

PROJEKT HURTOWNI DANYCH DLA PRZEDSIĘBIORSTWA PRODUKCYJNO-HANDLOWEGO W ŚRODOWISKU MS SQL SERVER

Katarzyna BŁASZCZYK, Ryszard KNOSALA

Streszczenie: Artykuł opisuje podstawową tematykę związaną z systemami hurtowni danych i możliwości ich implementacji w środowisku analitycznym Microsoft SQL Server 2000. W tym celu przedstawiono projekt hurtowni danych dla przedsiębiorstwa produkcyjno-handlowego w oparciu o to środowisko, w ramach którego omówiono: wymagania OLAP przedsiębiorstwa, schemat logiczny systemu, utworzenie i integracje źródeł danych oraz budowę i możliwości analizy wielowymiarowych kostek danych.

Słowa kluczowe: hurtownia danych, przedsiębiorstwo produkcyjno-handlowe, OLAP, Microsoft SQL Serwer, Data Warehousing Framework.

1. Wstęp

Tradycyjne systemy oparte o model przetwarzania danych w trybie on-line (OLTP - Online Transaction Processing), czyli przetwarzanie transakcji w czasie rzeczywistym są satysfakcjonujące jedynie w przypadku bieżącej obsługi działalności firmy i okazują się niewystarczające w podejmowaniu decyzji najważniejszych - decyzji strategicznych. Ten klasyczny model przetwarzania danych nie wspomaga procesów analizy danych, gdyż w znacznie mniejszym stopniu systemy te umożliwiają operacje agregacji danych, wykonywania podsumowań, czy też optymalizacji złożonych zapytań. Rozwiązaniem tego problemu jest skorzystanie z hurtowni danych - systemu opartego na innym modelu przetwarzania, nazwanym przetwarzaniem analitycznym w trybie on-line (OLAP- Online Analytical Processing).

Według Inmon'a hurtownia danych, zwana także magazynem danych (ang. data warehouse) to „*zorientowana podmiotowo, zintegrowana, zróżnicowana czasowo i trwała kolekcja (bazą) danych zaprojektowaną i zaimplementowaną dla potrzeb wspomaganie procesu podejmowania decyzji, w której dane odnoszą się do określonej chwili czasowej*”. Innymi słowy, jest to zbiór technologii będący podstawą systemów wspomaganie decyzji, określane często jako „serce systemu wspomaganie decyzji” [1, 2].

Omawiana struktura zawiera informacje zebrane z danych transakcyjnych pochodzących z różnych heterogenicznych źródeł, zorientowanych wokół przedmiotów analizy. Źródłami mogą być: scentralizowane lub rozproszone bazy danych, bazy relacyjno-obiektowe, bazy obiektowe bądź też inne źródła tj. arkusze kalkulacyjne, pliki tekstowe, dokumenty XML itp.. Takie rozwiązanie odseparowuje część analityczną od części transakcyjnej, czyli zmniejsza obciążenie systemu, a także daje możliwość wyłonienia potrzebnych informacji z różnych źródeł w celu odkrywania zależności między nimi oraz zoptymalizowania struktury pod kątem zadawanych zapytań.

Dane potrzebne do podejmowania decyzji strategicznych przeważnie nie wymagają uaktualnienia z dokładnością do minuty ale powinny zawierać historię przedmiotu. Wobec

tego hurtownia danych jest zasilana, według określonego planu, danymi pochodzącymi z innych systemów zewnętrznych. Informacje w systemie są zatem stałe i przeważnie nie pozwala się użytkownikom na ich uaktualnianie i kasowanie. Dzięki takiemu rozwiązaniu dane przechowywane w systemie nie są najświeższe, ale są aktualne na dany moment czasu, nawet wtedy jak w bazie transakcyjnej uległy one zmianie [2, 3, 4].

W powyższym rozumowaniu hurtowni danych można mylnie postrzec, jako jedynie dużą kolekcję danych o specyficznym charakterze. Pojęcie hurtowni danych należy jednak traktować jako długofalowy projekt o złożonej architekturze i strategicznym znaczeniu dla przedsiębiorstwa [5]. W architekturze hurtowni można bowiem wyróżnić szereg następujących elementów (por. [5]):

- heterogeniczne wewnętrzne i zewnętrzne źródła danych,
- proces i narzędzia do integracji danych z różnych źródeł tzw. proces ETL (z ang. Extraction, Transformation, Loading),
- oddzielną, stałą bazę danych wraz z metadanymi,
- narzędzia dostępne dla użytkowników: narzędzia eksploracji danych, narzędzia OLAP narzędzia do tworzenia raportów i zapytań [6], systemy SWD.

2. Analiza danych w MS SQL Server

Istnieje wiele systemów informatycznych, które umożliwiają usługi związane z obsługą hurtowni. Jedno z takich rozwiązań oferuje firma Microsoft, dołączając do systemu MS SQL Server szereg składników umożliwiających zbudowanie zintegrowanego systemu zarządzania i analizowania danych w przedsiębiorstwie – oprogramowanie Data Warehousing Framework. Tymi narzędziami są: usługi Analizy Danych (AS - Analysis Services), Microsoft Repository, usługi PivotTable, API OLE DB, usługi Data Transformation Services SQL Server (DTS), liczne kreatory SQL Server.

Analysis Services stanowi hybrydowy serwer OLAP, który może obsługiwać wielkie (wielkości setek gigabajtów), wielowymiarowe bazy danych, przechowywane jako struktury MOLAP, ROLAP lub HOLAP, albo jako dowolne kombinacje tych trzech struktur. Może on łączyć się nie tylko z tabelami programu SQL Server, ale również z innymi relacyjnymi bazami danych (np. Oracle), niezależnie od tego, czy są one uruchamiane w systemie Windows NT czy nie. Serwer OLAP oferuje funkcjonalność przeprowadzania obliczeń „ad hoc”, filtrowania i klasyfikacji, która może być wywoływana przez narzędzia klienckie korzystające z interfejsu OLE DB for OLAP API i obsługiwanego przez ten interfejs, bogatego języka kwerend wielowymiarowych, MDX (Multidimensional Expressions) [7, 8, 9].

Microsoft Repository jest przestrzenią przechowującą wszystkie narzędzia i źródła danych w postaci standardowego języka zapisu i odczytu – języka Open Information Model (OIM). Umożliwia to korzystanie z różnych narzędzi Microsoft i firm trzecich przy tworzeniu kompleksowego systemu wspomagania decyzji.

API OLE DB stanowi uniwersalny dostęp do różnych źródeł danych i informacji na wielu platformach [10]. Z tego dostępu korzysta m.in. DTS pozwalając na importowanie, eksportowanie i przekształcanie danych z prawie każdego źródła danych do innego źródła, począwszy od najprostszego – tekstu ASCII. Przekształcenia DTS mogą być tworzone przy pomocy wbudowanego kreatora, kodowane w języku skryptowym (VBScript, JavaScript) lub za pomocą języka Visual Basic.

Kreatorzy i projektanci stanowią także liczną grupę narzędzi graficznych, zawartych w Microsoft SQL Server 2000, pozwalających na łatwe wykonywanie różnych czynności

administracyjnych, tworzenie i przeglądanie struktur baz danych oraz wykonywanie analiz [11, 10, 8].

3. Projekt hurtowni danych dla przedsiębiorstwa produkcyjno-handlowego

Omówiony projekt jest realizowany na przykładzie firmy produkcyjno - handlowej, która zajmuje się wytwarzaniem i dystrybucją na szerszą skalę części, podzespołów i elementów konstrukcyjnych maszyn i urządzeń, służących do produkcji okien aluminiowych i plastikowych. Celem projektu jest opracowanie systemu scentralizowanej hurtowni danych, wspomagającej podejmowanie strategicznych decyzji. Zaprojektowana hurtownia danych ma być źródłem analiz i raportów, ale także miejscem, gdzie zarząd spółki znajduje nowe pomysły na rozwój biznesu i zweryfikuje własne hipotezy dotyczące zachowań rynku, klientów i dostawców. Może również wskazywać miejsca nadużyć, zbędnych działań i możliwości oszczędności w złożonym procesie produkcji, sprzedaży i zakupów. Opracowany system, aby oddawał pełny obraz działalności, pozwoli zebrać dane z wielu działów firmy.

Projektowanie hurtowni danych, jako proces złożony, wymaga wykonania następujących etapów (por. [2]):

- określenie wymagań przedsiębiorstwa,
- określenie fizycznego projektu hurtowni danych,
- analiza, utworzenie i integracja potrzebnych źródeł danych,
- utworzenie hurtowni danych,
- wybór i budowa narzędzi dostępowych dla użytkowników,
- testowanie i ewentualne poprawki.

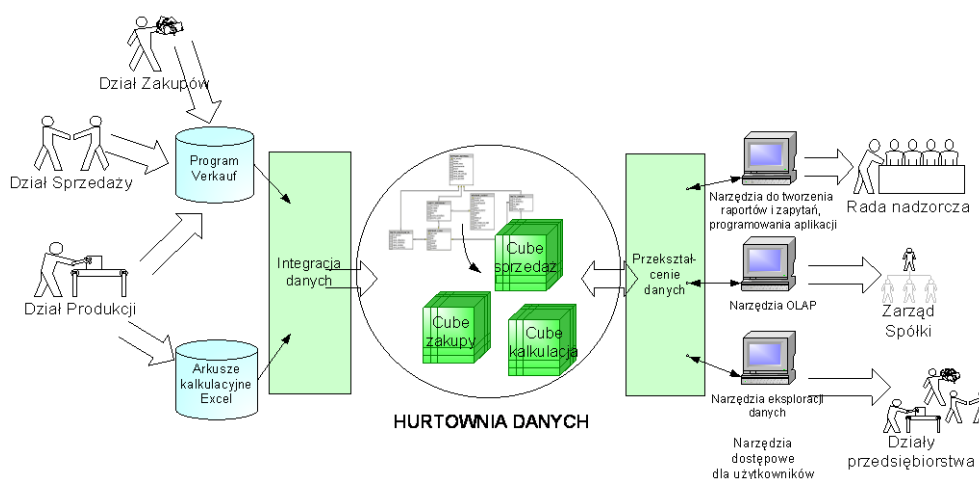
3.1. Wymagania OLAP dla firmy

Przedsiębiorstwo zajmuje się działalnością produkcyjno-handlową, w związku z czym wymagana jest tutaj analiza danych w środowisku SWD dotycząca klientów oraz sprzedawanych i produkowanych artykułów. Informacje te potrzebne są do zminimalizowania kosztów dystrybucji i lepszego zrozumienia potrzeb rynków, dotyczących najbardziej popularnych produktów firmy, jak również do zoptymalizowania oferty produkowanych artykułów poprzez ograniczenie najmniej opłacalnej produkcji, a skupienie się na artykułach generujących większy zysk. Duży odsetek tworzonych zysków w firmie pochodzi z działalności handlowej przedsiębiorstwa, zatem ważnym elementem prócz samej sprzedaży artykułów jest także sfera zakupów. Szczegółowe wymagania przetwarzania i reprezentacji danych w projekcie są następujące:

- Przeglądanie rzeczywistych jednostek sprzedaży oraz kalkulowanego zysku z ostatnich lat za pomocą hierarchii artykułów (wszystkie artykuły, rodzaje artykułów, grupy produktów itp.), hierarchii geograficznej klientów (kraje, regiony, miasta), specyficznych wymagań klientów dla różnych przedziałów czasowych (cały okres, lata, kwartały, miesiące).
- Kalkulacja opłacalności produkcji artykułów z możliwością przeglądania za pomocą hierarchii artykułów i hierarchii czasowej.
- Przeglądanie rzeczywistych jednostek wykonanych zakupów z uwzględnieniem hierarchii kupowanych artykułów, hierarchii i charakterystyki dostawców dla różnych przedziałów czasowych.

- Przeglądanie danych z uwzględnieniem bezpośredniego okresu tygodniowego, z uwagi na sposób składania zamówień w firmie.

Szczegółowe wymagania zostaną zrealizowane przez utworzenie trzech wielowymiarowych kostek danych, dotyczących sprzedaży, zakupów i kalkulacji. Wymaga to zebrania danych z trzech działów firmy: działu sprzedaży, produkcji i zakupów. Schemat źródła informacji hurtowni danych wraz z ich przeznaczeniem przedstawia rysunek 1.



Rys. 1. Źródła informacji hurtowni danych oraz ich przeznaczenie

Do zaimplementowania hurtowni danych zostanie użyta technologia MS SQL Serwer 2000 z wykorzystaniem narzędzi przetwarzania danych OLAP.

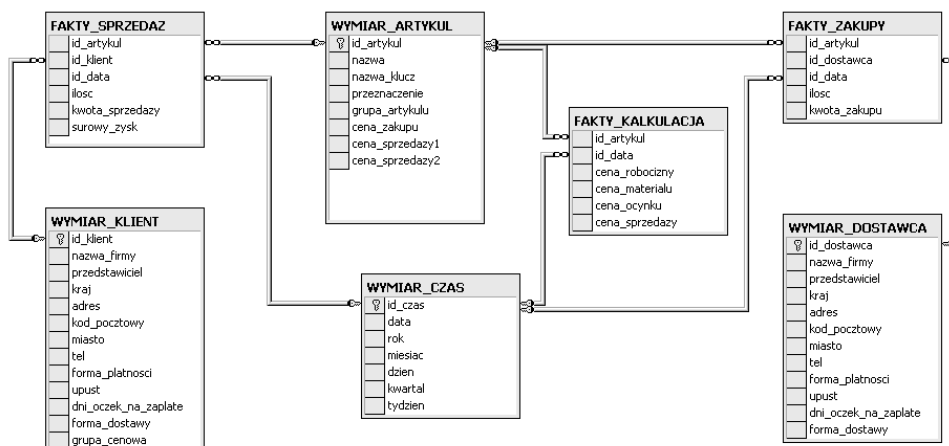
3.2. Logiczny projekt hurtowni danych

Projektując hurtownię danych do potrzeb analitycznych, projektant ma do wyboru trzy podstawowe schematy logiczne [12, 13, 14, 10]:

- schemat gwiazdy (ang. star schema),
- schemat płatka śniegu (ang. snowflake schema),
- schemat konstelacji faktów (ang. facts constellation).

Typ użytej struktury wynika z wymagań, które powinien spełniać tworzony system oraz ze stosowanego stylu projektowania i znajomości danych źródłowych. Jeżeli projektantem jest ta sama osoba, która tworzyła bazy operacyjne, może mieć ona większe problemy z wyobrażeniem sobie uproszczonego modelu – zastosuje zatem schemat płatka śniegu [10].

Schemat znajdujący się na rysunku 2 najlepiej odzwierciedla wymagania firmy, umożliwia wykorzystanie łatwo dostępnych, często niewykorzystanych danych w przedsiębiorstwie i może być efektywnie wykorzystany przez usługi wielowymiarowe.



Rys. 2. Logiczny projekt hurtowni danych

Tabele: WYMIAR_DOSTAWCA, WYMIAR_KLIENT, WYMIAR_CZAS, WYMIAR_ARTYKUL stanowią, jak sama ich nazwa wskazuje, źródła wymiarów kostki – tabele informacyjne bądź tabele strukturalne.

Tabele: FAKTY_SPRZEDAZ, FAKTY_ZAKUPY, FAKTY_KALKULACJA odzwierciedlają przedmiotową działalność przedsiębiorstwa i stanowią główne tabele faktów w schemacie hurtowni danych.

Schemat projektowanej hurtowni jest schematem wielogwieźdźdźistym. Składa się bowiem z tabel faktów i połączonych z nimi bezpośrednio tabel wymiarów. Tabele wymiarów są jednakże połączone z więcej niż jedną tabelą faktów. Ponieważ jest to struktura gwieździasta, tabele faktów są umieszczone w trzeciej formie normalnej, czyli dane w nich będą się powtarzać. Ma to duże znaczenie ze względu na poprawie wydajność systemu i umożliwienie obsługi wielowymiarowości.

3.3. Utworzenie i integracja potrzebnych źródeł danych

Dane odnośnie artykułów, sprzedaży, zakupów oraz kontrahentów, firma przechowuje w bazie niemieckiego programu Verkauf. Jest to program z niewielkimi możliwościami analitycznymi. Istnieje jednakże możliwość wyeksportowania danych zawartych w bazie programu do postaci plików tekstowych (flat). Taka postać informacji będzie mogła być bezpośrednio zaimportowana, poprzez narzędzia DTS, do bazy danych MS SQL Server, czyli do naszej hurtowni danych. Jednakże pliki wygenerowane przez program Verkauf zawierają sporo danych nadmiarowych i nieustrukturyzowanych: informacje dla drukarki, informacje o akapitach i przerwach, puste komórki, dane skumulowane bądź też dane w różnych formatach. W celu oczyszczenia i wydobywania potrzebnych informacji zostały napisane, przy wykorzystaniu języka programowania VBA (Visual Basic for Applications), odpowiednie skrypty programu Excel (Makra). Uruchomienie tych skryptów na plikach tekstowych wyeksportowanych z programu Verkauf umożliwi wstępne oczyszczenie danych.

Kalkulacje opłacalności produkcji wybranych artykułów, kierownicy produkcji przeprowadzają w arkuszach kalkulacyjnych programu Excel. Nie są to dane dotychczas

porównywane i wykorzystywane do analizy, gdyż kalkulacja pojedynczego artykułu zawarta jest w osobnym pliku, te zaś są rozrzucone są po różnych katalogach w zależności od numeru analizowanego produktu. Wymagany jest zatem duży nakład czasu, aby wszystkie te informacje zebrać w jednym miejscu. Rozwiązaniem problemu, podobnie jak poprzednio, jest utworzenie skryptu programu Excel, który umożliwi zebranie informacji z odpowiednich komórek z kilkuset arkuszy kalkulacyjnych i zapisanie ich w jednym pliku.

Kolejnym etapem, po oczyszczeniu danych w powyższym wstępnym procesie transformacji, jest transport danych do hurtowni. Ruch ten jest możliwy dzięki narzędziom DTS. Umożliwiają one zaprojektowanie systemu przeniesienia informacji ze źródeł (plików tekstowych i plików „.xml”), końcowej transformacji (np. usunięcie wierszy powtarzających się czy ujednoczenie formatów) oraz załadowania danych w docelowe struktury hurtowni danych, w taki sposób, aby dane w hurtowni były „czyste”, sprawdzone i właściwie zagregowane.

3.4. Budowa i analiza „wielowymiarowych kostek danych”

Jedną z podstawowych cech systemów OLAP jest wielowymiarowe ujęcie danych (ang. multidimensional data model - MDD) oparte na tzw. wielowymiarowych kostkach danych, będącymi kolekcjami miar numerycznych (faktów) i atrybutów (wymiarów) przypisanych odpowiednio do osi. Każdy wymiar kostki (oś) jest opisany zbiorem atrybutów i od niego zależy wartość każdej miary, będącej punktem w wielowymiarowej przestrzeni wymiarów. Ponadto atrybuty wymiaru mogą tworzyć hierarchię atrybutów, wspomagającą wykonywanie i przeglądanie operacji na kostce.

Schemat logiczny hurtowni danych (rys. 2) umożliwia analizę danych przy pomocy trzech kostek wielowymiarowych: kostki sprzedaży, kostki zakupów i kostki kalkulacji.

Podstawową tabelą faktów na której opiera się kostka „Cube_sprzedaz” jest tabela FAKTY_SPRZEDAZ. Wybrane kolumny tabeli faktów (np. kwota sprzedaży z Sum jako funkcją agregacji, ilość sprzedanych artykułów - Sum, Liczba zamówień - Count, Liczba klientów – Distinct Count) będą miarami kostki, na których będą wykonywane kalkulacje. Kolejnym elementem jest ustalenie wymiarów kostki. Aby ułatwić analizę sprzedaży i zysków ze sprzedaży artykułów najwygodniej jest utworzyć wymiar składający się z elementów tabeli WYMIAR_ARTYKUL, zachowując przy tym poszczególne poziomy wymiaru od ogólnego uszeregowania artykułów, czyli przeznaczenia poprzez poszczególne grupy artykułów, nazwy kluczowe po dane szczegółowe (id artykułu). Aby umożliwić przeglądanie analiz z uwzględnieniem specyficznych grupy klientów należy utworzyć kolejne wymiary tj.: wymiar hierarchiczny klientów pod względem położenia geograficznego, wymiar klientów bezpośrednio według nazw firm, wymiar określający formę dostawy artykułu do klienta, formę płatności za artykuły oraz wymiar uwzględniający listę cenową klientów. Ostatni wymiar będzie stanowić wymiar czasu. W celu głębszej analizy utworzone zostaną wymiary kalkulowane serwera OLAP (calculated members). Każdy z takich elementów powiązany jest obliczeniowo z miarami wymiarów np. miara kalkulowana struktury sprzedaży według obszarów (%) określona jest wyrażeniem:

$$\frac{[Measures].[Kwota Sprzedazy]}{(Ancestor([Klient_geo], [Klient_geo].[All]))}, [Measures].[Kwota Sprzedazy]},$$

natomiast struktura sprzedaży wg artykułów (%) określona jest następująco:

[Measures].[Kwota Sprzedazy]/ (Ancestor([Artykuły].[Hier_artykuły],
[Artykuły].[Hier_artykuły].[All])), [Measures].[Kwota Sprzedazy]).

Microsoft SQL Server zawiera moduł umożliwiający graficzne analizowanie wielowymiarowych kostek. Wymiary i miary kostki można tutaj dowolnie przesuwac, dzięki czemu mamy możliwości różnych analiz i raportów na bazie operacji wykonywanych na wielowymiarowych danych. Analizę kostki „Cube_sprzedaz” można rozpocząć od specyfikacji klientów wg kryteriów tj. grupy asortymentowe produktów, wartość sprzedaży, efektywność sprzedaży, obszary geograficzne sprzedaży (rysunek 3), specyficzne cechy odbiorców np. sposób płatności czy forma dostawy. Powstałe wymiary (umieszczone nad tabelką) pozwalają dodatkowo filtrować zestawienia. Podział odbiorców na jednorodne grupy pozwala przedsiębiorstwu określić i ulepszyć politykę produktową, cenową i transakcyjną, kanały dystrybucji oraz działania promocyjne.

The screenshot shows the 'Cube Browser - Cube_sprzedaz' window. At the top, there are several filter dropdowns: 'Czas' (All Czas), 'Forma dostawy kl' (All Forma dostawy kl), 'Forma płatności kl' (All Forma płatności kl), 'Lista cen' (All Lista cen), and 'Klient nazwa' (All Klient nazwa). Below the filters is a pivot table with the following structure:

| MeasuresLevel | | - Przeznaczenie | + Grupa | | | |
|---------------|----------------|----------------------------|---------------------------|---------------------------|----------------------|-----------------|
| | | Strukt. sprzed wg obsz (%) | | | | |
| | | - Handel | | + Produkcja | | |
| - Kraj | + Kod Pocztowy | Handel Total | - Detal- część, narzędzia | Maszyny i ich wyposażenie | + Wyrób gotowy-detal | Produkcja Total |
| + | Total | 0,39% | 0,43% | 0,01% | | |
| | D Total | 0,46% | 0,13% | 2,42% | 48,17% | 20,56% |
| | + 2826 | | | | | 0,02% |
| | + 56858 | 0,43% | 0,13% | 2,03% | 48,17% | 20,44% |
| | + 66954 | | | | | 0,02% |
| | + 72379 | 0,04% | | 0,39% | | 0,07% |
| | + 98617 | | | | | 0,00% |
| + | H Total | 0,00% | 0,00% | | | |
| | PL Total | 96,39% | 98,97% | 73,77% | 51,83% | 79,39% |
| | + | | | | | 0,05% |
| | + 00-043 | 0,64% | 0,28% | 3,97% | | |
| | + 00-102 | 0,21% | 0,19% | 0,35% | | 0,00% |
| | + 00-108 | 0,01% | | 0,06% | | |
| | + 00-113 | 1,78% | 1,48% | 4,62% | | 0,03% |
| | + 00-180 | 0,34% | 0,37% | | | |
| | + 00-339 | 0,09% | 0,10% | | | |
| | + 00-613 | 0,54% | 0,54% | 0,54% | | 0,06% |

At the bottom of the window, there is a note: "Double-click a member to drill up or down." and buttons for "Close" and "Help".

Rys. 3. Zestawienie struktury sprzedaży wg obszarów geograficznych z uwzględnieniem grup artykułów

Przy pomocy kostki „Cube_sprzedaz” możliwa jest także analiza ukierunkowana na artykuły. Dzięki czemu można przejrzeć artykuły, których sprzedaż jest największa bądź na które w ogóle nie ma popytu. Możliwe jest porównanie sprzedaży tych artykułów w poszczególnych okresach czasu czy też ze względu na grupy klientów itp.

Kostka „Cube_zakupy”, z tabelą FAKTY_ZAKUPY jako źródło miar i tabelami WYMIAR_ARTYKUŁ, WYMIAR_DOSTAWCA i WYMIAR_CZAS jako źródło atrybutów i hierarchii, umożliwi analogiczną analizę danych w stosunku do dostawców firmy i zakupywanych przez firmę artykułów. Analiza ta pozwoli określić strategicznych

dostawców, strukturę i wartość dokonywanych zakupów, wskazywać miejsca nadużyć i możliwości zaoszczędzenia.

Tabele FAKTY_KALKULACJA, WYMIAR_ARTYKUŁ, WYMIAR_CZAS będą użyte do utworzenia kolejnej kostki – wielowymiarowej kostki „Cube kalkulacja”. Umożliwi ona ocenę opłacalności produkcji wybranych artykułów. Z tabeli faktów należy wyszczególnić szereg miar zagregowanych, począwszy od sum (Sum) poszczególnych kosztów produkcji artykułów np. ceny robocizny czy materiału, po ilość (Count) analizowanych elementów. Wartości te nie będą wyświetlane ale są niezbędne do obliczeń wartości średnich, a pośrednio do kalkulacji zysków. Do analizy tej kostki można użyć wymiarów już utworzonych poprzednio, bądź można utworzyć dodatkowe wymiary np. wymiar zawierający tylko nazwę i numer artykułu, dzięki czemu będziemy mogli przeglądać elementy kalkulacyjne artykułów bezpośrednio identyfikowanych po nazwie. Wyszczególnione miary pozwolą nam obliczyć miary kalkulowane tj.: średnią cenę robocizny, materiału, ocynku czy cenę sprzedaży, zysk jednostkowy, procentowy zysk jednostkowy, procentowy udział kosztu materiałów i robocizny w cenie sprzedaży itp. Uwzględnienie wymiarów i hierarchii pozwoli nam przedstawić wszystkie analizowane kalkulacje w różnych przedziałach czasowych (dni, tygodni, miesięcy, kwartałów itp.) dla wybranych artykułów według grup, kluczy przewodnich, nazw lub dla poszczególnego artykułu.

Przeglądanie danych za pomocą dostępnej przeglądarki kostek oraz wyszukiwanie artykułów, których produkowanie jest dla przedsiębiorstwa nieopłacalne czy też najbardziej opłacalne, było by trochę niewygodne. Jednakże mamy możliwość tworzenia dowolnych zapytań i widoków danych określających poszczególne punkty kostki za pomocą rozszerzeń wielowymiarowych (MDX). Elementy te pozwalają na dostosowanie procesów OLAP do specyficznych wymagań za pomocą stawianych zapytań rozszerzonego języka SQL. W usługach OLAP zawarte jest narzędzie o nazwie MDX Sample, które pozwala na przeglądanie rezultatów zapytań MDX. Struktura przykładowego zapytania jest następująca:

```
select
  {[Measures].[Zysk jednostkowy],[Measures].[Sr Cena Materiału],
  [Measures].[Zysk jednostkowy %],[Measures].[Udział % materiału]} on columns,
  filter(
    [Art_szkrocony].[Id Artykuł].members,[Measures].[Udział % materiału]>50),
  [Measures].[Sr Cena Materiału]>20
  ) on rows
from Cube_kalkulacja
where [Czas].[All Czas].[2004],[Art_grupa].[All Art_grupa].[Wyrób gotowy-detaj]).
```

Wynikiem zapytania jest grupa artykułów, dla których udział kosztu materiału w cenie sprzedaży tego produktu jest większy niż 50% oraz średnia cena tego materiału jest większa od 20 (Euro).

3.5. Zastosowanie usług PivotTable do przeglądania kostek OLAP

Usługi PivotTable zapewniają połączenie między hurtownią danych (serwerem OLAP) i wszystkimi aplikacjami zewnętrznymi, które się z nią komunikują. Mogą to być aplikacje C++, aplikacje Visual Basic, ale także Excel lub komponenty Web.

Przykładowo Excel 2000 umożliwia tworzenie kostek po stronie klienta na drodze bezpośredniego dostępu do hurtowni danych. W związku z czym umożliwia on prace w arkuszu kalkulacyjnym bezpośrednio za pomocą miar i wymiarów. Kostką, podobnie jak w przeglądarce danych na serwerze MS SQL, można dowolnie manipulować, zagłębiając się w poszczególne wymiary kostki, filtrując dane w odpowiedni sposób lub zmieniając wymiary w kolumnach i wierszach tabeli. Ponadto komponenty Web Office 2000 pozwalają na tworzenie stron WWW z arkuszy Excel i w ten sposób umożliwiają zdalne przeglądanie rezultatów analizy.

4. Podsumowanie

Jak wynika z rozważań, system hurtowni danych jest czymś znacznie więcej niż prostym narzędziem do raportowania. Jest to środowisko zbierające dane ze wielu systemów działających w organizacji. W hurtowni przechowywane są dane historyczne, charakteryzujące się wysoką jakością i zoptymalizowane pod kątem dostarczenia ich w jak najkrótszym czasie. Takie informacje stanowią doskonałe źródło wiedzy na temat działania organizacji oraz wiedzy, która umożliwia podejmowanie właściwych decyzji, zapewnia monitorowanie i kontrolę realizowanych przez organizację funkcji.

W części praktycznej pracy przedstawiono projekt hurtowni danych na przykładzie zastosowania w firmie produkcyjno-handlowej oraz implementację tego projektu w środowisku Microsoft SQL Server 2000. Zrealizowane zostały wszystkie wymagane etapy budowy hurtowni danych oraz przedstawiono przykładowe możliwości analityczne systemu.

Jak można wywnioskować z tej części pracy, jedną z najtrudniejszych kwestii w tworzeniu hurtowni danych jest integracja danych z wielu potencjalnych źródeł. System umożliwia analizę informacji pochodzących z działu sprzedaży, produkcji i zakupów, które są zapisane w postaci plików tekstowych oraz arkuszy kalkulacyjnych. Jednakże dane wejściowe wymagają intensywnej manipulacji w celu ich użycia w hurtowni danych. W związku z czym, aby przygotować odpowiednio informacje zostały napisane skrypty VBA programu Excel.

Utworzona hurtownia danych oparta została o schemat logiczny - wielogwieździsty, co może niekorzystnie wpłynęło na rozmiar danych, jednakże zwiększyło wydajność i prędkość dostępu do danych w systemie. Zbudowane kostki zostały zapisane jako MOLAP – strukturę hybrydową modelu relacyjno - wielowymiarowego. Umożliwiło to umieszczenie centralnego magazynu w strukturze relacyjnej a stworzone agregacje w wielowymiarowej bazie danych, co spowodowało zmniejszenie transferu danych ze źródeł.

System spełnia wymagania OLAP przedsiębiorstwa dając możliwość analizy klientów i dostawców oraz sprzedawanych i produkowanych artykułów. Umożliwia także kalkulację opłacalności produkcji wybranych artykułów. Opisany projekt może zatem stanowić doskonałe źródło danych dla aplikacji klienckich oraz interfejsów systemów wspomagania decyzji, kierujących zapytania do utworzonej hurtowni danych i oferujących intuicyjną obsługę systemu.

Literatura

1. Ostrowska T.: Wielowymiarowe bazy danych w procesach podejmowania decyzji. Materiały II Konferencji Naukowej pt. Komputerowo Zintegrowane Zarządzanie, Red. Knosala R.: „Komputerowo zintegrowane zarządzanie”, WNT, Warszawa, 1999.
2. Poe V., Klauer P., Brobst S.: Tworzenie hurtowni danych. Wydawnictwo Naukowo-Techniczne, Warszawa, 2000.
3. Gorawski M.: Hurtownia danych, Informatyka, 3,2000, str.30-33.
4. Głowacki G.: Mechanizmy wspierające tworzenie i użytkowanie Hurtowni Danych w systemie zarządzania bazą danych Oracle. Materiały konferencyjne: VII Konferencja PLOUG, Zakopane, 2001.
5. Romanowski D., Szczerbinki E.: Magazyny danych ROLAP jako źródło systemów wspomaganie decyzji. Dlaczego systemy OLTP nie wystarczają? Materiały VII Konferencji Naukowej pt. Komputerowo Zintegrowane Zarządzanie, Red. Knosala R.: Komputerowo zintegrowane zarządzanie. WNT, Warszawa, 2004.
6. Cannolly T.: Systemy baz danych: praktyczna metoda projektowania. Uniwersytet Wrocławski, Wrocław, 2004.
7. Rankins R., Jensen P., Bertucci P.: Microsoft SQL Server 2000. Księga eksperta. Helion, Warszawa, 2003.
8. Sturm J.: Hurtownie danych Microsoft SQL Server 7.0. Przewodnik Techniczny. APN Promise Sp. Z o.o., Warszawa, 2000.
9. Przewodnik po Microsoft SQL Server 7.0, <http://www.microsoft.com/poland/sql/70/dodatki/default.msp>.
10. Whalen E., Gaarcia M., Deluca S., Thompson D.: Microsoft SQL Server 2000. Konfiguracja i optymalizacja. Opis techniczny. Wyd. I, Read Me Oficyna Wydawnicza, Warszawa, 2002.
11. Rabeler C., WickertMicrosoft D.: SQL Server 2000 Analysis Services Operations Guide, 2003.
12. Lane P.: Oracle9i Data Warehousing Guide, Release 2 (9.2), Oracle Corporation, 2002.
13. Morzy T.: Systemy informatyczne. Przetwarzanie danych w magazynach danych. Materiały konferencyjne: V Seminarium PLOUG – Systemy informatyczne. Projektowanie oraz implementacja magazynów (hurtowni) danych. Warszawa, 2002, http://www.ploug.org.pl/seminarium/seminarium_V/pliki/t_morzy.pdf.
14. Oracle 9i OLAP Users Guide, Release 2 (9.2.0.2), Oracle Corporation, 2002.

Prof. dr hab. inż. Ryszard KNOSALA

Mgr inż. Katarzyna BŁASZCZYK

Instytut Inżynierii Produkcji

Politechnika Opolska

45-370 Opole, ul. Ozimska 75

tel.: (0-77) 423-40-35

e-mail: blaszczyk@po.opole.pl

knosala@po.opole.pl