klienta i in.

Jeżeli przedsiębiorstwo decyduje się na posiadanie, obok istniejącego systemu transakcyjnego (OLTP – ang. *On Line Transaction Processing*), również system DSS, to jego zarządca musi mieć świadomość faktu, że ważność procesu administrowania (zarządzania) takim systemem wzrasta proporcjonalnie do poziomu ważności generowanej informacji.

Z punktu widzenia technicznego administrowanie hurtownią danych jest z jednej strony podobne do, a z drugiej strony różni się od, zarządzania systemem transakcyjnym. Podobieństwa można upatrywać w tym, że dane hurtowni są gromadzone i zarządzane często przez ten sam typ systemu zarządzania bazą danych DBMS (choć nie musi to być regułą). W tym przypadku wydaje się, że zadania administratora powinny mieć podobny charakter, a narzędzia wykorzystywane do administracji dokładnie te same. Z drugiej jednak strony różnica w administrowaniu wynika z charakterystyki danych, z którymi przychodzi nam pracować – w przypadku systemów OLTP dane mają charakter zmienny i są na bieżąco uaktualniane, natomiast w przypadku hurtowni danych są one trwałe i mają charakter historyczny. Różna może być też tzw. granulacja lub ziarnistość danych. Różnice w procesie administrowania pogłębiają się dodatkowo, jeśli proces administracji hurtownią danych poszerzymy o zarządzanie architekturą OLAP (ang. *On Line Analytical Processing*) tworzącą środowisko DSS. W takim przypadku dochodzą nowe elementy związane z przetwarzaniem OLAP czy też różnego rodzaju odwzorowaniami, schematami raportów, filtrami bezpieczeństwa i in.

Aspekt techniczny administrowania systemami dotyczy w szczególności zagadnienia metadanych. Metadane (ang. *metadata*), często określane jako „dane o danych”, to pewien rodzaj informacji na temat tego, jakie dane znajdują się w naszym systemie i jaka jest ich struktura. Z punktu widzenia praktycznego użycia jest to każdy rodzaj informacji, który przyczynia się do minimalizacji wysiłków i kosztu administrowania systemami hurtowni danych.

W artykule tym zaprezentowano możliwości, jakie daje zbudowany przez autorów podsystem zarządzania hurtowniami danych. Jego uniwersalizm pozwala na współpracę i pozyskiwanie metadanych z różnych systemów DBMS oraz komercyjnych narzędzi klasy OLAP. W dalszej części niniejszego opracowania przedstawiono zadania, jakie składają się na proces administrowania hurtowniami danych i środowiskami DSS (rozdz. 2) oraz modele administracji (rozdz. 4). W rozdz. 3 sklasyfikowano metadane ze względu na ich poziom opisu oraz określono źródła pochodzenia metadanych. W rozdz. 5 przedstawiono zbudowany przez autorów uniwersalny podsystem zarządzania hurtowniami danych. W rozdz. 6 omówiono perspektywy rozwojowe podsystemu.

## 2. Proces administrowania i zadania administratora

Głównymi czynnikami decydującymi o tym, czy informacja otrzymywana w procesie jej przetwarzania OLAP w systemie DSS jest przydatna, czy nie, są: szybkość, z jaką możemy otrzymać żadaną informację oraz jakość samej informacji. Administrator w procesie zarządzania poszczególnymi elementami tego systemu może mieć wpływ zarówno na pierwszy z wymienionych czynników, jak i na drugi – jeśli tylko jest członkiem zespołu projektującego. Administrator powinien mieć pełną świadomość logicznych powiązań pomiędzy poszczególnymi elementami środowiska i być koordynatorem współtworzącym te elementy. Niejednokrotnie bierze czynny udział w procesie projektowania architektury środowiska i przepływu danych, zamiast np. wykonywać tylko wyspecyfikowane zadania. Oto niektóre zadania realizowane przez administratora systemów hurtowni danych/OLAP/DSS:

- wdrożenie odpowiedniej strategii indeksowania i partycjonowania bazy hurtowni,
- tworzenie kopii zapasowych hurtowni,
- strojenie hurtowni danych,
- koordynacja lub sterowanie procesem ekstrakcji danych do hurtowni,
- tworzenie strategii aktualizacji danych w hurtowni i automatyzacja tego procesu,
- wyodrębnianie nowych obszarów pozyskiwania danych,
- kontrola jakości danych,
- prowadzenie audytu rekordów co do czasu wprowadzenia i miejsca pochodzenia,
- przeglądanie dzienników systemowych, dzienników aplikacji i zadań,
- monitorowanie sukcesów i porażek procesu ładowania,
- rozwiązywanie konfliktów danych w procesie ETL oraz korygowanie błędów,
- zarządzanie elementami struktury wielowymiarowej, m.in.: wymiarami, hierarchiami, atrybutami wymiarów, miarami faktów,
- zarządzanie relacjami pomiędzy elementami struktury wielowymiarowej,
- zarządzanie schematami, układem i księgą raportów dla decydentów,
- tworzenie szablonów raportów,
- zapewnienie odpowiedniej polityki bezpieczeństwa na poszczególnych etapach tworzenia i korzystania z systemu.

To tylko niektóre, wybrane zadania administratora. Zakres odpowiedzialności administratora też nie jest jednoznacznie sprecyzowany - wykonanie niektórych z wyżej wymienionych zadań może zostać oddelegowane do innych projektantów. Jednakże administrator systemu powinien mieć kontrolę nad tymi zadaniami i móc interweniować w wielu nieprzewidzianych sytuacjach.

### 3. Rola metadanych w procesie administracji HD

Środowisko hurtowni danych podlega ciągłym zmianom – następuje przepływ nowych danych w systemie, tworzone są nowe elementy lub modyfikowane istniejące, środowisko zostaje poszerzone tematycznie, powstają nowe obszary pozyskiwania danych i nowe kanały dystrybucji. Twórcy mówią o tzw. cyklu życia hurtowni (ang. *Data Warehouse Lifecycle*) [1]. Ciągłe zmiany generują nowe informacje i nowe metadane. Metadane powinny być grupowane i kolekcjonowane w specjalnych repozytoriach, nad którymi kontrolę powinien również sprawować administrator środowiska.

Ze względu na poziom opisu metadane możemy podzielić na: techniczne i semantyczne (biznesowe). Metadane techniczne (zwane również administracyjnymi lub strukturalnymi) gromadzą informacje o architekturze i schemacie m.in. systemu transakcyjnego, hurtowni danych i motoru OLAP. Obejmują one np. informacje o tabelach i strukturze poszczególnych wierszy tabel, powiązania pomiędzy atrybutami, wyzwalacze, widoki zdefiniowane w różnych bazach danych i różnych systemach plików, zależności i odwzorowania pomiędzy źródłami transakcyjnymi, hurtownią danych na poziomie fizycznym i implementacyjnym. Zawierają informacje o wszystkich filtrach ekstrakcji, transformacji i agregacji danych pochodzących np. z plików i tabel fizycznych bazy danych. Metadane semantyczne przeznaczone są do biznesowego opisu zawartości hurtowni danych. Ten rodzaj informacji może dotyczyć: pojęciowego modelu przedsiębiorstwa (opisującego model danych i strukturę przedsiębiorstwa na wysokim poziomie), wielowymiarowego modelu danych (m.in. wymiary, hierarchie, atrybuty) oraz wszelkiego rodzaju zależności pomiędzy wymienionymi modelami a fizycznymi danymi [4].

Źródłem metadanych mogą być różnego rodzaju narzędzia działające w środowisku. Najczęściej operują one na pewnej wybranej grupie metadanych lub tworzą ten rodzaj informacji w odniesieniu do określonej dziedziny problemu (np. metadane procesu ekstrakcji). Narzędzia te wykorzystują własne metody przechowywania wygenerowanych metadanych. Do narzędzi, które operują lub tworzą metadane w środowisku hurtowni danych, możemy zaliczyć: narzędzia modelowania danych i narzędzia typu CASE (np. SYBASE POWER DESIGNER, PLATINUM ERWIN, RATIONAL ROSE, ENTERPRISE ARCHITECT), narzędzia ekstrakcji, transformacji i przenoszenia danych z systemów transakcyjnych do hurtowni danych (np. ORACLE WAREHOUSE BUILDER, MICROSOFT DATA TRANSFORMATION SERVICES, PLATINUM DECISION BASE), narzędzia OLAP, raportujące i narzędzia *business intelligence* (np. MICROSTRATEGY DESKTOP, COGNOS IMPROMPTU, BUSINESS OBJECTS), repozytoria metadanych dla hurtowni danych lub ogólnego przeznaczenia, systemy zarządzania bazą danych DBMS i grupy narzędzi administracyjnych dla tych systemów [4].

Kontrola i zarządzanie katalogiem metadanych powinny również należeć do zadań administratora, powinna istnieć także możliwość wymiany metadanych między różnymi narzędziami. Cały proces administracji w gruncie rzeczy oparty jest na metadanych – niektórzy twierdzą, że administracja w środowisku hurtowni danych to właściwie zarządzanie metadanymi [1] i trudno się z tym nie zgodzić. Z tego powodu, tak ważnym zagadnieniem jest przyjęcie odpowiedniego modelu administrowania w zależności od elementów składających się na architekturę systemu.

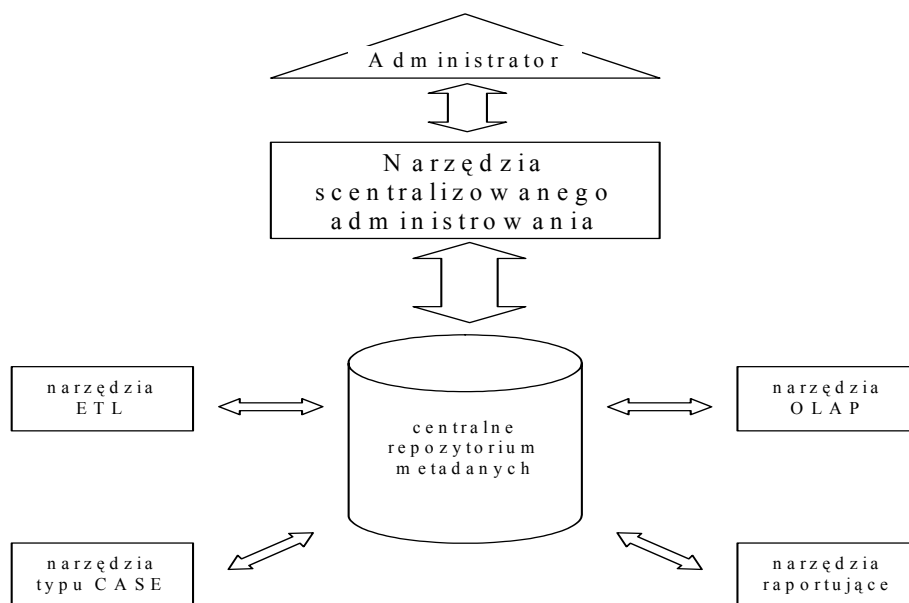
## **4. Modele administrowania systemami DSS**

Model administrowania systemami DSS determinuje rodzaj zadań, z jakimi przyjdzie zmierzyć się administratorowi systemu, rodzaj narzędzi, jakimi będzie musiał posługiwać się w tym procesie, oraz politykę rozwoju architektury. Poniżej przedstawiono administrowanie środowiskiem DSS w przedsiębiorstwie.

### **4.1. Model zintegrowany**

Model zintegrowany (Rys. 1), charakteryzuje się występowaniem centralnego repozytorium jako składnicy metadanych pochodzących z różnych źródeł. Producentem metadanych do repozytorium mogą być różnego rodzaju narzędzia środowiska DSS, poczynając od narzędzi typu CASE, systemy zarządzania bazą danych DBMS bazy źródłowej i bazy hurtowni danych, narzędzia ETL, OLAP i narzędzia raportujące, kończąc na innych repozytoriach pełniących w tym wypadku rolę podrzędną w stosunku centralnego repozytorium. Wszystkie narzędzia korzystają z tego samego zestawu metadanych w trakcie swojej pracy, a wszelkie zmiany są na bieżąco uaktualniane. W modelu zintegrowanym administrator systemu posługuje się jednym narzędziem scentralizowanego zarządzania lub pakietem takich narzędzi – jest to najczęściej tzw. konsola administracyjna.

Niewątpliwą zaletą takiego modelu jest koncentracja metadanych w jednym miejscu, dzięki czemu administrator ma globalną kontrolę nad wszystkimi elementami systemu z poziomu jednego narzędzia i może mieć pewność, że to, co właśnie widzi na ekranie, jest tym, co faktycznie znajduje się w systemie. Dodatkowym atutem jest możliwość tworzenia nowych elementów środowiska z poziomu konsoli administracyjnej (czyli wprowadzania nowych metadanych do repozytorium) – mamy tu szczególnie na myśli takie elementy, które niekoniecznie uda się stworzyć za pomocą wspomnianych narzędzi działających w środowisku. Mogą to być np. różnego rodzaju komentarze, specyfikacje, nowe hierarchie atrybutów i in. [1].



Rys. 1. Model zintegrowany administracji środowiskiem DSS

Fig. 1. Administering in DSS environment in the integrated model

Niestety, w praktyce model tego typu sprawdza się jedynie w środowiskach, w których stosujemy narzędzia jednego producenta, czyli tzw. zintegrowane rozwiązania dla OLAP. Powodem jest kompatybilność poszczególnych narzędzi z centralnym repozytorium, a to gwarantują głównie całe pakiety produktów zintegrowanych. Generuje to jednak kolejny problem – nie zawsze rozwiązania zintegrowane, pochodzące od jednego producenta gwarantują oczekiwaną funkcjonalność. Często zdarza się, że korzystamy z narzędzi kilku producentów ze względu na ich lepszą funkcjonalność. Problem ten już dawno zaczęli dostrzegać producenci i dostawcy narzędzi klasy OLAP. W wyniku podejmowanych wysiłków powstało kilka różnych standardów, które miały umożliwić wymianę metadanych między narzędziami. Wiąże się z tym kolejny model administracji, który przedstawiono poniżej.

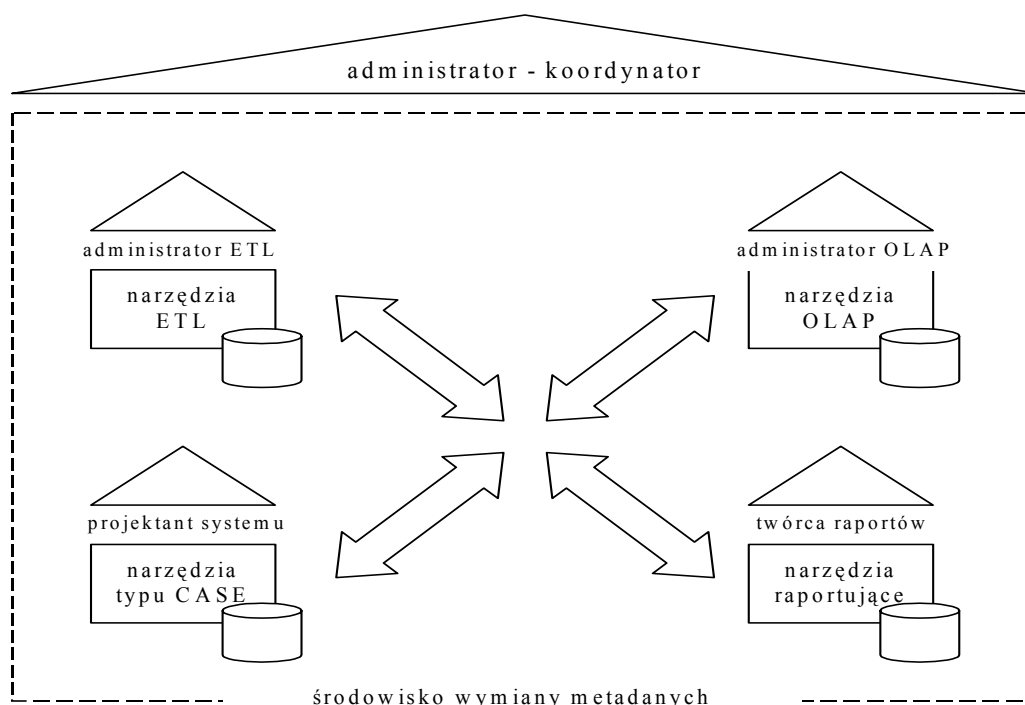
#### 4.2. Model zdecentralizowany

W odróżnieniu od modelu zintegrowanego, model zdecentralizowany (Rys. 2) cechuje się brakiem centralnego repozytorium metadanych i występowaniem tzw. repozytoriów lokalnych, które należy rozumieć jako lokalne zbiory metadanych często o różnej postaci. Repozytoria lokalne są związane z narzędziami DSS wchodzącymi w skład architektury (z tymi narzędziami, które produkują metadane). W takim modelu administracja ma charakter rozproszony - można tu wyróżnić dwa podejścia:

Pierwsze zakłada wzajemną wymianę metadanych pomiędzy lokalnymi repozytoriami, kiedy tylko jest to konieczne (np. narzędzia ETL potrzebują informacji na temat struktury bazy źródłowej i docelowej). Przy takim podejściu lokalne repozytoria mogą mieć różną budo-

wę, metadane mogą być przechowywane w różnej postaci, natomiast istotne jest, aby istniał wspólny format wymiany tych metadanych rozumiany przez każdą ze stron linii wymiany.

Drugie rozwiązanie, w zamian za brak centralnego repozytorium, przewiduje istnienie globalnego widoku metadanych – pewnej logicznej struktury centralnej, wewnątrz której metadane są różnie rozlokowane fizycznie. W tym podejściu wszystkie lokalne repozytoria powinny mieć taką samą strukturę (tzw. metamodel) oraz powinien istnieć określony format wymiany metadanych. Format ten będzie rozumiany przez wszystkie komponenty architektury rozproszonej, gdyż reprezentacja metadanych w każdym jej elemencie będzie taka sama. Taki zdecentralizowany model administrowania ze wspólną reprezentacją metadanych (metamodelem) nosi nazwę **modelu federacyjnego** [5].



Rys. 2. Model zdecentralizowany administrowania środowiskiem DSS

Fig. 2. Administering in DSS environment in the distributed model

Model federacyjny administrowania zyskał w ostatnich latach bardzo dużą popularność. Był on odpowiedzią na trudności, jakie generował model zintegrowany. Na przełomie ostatnich 10 lat powstało kilka organizacji o charakterze niekomercyjnym, których celem było ujednolicenie reprezentacji metadanych w środowiskach hurtowni danych/OLAP. Wielu producentów narzędzi zgłosiło swoją przynależność do tych organizacji i wspierało ich działalność, jednocześnie akceptując wspólne postanowienia co do formy, modelu i architektury systemów wymiany metadanych. Na szczególną uwagę zasługują tutaj dwa standardy:

- Common Warehouse Metamodel (CWM), promowany przez Object Management Group (OMG), którego celem jest umożliwienie wymiany metadanych środowisk hurtowni danych i środowisk *business intelligence* pomiędzy różnymi narzędziami i platformami hurtowni danych oraz repozytoriami metadanych w rozproszonych, heterogenicznych środowiskach [7].
- Open Information Model (OIM), promowany przez grupę Meta Data Coalition (MDC), który miał obejmować wszystkie fazy rozwijania systemów IT poczynając od analizy, kończąc na wdrożeniu kompletnego systemu [6].

W obu standardach wymiana metadanych odbywała się z wykorzystaniem języka XML. Na dzień dzisiejszy standard OIM jest już częścią standardu CWM.

Administrowanie modelem zdecentralizowanym nie jest proste (Rys. 2) - administrator być może pełni tylko rolę koordynatora wymiany w środowisku wymiany metadanych.

Z drugiej strony, gdyby powstały narzędzia administracyjne umożliwiające zarządzanie informacją pochodzącą z globalnego widoku metadanych, wówczas administracja znów skupiłaby się w jednych rękach. Problem ten wydaje się prawdziwym wyzwaniem. Kolejnym trudnym zagadnieniem jest kompatybilność wstecz, z narzędziami, które oparte były na innych standardach przechowywania i wymiany metadanych. I na koniec jeszcze jedna, dość istotna wada – model zdecentralizowany nie umożliwia tworzenia i przechowywania metadanych spoza narzędzi wchodzących w skład architektury środowiska wymiany. Przykładowo, jeśli z konsoli administratora, działającej w oparciu o globalny widok metadanych, tworzona jest nowa informacja opisująca dane środowisko DSS (czyli nowa metadana) - taka informacja, której nie można utworzyć za pomocą żadnego narzędzia w środowisku (może to być np. dodatkowy komentarz, specyfikacja jakiegoś elementu, szczególny rodzaj hierarchii atrybutów) - gdzie i w jakiej postaci zostanie ona zapamiętana?

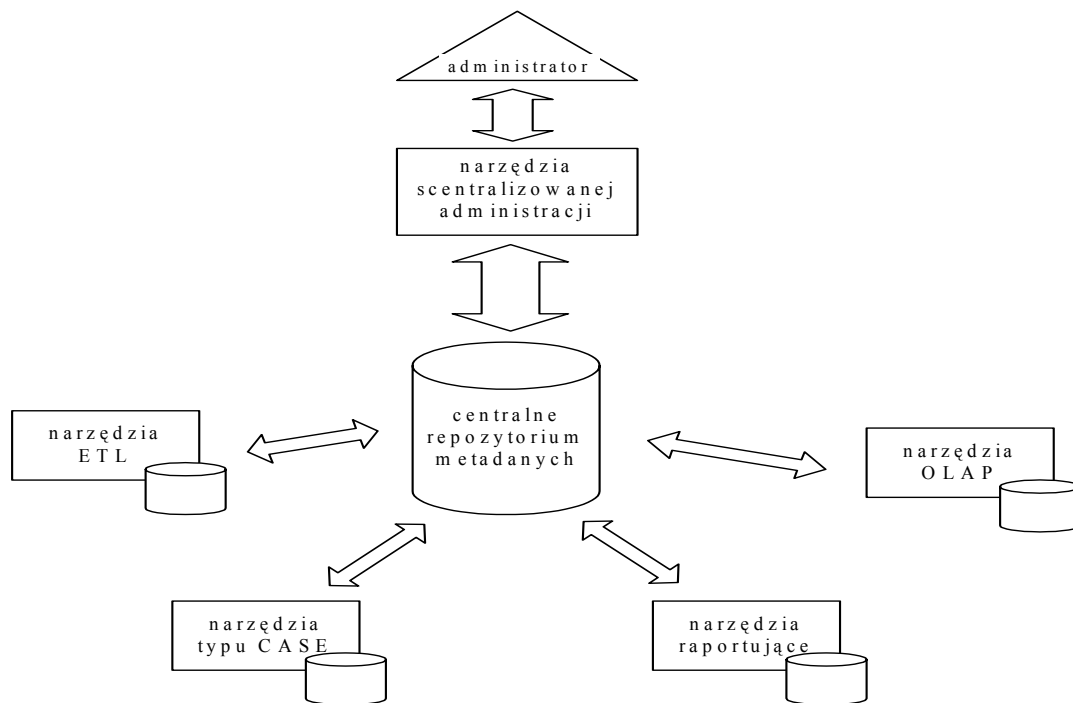
Poniżej autorzy przedstawiają nowy model administrowania systemami hurtowni danych.

#### 4.3. Model kooperacyjny z redundancją metadanych

Model kooperacyjny z redundancją metadanych (Rys. 3), inaczej zwany **integrującym**, to model administrowania, który pozwala rozwiązać problemy występujące w środowiskach opartych na modelach zintegrowanych i zdecentralizowanych. Głównym elementem architektury jest centralne repozytorium metadanych, na którym bazuje narzędzie lub grupa narzędzi scentralizowanej administracji. Kontrola nad systemem znów powraca więc w ręce jednej osoby, którą jest administrator środowiska DSS. Każde z narzędzi wchodzących w skład środowiska generuje metadane do lokalnego repozytorium. Następnie centralne repozytorium jest okresowo synchronizowane z repozytoriami lokalnymi - bądź to z określonym interwałem czasowym, bądź też następuje synchronizacja na żądanie. Ponieważ



model ma charakter integrujący, również z narzędziami starszymi, nie wspierającymi obecnego standardu CWM (OMG), nie możemy założyć, że lokalne repozytoria będą się opierać na wspólnym metamodelu. Musimy jednak zapewnić wspólny format wymiany metadanych (synchronizacji z repozytorium centralnym). Centralne repozytorium powinno zostać stworzone najlepiej na podstawie aktualnego standardu CWM – zapewni to kompatybilność z nowoczesnymi narzędziami. Można też wykorzystać, jako centralne repozytorium, jedno ze specjalizowanych dla celów hurtowni danych repozytoriów dostarczanych przez takich producentów jak: MICROSOFT (REPOSITORY), PLATINUM (REPOSITORY), VIASOFT (ROCHADE) lub ARDENT (METASTAGE) i in. Firmy te dbają zazwyczaj na bieżąco o to, by ich produkty były zgodne z obowiązującymi standardami.



Rys. 3. Model integrujący administracji środowiskiem DSS

Fig. 3. Administering in DSS environment in the integrating model

Korzyści płynące z zastosowania takiego modelu mogą być bardzo znaczące, szczególnie w przedsiębiorstwach, w których występuje duże zróżnicowanie używanych narzędzi. Niedogodności modelu wymuszają konieczność powielania metadanych z lokalnych repozytoriów do repozytorium centralnego. Nie chodzi tu wcale o miejsce, które będzie potrzebne do przechowywania redundantnych informacji – przy możliwościach dzisiejszych stacji roboczych czy serwerów nie powinno być z tym problemu. Chodzi raczej o mechanizmy wymiany, które trzeba będzie zastosować, w szczególności do narzędzi starszej generacji. Nie powinno to jednak zniechęcać potencjalnych administratorów.

Możliwości wykorzystania modelu kooperacyjnego z redundancją metadanych przedstawiono w następnym rozdziale.

## 5. Opis systemu zarządzania HD

Opierając się na zaproponowanym integrującym modelu administrowania środowiskiem hurtowni danych, zbudowano w pełni działający system centralnego zarządzania. W trakcie realizacji projektu postawiono kilka kluczowych założeń - niektóre z nich to:

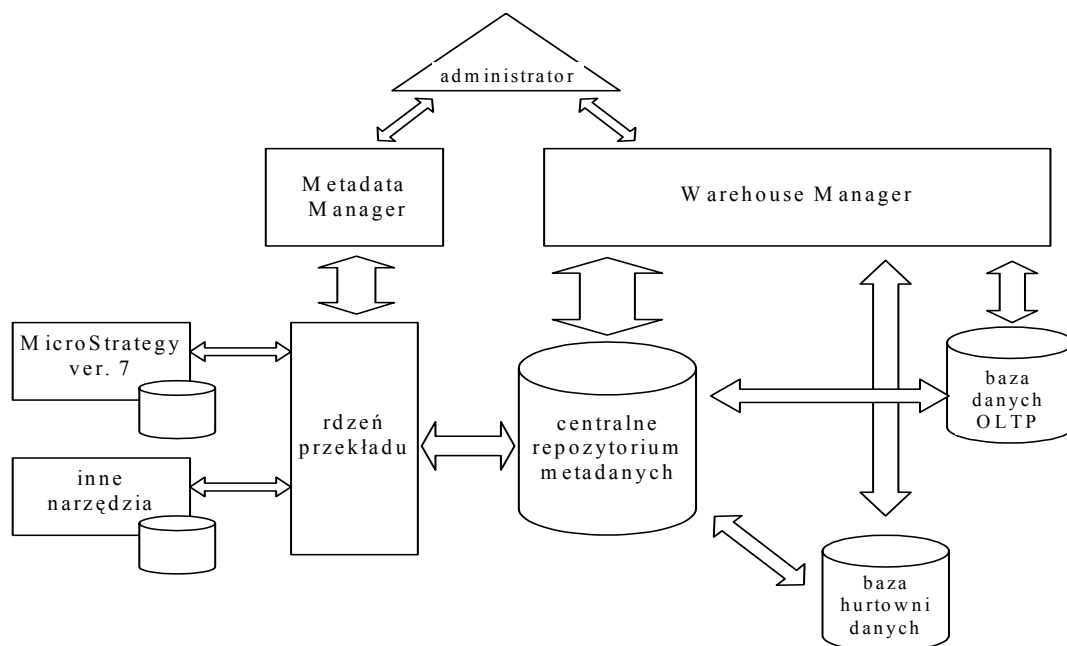
- zarządzanie w środowisku przy użyciu jednego narzędzia,
- kooperacja w środowisku narzędzi różnych producentów, w różnych wersjach,
- kompatybilność narzędzi na poziomie metadanych i wzajemna wymiana metadanych,
- koncentracja metadanych w jednym miejscu dla celów zarządzania.

Przepływ danych w zaprojektowanym systemie ilustruje Rys. 4. Konsola administracyjna składa się z dwóch współpracujących ze sobą programów:

- *Metadata Manager* – odpowiedzialny za wymianę metadanych oraz synchronizację repozytoriów lokalnych z centralnym repozytorium metadanych,
- *Warehouse Manager* – odpowiedzialny za zarządzanie elementami środowiska DSS.

Wymiana metadanych w środowisku odbywa się bądź w sposób bezpośredni, bądź z wykorzystaniem specjalnej warstwy pośredniej, zwanej rdzeniem przekładu. Bezpośrednia wymiana może zaistnieć pomiędzy narzędziami wspierającymi ten sam standard wymiany lub posiadającymi taki sam metamodel repozytorium. W przypadku autorów, centralne repozytorium zostało zaprojektowane zgodnie ze standardem Common Warehouse Metamodel grupy OMG [7] i umożliwia wymianę metadanych z wykorzystaniem formatu XMI (XML Metadata Interchange) [8]. W przypadku narzędzi starszej generacji, wymiana metadanych z centralnym repozytorium następuje pośrednio przez rdzeń przekładu. Rdzeń przekładu stanowi rodzaj kontenera, do którego są wczytywane metadane z lokalnych repozytoriów zanim nastąpi ich eksport do formatu XMI. Jest to konieczne ze względu na to, że starsze narzędzia posiadają własny format metadanych, który najczęściej nie jest zgodny z obowiązującymi standardami. Do testów użyliśmy narzędzi MICROSTRATEGY w wersji 7 (narzędzie klasy OLAP/DSS), które generują własny, lokalny zbiór metadanych opisujących strukturę wielowymiarową i elementy środowiska DSS (raporty, szablony, filtry i in.). MICROSTRATEGY udostępnia zestaw interfejsów (API) umożliwiających dostęp do swojego repozytorium i manipulację informacjami znajdującymi się wewnątrz. Pociąga to za sobą konieczność tworzenia modułów sprzęgających (interfejsów wymiany), które pozwolą na załadowanie lokalnego zbioru metadanych MICROSTRATEGY do rdzenia przekładu. Dla innych narzędzi trzeba stworzyć odpowiedni interfejs wymiany do rdzenia przekładu, ko-

rzyszając ze specyfikacji danego producenta. Rdzeń przekładu jest w tym przypadku hermetyczną strukturą, do której dostęp jest realizowany tylko przez odpowiednie interfejsy wejścia/wyjścia. Nad całością procesu translacji i eksportu czuwa moduł *Metadata Manager*, w którym możemy dokonać odpowiednich odwzorowań między formatami oraz wyboru informacji do przetworzenia.



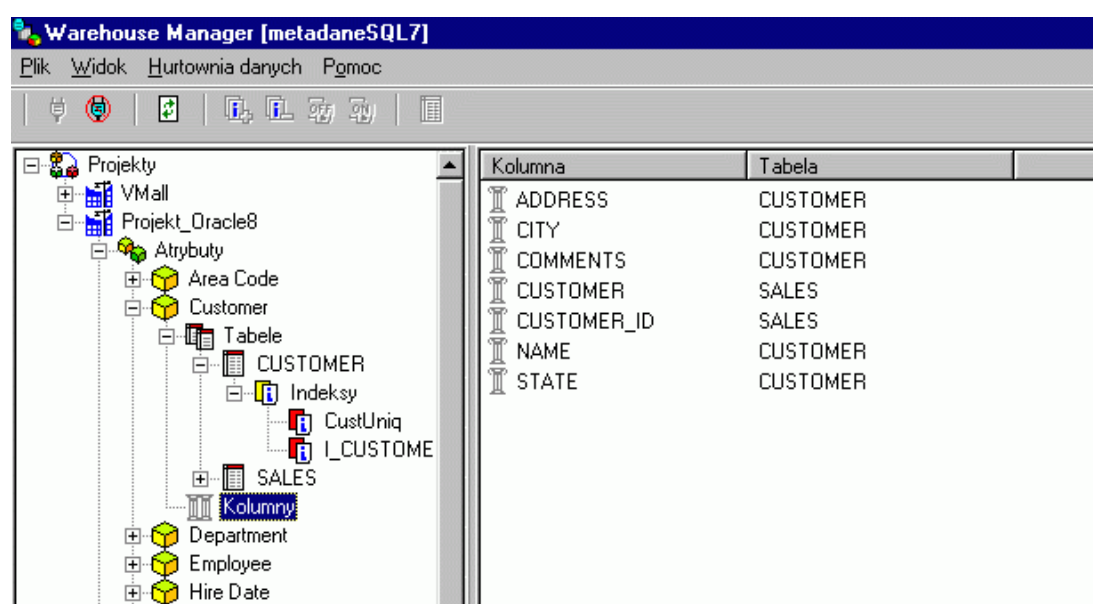
Rys. 4. Architektura systemu z wykorzystaniem modelu kooperacyjnego

Fig. 4. System architecture in the co-operative model

W testowanym systemie następuje również wymiana metadanych pomiędzy systemami zarządzania bazą danych (DBMS) hurtowni danych oraz bazy źródłowej OLTP a centralnym repozytorium. Można również zauważyć (Rys. 4) linie wymiany danych pomiędzy konsolą administratora (*Warehouse Manager*) a bazą hurtowni oraz bazą danych transakcyjną (OLTP). Dzięki tak skonstruowanej magistrali wymiany informacji administrator środowiska ma dostęp zarówno do danych w bazie źródłowej i bazie hurtowni, jak i do struktury tych danych (czyli do metadanych) poprzez centralne repozytorium. Rozwiązanie takie ma wiele zalet - mając dostęp do bazy danych OLTP administrator może np. kontrolować przebieg procesu ETL, a mając dostęp do bazy hurtowni zarządzać procesami agregacji danych. Wadą rozwiązania jest natomiast redundancja metadanych w centralnym repozytorium – wynika ona z przyjętego modelu - jednakże przy obecnych parametrach komputerów jest ona do pominięcia. Bardziej istotnym mankamentem jest wymiana metadanych, pochodzących ze starszych narzędzi, tylko w jednym kierunku, tzn. do centralnego repozytorium. Niestety, niektórzy producenci narzędzi umożliwili dostęp do lokalnych repozytoriów tylko w jedną stronę (np. producenci zrzeszeni niegdyś w OLAP Council stosowali tzw. Multidimensional-API

o charakterze „tylko do odczytu”). Oznacza to w praktyce, że możliwe jest odczytanie metadanych i ich skopiowanie do repozytorium centralnego, natomiast zmiany w metadanych w centralnym repozytorium nie mogą zostać z powrotem zaktualizowane w repozytoriach lokalnych. Nie jest to jednak wada modelu, tylko samych narzędzi-producentów metadanych.

Okno konsoli administratora przedstawia Rys. 5. Program bazuje na informacjach (metadanych) znajdujących się w centralnym repozytorium, które wczytuje i odpowiednio interpretuje przydzielając miejsce w określonym folderze (lewy panel programu *Warehouse Manager*). Układ paneli programu przypomina program *Eksplorator Windows*, a informacje są pogrupowane w odpowiednich folderach.



Rys. 5. Konsola administratora

Fig. 5. Administrative console

Głównym folderem w drzewie jest folder *Projekty*, w którym znajdują się informacje na temat środowisk DSS, którymi w danym momencie możemy zarządzać. Wszystkie elementy środowiska DSS są odpowiednio pogrupowane i znajdują się w odpowiednich folderach, np. atrybuty wymiarów w folderze *Atrybuty*, tabele w folderze *Tabele*, indeksy w folderze *Indeksy* itd. Kolejne elementy środowiska można eksplorować w głąb do elementów podrzędnych, tzn. takich, na których opiera się struktura elementu nadrzędnego, np. rozwijając atrybut wymiaru możemy zejść niżej do poziomu tabeli, na bazie której został zbudowany atrybut oraz kolumn stanowiących o formie atrybutu, możemy również otrzymać informację o założonych w tym przypadku indeksach na tabelach wymiaru w hurtowni danych (Rys. 5). Taka prezentacja obiektów i ich wzajemnych powiązań możliwa jest tylko dzięki istnieniu w repozytorium metadanych odpowiednich odwzorowań warstwy wielowymiarowej do warstwy technicznej, a więc powiązaniu metadanych biznesowych z technicznymi. Występowanie w strukturze technicznej centralnego repozytorium nie tylko

centralizuje proces administrowania, ale daje również dodatkowe korzyści, o których już wspominaliśmy, dotyczące wprowadzania dodatkowych elementów środowiska DSS. Za przykład niech posłużą nam indeksy tabel – w prezentowanym systemie nie tylko mamy możliwość ich tworzenia, ale również włączania i wyłączania (co umożliwiają tylko niektóre DBMS). Wszystko to odbywa się za pomocą odpowiednich zapisów w repozytorium. Dodatkowe elementy mogą być tworzone na różnym poziomie opisu informacji znajdującej się w środowisku DSS, tzn. na różnych poziomach metadanych.

## 6. Podsumowanie

Proces administrowania jest niezmiernie ważnym aspektem w cyklu życia systemów hurtowni danych. Obejmuje on wiele skomplikowanych zagadnień, które przytoczono w niniejszym artykule. Zadania administratora i jego miejsce w strukturze przedsiębiorstwa będą zdeterminowane przez przyjęty model administracji. Kluczową rolę w całym procesie administracji odgrywają metadane, które mogą pochodzić z różnych źródeł i być przechowywane w różnym formacie. Centralizacja metadanych w jednym miejscu dla procesu zarządzania jest bardzo trudna, jednak daje najlepsze możliwości administracyjne. Wykorzystanie jednego repozytorium metadanych i wymiana metadanych między narzędziami jedynie poprzez to repozytorium są możliwe praktycznie tylko przy wykorzystaniu narzędzi jednego producenta. Zależność tą można zaobserwować w modelu zintegrowanym. Z drugiej strony, stosując model federacyjny, umożliwia się niezależną wymianę metadanych w środowisku, tworząc pewien globalny widok metadanych kosztem możliwości zarządczych. Odpowiedzią na wady obu modeli i próbą rozwiązania ich problemów jest proponowany przez autorów model kooperacyjny z redundancją metadanych. Model ten nie tylko koncentruje proces administrowania w jednym miejscu, ale dodatkowo tworzy środowisko kompatybilne dla starszych wersji narzędzi-producentów metadanych. Zapewnia również swobodną wymianę metadanych z narzędziami nowej generacji, które wykorzystują ogólnie przyjęte standardy wymiany. Dla tak zaprojektowanego modelu zbudowano konsolę administracyjną, która umożliwia zarządzanie w systemach hurtowni danych. Jak dotąd, rozwiązanie to potwierdziło swą dużą przydatność w realizowanych projektach hurtowni danych.

Prace mające na celu poszerzanie możliwości funkcjonalnych zbudowanego środowiska administracyjnego trwają nadal. Autorzy mają świadomość pewnych wad prezentowanego rozwiązania, jednakże bilans zysków i strat wciąż jest zdecydowanie korzystny, a to uzasadnia przyjęty kierunek rozwoju.

## LITERATURA

1. Kimball R., Reeves L., Ross M., Thornthwaite W.: The Data Warehouse Lifecycle Toolkit, Expert Methods for Designing, Developing and Deploying Data Warehouses. John Wiley & Sons, Inc., New York, 1998.
2. Gorawski M.: Praktyczne aspekty projektowania hurtowni danych. *Studia Informatica*, vol.24, nr 4 (56), 2004.
3. Staudt M., Vaduva A., Vetterli T.: The Role of Metadata for Data Warehousing, Proc. International Database Engineering and Applications Symposium (IDEAS '01), Grenoble, France, July 2001.
4. Stöhr T., Müller R., Rahm E.: An Integrative and Uniform Model for Metadata Management in Data Warehousing Environments, Preceedings of the International Workshop on Design and Management of Data Warehouses (DMDW'99).
5. Staudt M., Vaduva A., Vetterli T.: Metadata Standards for Data Warehousing: Open Information Model vs. Common Warehouse Metamodel, *SIGMOD Record* 29(3): 68-75 2000.
6. Meta Data Coalition. Open Information Model. Version 1.0, August 1999, <http://www.mdcinfo.com>
7. Object Management Group, Common Warehouse Metamodel (CWM) Specification, September 1999, <http://www.omg.org>
8. Object Management Group, XML Metadata Interchange (XMI) Specification, Version 2.0, May 2003, <http://www.omg.org>

Recenzent: Dr inż. Arkadiusz Sochan

Wpłynęło do Redakcji 29 styczeń 2004 r.

## Abstract

Administering process takes very important place in the data warehouse lifecycle. The nature of the data warehouse environment is not solid, but changes all the time. Developers and creators of the environment have to exchange necessary information about structure of data in the data warehouse. A control over all taken actions and exchanged information became an absolute necessity. The article introduces crucial aspects of the management process in the data warehouse environment. It describes why the process is so important. We define fundamental tasks the data warehouse system administrator is involved in and

we investigate existing models of administering - integrated model and distributed model (with a special attention to federated model). We compare these two models to show some advantages and disadvantages of using them in the exchange information structures inside the company. We also present our cooperative management model as a third possibility, which eliminates inconveniences of the integrated and classic distributed models. Special place we dedicate to role of metadata that describes the structure of information in the data warehouse in various level of abstraction. We show the way we can store metadata and exchange it in mentioned models. Additionally, we introduce our management console application that is based on the cooperative management model. The console is an existing, real application that was built to simplify the process of administering in the distributed, heterogeneous environments.

### **Adresy**

Marcin GORAWSKI: Politechnika Śląska, Instytut Informatyki, ul. Akademicka 16,  
44-101 Gliwice, Polska, Marcin.Gorawski@polsl.pl

Dariusz MROZEK: Politechnika Śląska, Instytut Informatyki, ul. Akademicka 16,  
44-101 Gliwice, Polska, mrozek@zti.iinf.polsl.gliwice.pl