

Katarzyna HAREŹŁAK, Stanisław KOZIELSKI, Sebastian SMERD
Politechnika Śląska, Instytut Informatyki

APLIKACJE ANALITYCZNE DLA HURTOWNI DANYCH DZIEKANATU

Streszczenie. W artykule przedstawiono aplikacje analityczne zbudowane przy wykorzystaniu Oracle dla hurtowni danych dziekanatu. Zrealizowano dwa warianty budowy aplikacji, odpowiadające dwu różnym architekturom hurtowni danych: ROLAP i MOLAP. Opracowane aplikacje wyznaczają statystyki dotyczące liczby studiujących, wyników zaliczeń i egzaminów oraz kwot wypłacanych stypendiów.

OLAP APPLICATIONS FOR UNIVERSITY DATA WAREHOUSE

Summary. OLAP applications, built for the university data warehouse are presented in the paper. Two applications were built accordingly to the two-level data warehouse architecture: ROLAP and MOLAP. These applications allow perform histograms relative to number of students, exam's results and sums of scholarships.

1. Wstęp

Aplikacje analityczne – to aplikacje przeznaczone dla analityków i decydentów przedsiębiorstwa lub organizacji, współpracujące z serwerem OLAP i pozwalające na efektywne przeglądanie oraz analizowanie danych zawartych w hurtowni danych. Stanowią one element systemu OLAP, będącego istotnym narzędziem przedsiębiorstwa, pomagającym w ocenie jego działalności. Obok aplikacji należą do niego: serwer OLAP, hurtownia danych oraz programy ekstrakcji i ładowania danych do hurtowni.

Aby system taki mógł dostarczać użytecznych i szczegółowych informacji na temat procesu gospodarczego w przedsiębiorstwie, musi spełniać następujące wymagania:

- informacja zawarta w systemie musi mieć zastosowanie w rozwiązywaniu problemów przedsiębiorstwa lub być w inny sposób użyteczna dla niego, sama ilość zebranej informacji nie gwarantuje jej użyteczności,
- informację musi cechować punktualność – informacja użyteczna ze względu na treść, ale uzyskana zbyt późno, może być bezwartościowa,
- system musi zapewniać łatwe porównanie informacji na podstawie wspólnych cech,
- system musi umożliwiać analizę danych na wiele sposobów; do typowych operacji w aplikacjach OLAP należą:
 - wyznaczanie punktu centralnego (pivoting) – polegające na wskazaniu miary, która nas interesuje i wymiarów, w których ma być prezentowana,
 - rozwijanie (drill-down), zwijanie (roll-up lub drill-up) – polegające na nawigacji wzdłuż hierarchii danego wymiaru, aby rozbić aktualnie prezentowane agregaty na składowe przypisane do atrybutów danego wymiaru na niższym poziomie hierarchii. Operacja zwijania działa odwrotnie do rozwijania, następuje przesuwanie się w górę po hierarchii w celu uzyskania większego stopnia agregacji danych,
 - obracanie (rotating),
 - wycinanie (slice and dice) – polegające na logicznym redukowaniu liczby wymiarów i dokonywaniu projekcji danych w pozostałych wymiarach, (przetwarzaniu podlegają tylko te dane, które leżą na zadanych współrzędnych zredukowanego wymiaru),
 - wyznaczanie rankingu – polegające na układaniu elementów wymiaru zgodnie z malejącymi lub rosnącymi wartościami agregatów,
- informacja winna być prezentowana w języku charakterystycznym dla jej odbiorcy, a nie w języku systemów analitycznych.

2. Etapy tworzenia systemu analitycznego wspomagającego pracę dziekanatu

Przygotowanie systemu analitycznego składa się z następujących kroków:

1. projektowanie struktury danych dla hurtowni danych,
2. projektowanie przekształceń danych w celu przeniesienia danych historycznych oraz bieżących baz danych OLTP do hurtowni danych,
3. projektowanie aplikacji do przeglądania danych w hurtowni,
4. stworzenie hurtowni danych i przeniesienie danych,
5. budowa aplikacji.

W artykule tym przedstawione zostaną zagadnienia związane z projektowaniem i tworzeniem aplikacji analitycznych dla hurtowni danych dziekanatu. Natomiast opis dwóch pierwszych etapów w odniesieniu do systemu danych dziekanatu zawarty został w pracy [6].

W pierwszej fazie prac określone zostały cele, jakie aplikacje te muszą spełniać. Typowe potrzeby danych dziekanatu, adresowane do hurtowni danych i aplikacji analitycznych, to m.in. wyznaczanie następujących statystyk:

- liczba studentów studiujących na danym wydziale,
- liczba studentów, którzy zaliczyli/nie zaliczyli dany semestr,
- liczba studentów, którzy otrzymali wpis warunkowy na dany semestr,
- sumaryczne kwoty stypendiów określonego rodzaju wypłacane studentom w zadanym okresie.

Dla wszystkich powyższych statystyk powinna istnieć możliwość prezentacji w rozbiciu kolejno na kierunki studiów, lata studiów, lata akademickie, semestry itp.

Inna grupa statystyk interesujących dziekanat to wyniki zaliczeń i egzaminów, w tym:

- średnie ocen uzyskane przez poszczególnych studentów w zadanych okresach czasu,
- listy rankingowe studentów na podstawie średnich ocen w zadanych przedziałach czasu w rozbiciu na kierunki i lata studiów,
- średnie ocen z poszczególnych przedmiotów w rozbiciu na lata akademickie i semestry.

3. Narzędzia do tworzenia aplikacji OLAP

Do budowy hurtowni danych dla dziekanatu zaproponowano dwupoziomą hurtownię danych [6]. Poziom podstawowy stanowi hurtownia danych o modelu ROLAP gromadząca tylko dane szczegółowe. Hurtownia ta została zbudowana przy wykorzystaniu serwera Oracle8. Wyższy poziom tworzy hurtownia o modelu MOLAP mieszcząca tylko dane zagregowane, pamiętane w wielowymiarowej bazie danych Personal Express'a.

Dla tak przyjętego modelu zrealizowano dwa warianty budowy aplikacji, odpowiadające dwu różnym architekturom hurtowni danych, wykorzystując odpowiednio narzędzia Oracle Discoverer oraz Oracle Objects i Express Analyzer. Poniżej zamieszczono krótką charakterystykę każdego z tych narzędzi.

3.1. Oracle Express

Narzędziem firmy Oracle służącym do tworzenia i rozwijania aplikacji OLAP w architekturze klient/serwer jest pakiet Oracle Express Objects. Umożliwia on tworzenie

aplikacji w trybie graficznym (metodą drag and drop), jak również za pomocą narzędzi programistycznych. Dzięki Oracle Express Analyzer, narzędziu służącemu do analizy i raportowania, użytkownik końcowy może w prosty sposób uruchamiać, modyfikować i oglądać aplikacje stworzone za pomocą pakietu Express Objects. Posiada on wbudowane mechanizmy tworzenia tablic i wykresów. Twórca aplikacji może w prosty sposób obracać osie raportu, dodawać rozwijane listy z pozycjami wymiarów oraz używać hierarchii dla lepszego uporządkowania wyświetlanych danych. Do wyboru wartości prezentowanych w raportach i na wykresach służy narzędzie zwane Selektorem. Wykorzystując mechanizmy do przeszukiwania według nazwy, poziomu w hierarchii, kryteriów wyjątków – Selektor umożliwia tworzenie podzbiorów wartości wyświetlanych na ekranie.

Zarówno Express Objects, jak i Express Analyzer mogą pracować w różnych konfiguracjach:

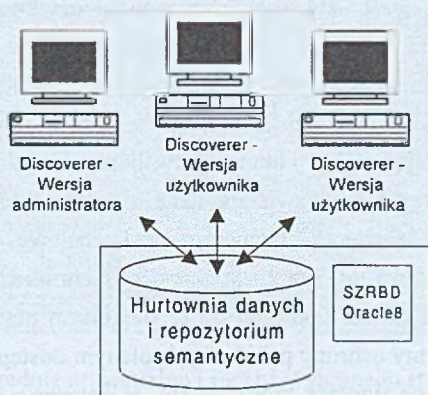
- w oparciu o Personal Express – serwer pracujący na pojedynczych komputerach klasy PC i w sieciach LAN.
- w oparciu o Oracle Express Server – serwer pracujący na komputerach z systemem UNIX, mainframe'ach i serwerach Windows NT.

3.2. Narzędzia Oracle Discoverer

Serwer relacyjnej bazy danych Oracle8 oraz narzędzie Oracle Discoverer stanowią połączenie, z pomocą którego można zbudować relacyjną hurtownię danych i aplikacje realizujące zadania analizy danych. Rysunek 1 pokazuje typową konfigurację omawianych narzędzi.

Oracle Discoverer składa się z dwóch części (rys. 1): wersji użytkownika (ang. *User Edition*) – przeznaczonej dla użytkowników końcowych – oraz wersji administratora (ang. *Administration Edition*) – przeznaczonej dla administratorów. Wersja użytkownika wspomaga interaktywne, wydajne i szybkie przeszukiwanie danych, uszczegóławianie, analizę wielowymiarową i tworzenie wykresów. Wersja administratora umożliwia tworzenie i modyfikację składowanego w bazie danych repozytorium opisu semantycznego. W repozytorium przechowywany jest czytelny dla użytkowników końcowych opis struktur i związków danych przeglądanych za pomocą obiektów wersji użytkownika. Discoverer pracuje w architekturze klient-serwer i zapewnia typową dla systemów scentralizowanych skalowalność oraz ochronę danych przed niepożądanym dostępem. Użytkownik posługuje się wygodnym w obsłudze graficznym interfejsem. Wybiera interesujące go informacje, zaś generator zapytań tworzy i przekazuje do bazy danych odpowiednio zoptymalizowane zapytania. Wbudowany samowstrząs mechanizm przewidywania czasu wykonania zapytań ułatwia użytkownikowi wybranie właściwego zapytania. W czasie gdy zapytanie jest wykonywane, użytkownik może pracować z innymi zestawami danych. Może także przerwać

zapytanie trwające zbyt długo. Otrzymane wyniki są odczytywane z bazy danych partiami, co przyspiesza pracę i zmniejsza obciążenie sieci. Dane pobrane z bazy są przechowywane przez Oracle Discoverer w wydzielonym obszarze pamięci na stacji roboczej. Są one indeksowane i kompresowane tak, aby ułatwić wykonywanie – typowych dla analizy wielowymiarowej – operacji obracania i uszczegóławiania danych oraz zmniejszyć liczbę zapytań zadawanych do bazy. Oracle Discoverer umożliwia użytkownikom wielokierunkowe uszczegóławianie oraz wielowymiarową analizę danych pobranych z hurtowni. W trakcie analizy dane mogą być oglądane w różnych układach, bez konieczności wykonywania jakichkolwiek operacji w języku SQL i znajomości wewnętrznych struktur bazy danych.



Rys. 1. SZRBD Oracle8 i Discoverer – sposób współpracy

Fig. 1. The DBMS Oracle8 and Discoverer – the way of cooperation

Oracle Discoverer jest w pełni zintegrowany z bazą danych Oracle oraz z innymi narzędziami firmy Oracle. Raporty tworzone za pomocą Oracle Discoverer mogą być przenoszone automatycznie do Oracle Reports. Inne narzędzia mogą korzystać z informacji przechowywanych w warstwie użytkowej Oracle Discoverer poprzez otwarty interfejs Open Meta Layer API. Oracle Discoverer jest zgodny ze standardem OLE2, co ułatwia wymianę danych z aplikacjami biurowymi, takimi jak Microsoft Word czy Microsoft Excel. Możliwy jest także eksport danych w różnych formatach oraz dostęp do danych pochodzących z innych niż Oracle baz danych zgodnych ze standardem ANSI SQL92.

3.2.1. Oracle Discoverer – wersja użytkownika

Oracle Discoverer jest narzędziem do interaktywnego przeglądania baz danych, raportowania i analizy z możliwością tworzenia wykresów ułatwiających odkrywanie zależności i wyjątków ukrytych w przechowywanych danych. Oracle Discoverer umożliwia też przeglądanie danych, takich jak fragmenty wideo czy fonii, pliki tekstowe, dokumenty

i arkusze kalkulacyjne. Przeglądane dane mogą być przechowywane w bazie danych lub w oddzielnych plikach systemu operacyjnego.

Uzyskane za pomocą Discoverer'a wyniki można zapisywać w formatach html i pdf. Dzięki temu mogą być one udostępniane odbiorcom korzystającym z przeglądarek internetowych, takich jak Netscape Navigator czy Microsoft Explorer. Oprócz tego dane mogą być zapisywane w innych formatach, na przykład w plikach tekstowych lub w postaci arkuszy kalkulacyjnych.

3.2.2. Oracle Discoverer – wersja administratora

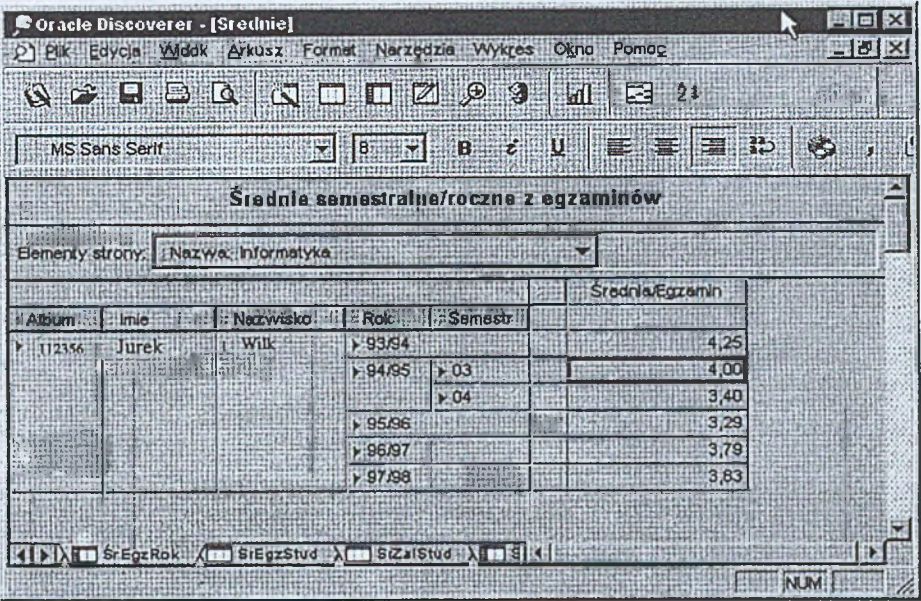
Zadaniem administratora jest stworzenie warstwy użytkowej – repozytorium semantycznego, które zawiera definicje obszarów tematycznych i folderów. Obszary tematyczne grupują logicznie powiązane ze sobą foldery. Foldery podstawowe, to udostępniane użytkownikowi perspektywy opisujące zawartość tablic w bazie danych. Elementy folderów odpowiadają zwykle kolumnom w bazie danych. Administrator może tworzyć także foldery złożone, które zawierają dane z wielu folderów podstawowych. Foldery mogą dodatkowo zawierać elementy wyliczeniowe, definiowane przez administratora. Warstwa użytkowa jest przechowywana w scentralizowany sposób w bazie danych Oracle Universal Data Server. Dostępu do repozytorium oraz do danych bronią – typowe dla Oracle – mechanizmy ochrony przed niepowołanym dostępem. Budowę warstwy użytkownika końcowego ułatwia automat, który tworzy jej wstępną wersję, w szczególności domyślnie nazwy i opisy folderów, nazwy i opisy elementów, listy wartości, definicje złączeń oraz operacji uszczegóławiania. Tak utworzona warstwa użytkowa może być następnie dowolnie zmieniana przez administratora. Warstwę użytkową można także utworzyć automatycznie w oparciu o zawartość repozytorium Designer/2000.

4. Przykłady wyników wybranych analiz

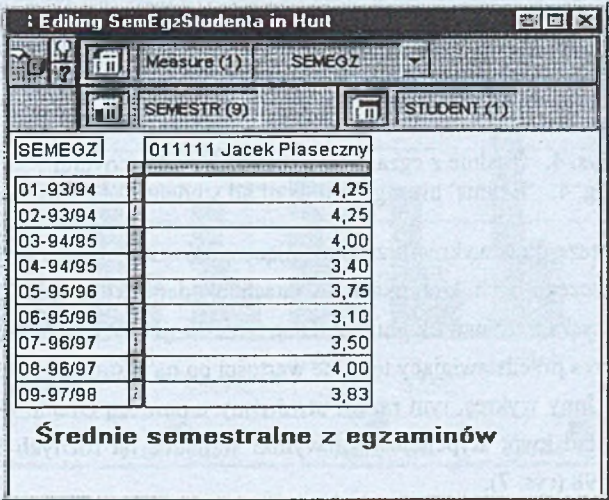
W tym rozdziale przedstawiono przykładowe elementy aplikacji analitycznych zrealizowanych za pomocą narzędzi Oracle Object i Oracle Discoverer oraz wnioski dotyczące wykorzystania tych narzędzi.

Rysunek 2 przedstawia raport utworzony z pomocą Discoverer'a prezentujący średnie semestralne i roczne wybranego studenta. Przechodzenie od średnich rocznych do średnich semestralnych i odwrotnie można realizować dzięki rozwijaniu i zwijaniu (przykładem jest rozwinięty rok akademicki 94/95, w którym widoczne są średnie w trzecim i czwartym semestrze). Z pomocą Oracle Objects można uzyskać podobne raporty, z tą różnicą że przy przyjętym w [6] modelu danych bazy wielowymiarowej nie jest możliwe drażnienie od średnich rocznych do średnich semestralnych. Aby uzyskać obydwie omawiane średnie,

należy utworzyć dwa oddzielne raporty – rysunek 3 przedstawia przykładowe średnie semestralne wybranego studenta.

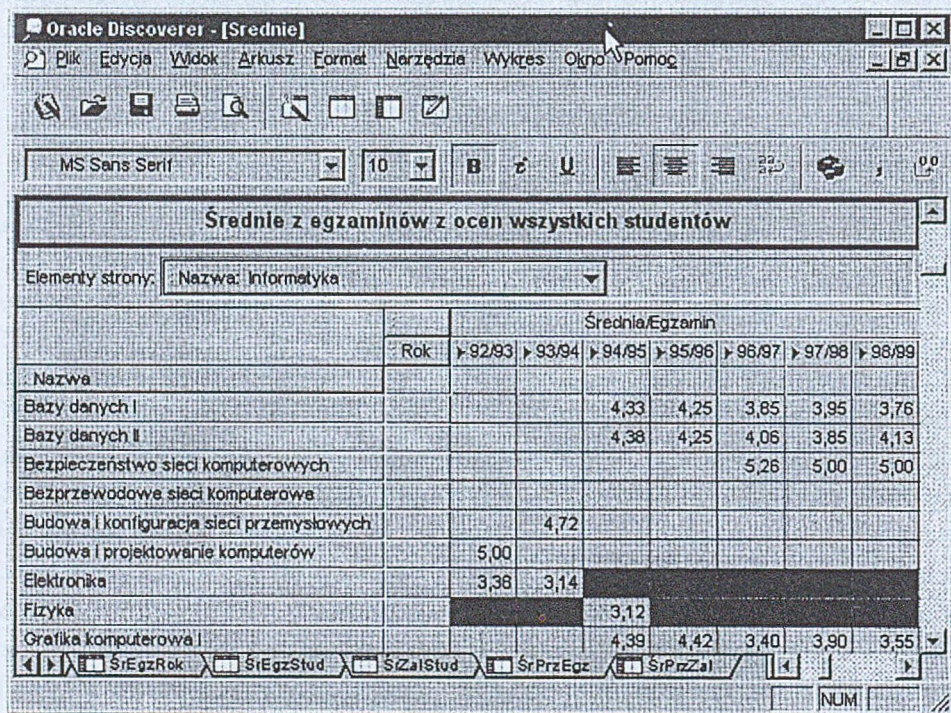


Rys. 2. Średnie semestralne i roczne wybranego studenta – Discoverer
Fig. 2. Average grades of the chosen student – Discoverer



Rys. 3. Średnie semestralne wybranego studenta – Oracle Objects
Fig. 3. Average grades of the chosen student – Oracle Objects

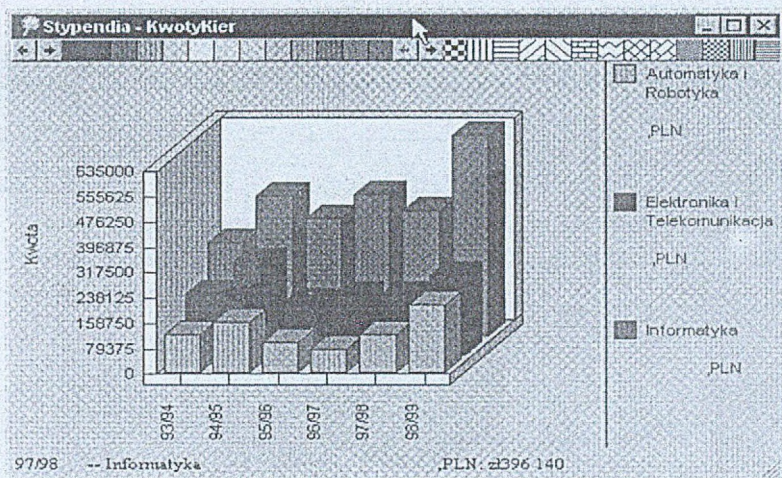
Przy wykorzystaniu Discoverer'a utworzony został raport prezentujący średnie z wszystkich ocen uzyskanych przez studentów z egzaminów z danego przedmiotu w danym roku akademickim. W raporcie zdefiniowany został wyjątek ukazujący średnie mniejsze od oceny dostatecznej – oznaczone ciemniejszym kolorem (rys.4).



Rys. 4. Średnie z egzaminów studentów – Discoverer

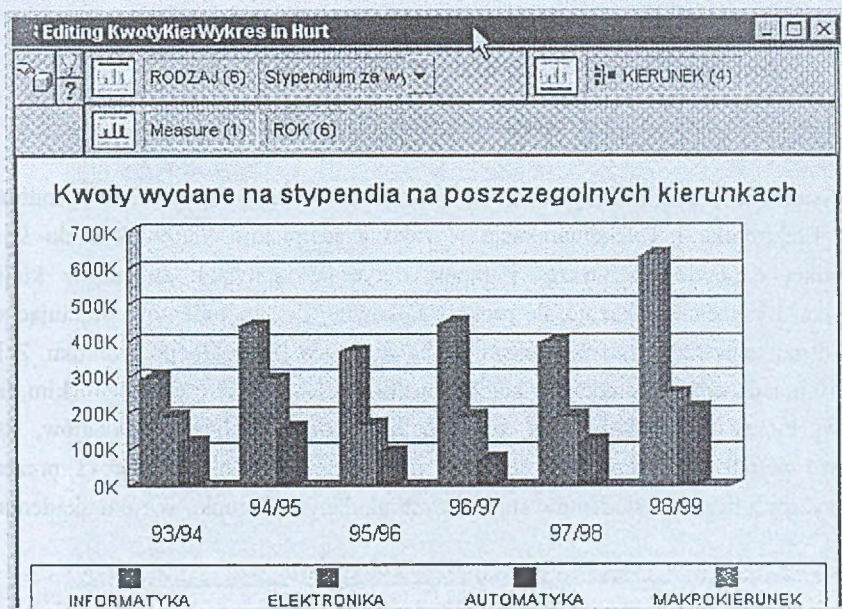
Fig. 4. Exams' average results of all students – Discoverer

Rysunek 5 prezentuje wykres przedstawiający porównanie kwot stypendiów za wyniki w nauce na poszczególnych kierunkach w latach akademickich. Taki wykres oraz wiele innych można uzyskać zarówno z pomocą Discoverer'a, jak i Oracle Objects (na rysunku 6 znajduje się wykres przedstawiający te same wartości co na rysunku 5, utworzony za pomocą Oracle Objects). Inny wykres, tym razem utworzony z pomocą Oracle Objects, przedstawia procentowy rozkład kwot stypendiów za wyniki w nauce na różnych kierunkach w roku akademickim 97/98 (rys. 7).



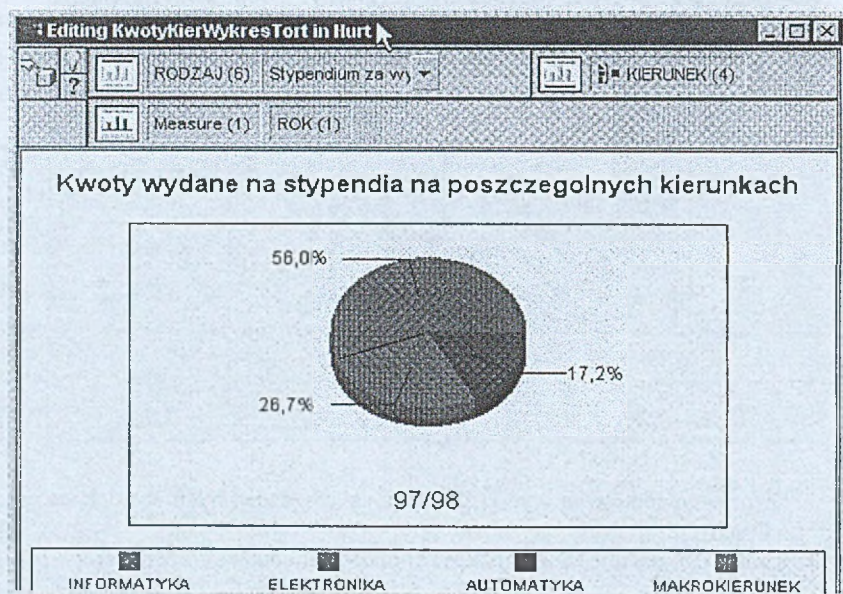
Rys. 5. Kwoty stypendiów za wyniki w nauce na poszczególnych kierunkach – Discoverer

Fig. 5. Knowledge obtained scholarships for particular courses – Discoverer



Rys. 6. Kwoty stypendiów za wyniki w nauce na poszczególnych kierunkach – Oracle Objects

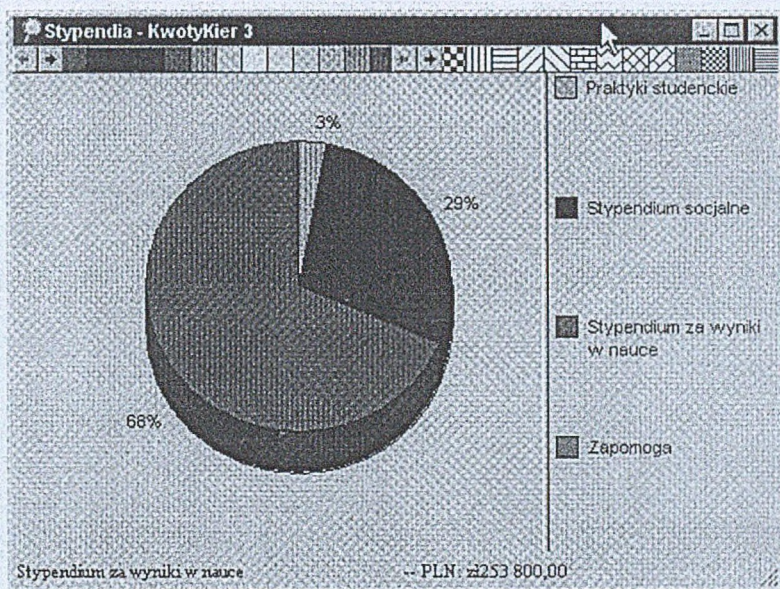
Fig. 6. Knowledge obtained scholarships for particular courses – Oracle Objects



Rys. 7. Procentowy rozkład kwot stypendiów za wyniki w nauce na różnych kierunkach w roku akademickim 97/98 – Oracle Objects

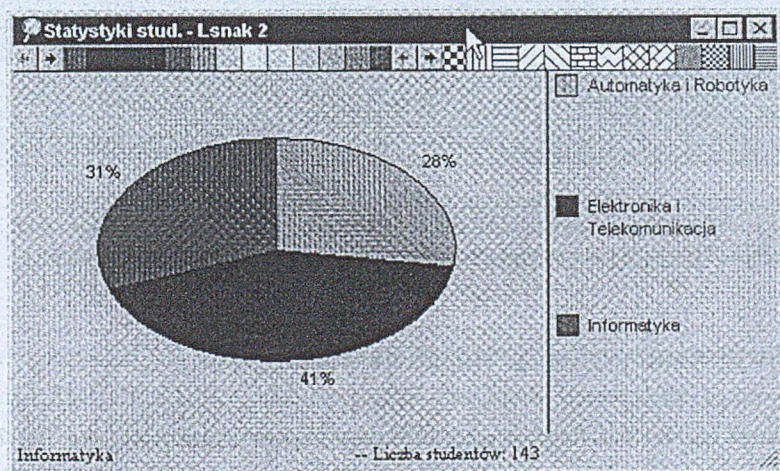
Fig. 7. Percentage distribution of scholarships for particular courses in academic year 97/98 – Oracle Objects

Rysunek 8 przedstawia procentowy rozkład kwot poszczególnych stypendiów na kierunku Elektronika i Telekomunikacja w roku akademickim 98/99. Rysunki 9 i 10 przedstawiają odpowiednio wykres i raport dotyczący statystyk studentów kierunku Informatyka. Rysunek 9 pokazuje, ile procent całkowitej liczby studentów, studiujących w roku 98/99 na semestrze trzecim, stanowi liczba studentów konkretnego kierunku. Z kolei rysunek 10 pokazuje liczbę studentów studiujących w danym semestrze akademickim, liczbę studentów, którzy zaliczyli lub nie zaliczyli semestru oraz liczbę studentów, którzy warunkowo zostali dopuszczeni do semestru następnego. Ostatni rysunek 11 prezentuje raport i wykres z liczbami studentów studiujących na danym kierunku w roku akademickim 98/99.



Rys. 8. Kwoty stypendiów na kierunku Elektronika i Telekomunikacja – Discoverer

Fig. 8. Scholarship amounts for Electronics and Telecommunication – Discoverer



Rys. 9. Procentowy skład liczby studentów studiujących w roku 98/99 na semestrze trzecim dla poszczególnych kierunków – Discoverer

Fig. 9. Percentage composition of number of students studying on the third semester in academic year 98/99 for particular courses – Discoverer

Statystyki studentów na kierunkach

Elementy strony: Nazwa: Informatyka

Rok	Semestr	Liczba studentów	Zaliczyło	Nie zaliczyło	Otrzymało warunk
95/96	01	128	128	0	0
	02	129	128	1	0
	03	132	129	0	3
	04	129	128	1	0
	05	112	111	0	1
	06	109	102	0	7
	07	126	126	0	0
	08	124	124	0	0
	09	111	111	0	0
	10	89	89	0	0
96/97	01	141	141	0	0
	02	139	137	0	1
	03	125	120	0	5
	04	124	121	0	3
	05	130	129	0	1
	06	124	119	0	5
	07	111	110	1	0

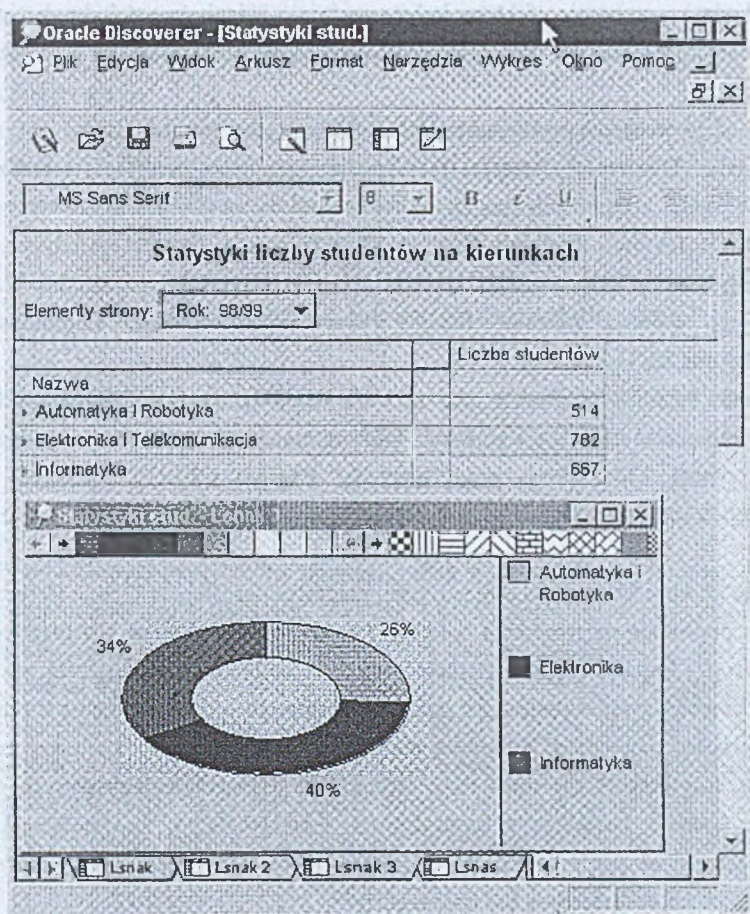
Rys. 10. Statystyki dotyczące studentów kierunku Informatyka – Discoverer

Fig. 10. Statistics concerning students of Computer Science – Discoverer

5. Porównanie wykorzystanych narzędzi

Założone na wstępie funkcje projektowanego systemu analitycznego dla dziekanatu zrealizowano w dwu alternatywnych postaciach, korzystając z narzędzi Oracle Express Objects oraz Discoverer. Porównując oba narzędzia, należy podkreślić, że zakres ich użycia oraz ich możliwości nie są identyczne.

Discoverer może być stosowany tylko do analizy danych w relacyjnych bazach danych oraz hurtowniach relacyjnych (typu ROLAP). Natomiast Express Objects i Express Analyzer



Rys. 11. Raport i wykres prezentujący liczby studentów różnych kierunków studiujących w roku 98/99 – Discoverer

Fig. 11. Report and graph of number of students studying in academic year in 98/99 for particular courses – Discoverer

służą przede wszystkim do analizy danych przechowywanych w wielowymiarowej bazie danych (MOLAP). Dodatkowo posiadają także możliwość współpracy z hurtowniami relacyjnymi (ROLAP), którą zapewnia im moduł RAM (ang. Relational Access Manager).

Możliwość zastosowania obu tych narzędzi w systemie prezentowanym w niniejszej pracy wynika z przyjętej, dwupoziomowej organizacji hurtowni danych. W takiej konfiguracji Discoverer może być wykorzystywany tylko na niższym poziomie, natomiast Oracle Express Objects na obu poziomach. Naturalnym efektem takiej organizacji hurtowni jest znacznie szybsze wykonywanie analiz dotyczących wartości zagregatyzowanych przy wykorzystaniu aplikacji Oracle Express Objects, korzystających z wielowymiarowej bazy

danych. W badanym systemie efekt ten był zauważalny, jednak ze względu na niezbyt duże ilości zgromadzonych danych (rzeczywistych) ostatecznie zrezygnowano z eksperymentów pomiarowych.

Oba narzędzia posiadają pewne ograniczenia funkcjonalne lub ilościowe.

W pakiecie Discoverer brakuje możliwości tworzenia rankingów analizowanych wartości. W celu osiągnięcia takich efektów należy wykonać sortowanie wartości prezentowanych faktów. Zarówno Express Objects, jak i Express Analyzer posiadają narzędzie Top/Bottom, które pozwala uzyskać ten efekt w sposób bezpośredni. Użytkownik może za jego pomocą określać liczbę wyświetlanych wartości najmniejszych lub największych oraz ich kombinacji.

Stosowanie Personal Express'a jako serwera wielowymiarowej bazy danych nakłada natomiast ograniczenie na liczbę wymiarów, którymi może być wymiarowana jedna zmienna. Liczba ta wynosi 32. Podczas pracy narzędzi Express Objects i Express Analyzer w konfiguracji z modułem RAM liczba wspomnianych wymiarów jest jeszcze mniejsza i wynosi 10.

Oceniając łatwość stosowania obu narzędzi, należy podkreślić, że Oracle Express Objects i Discoverer pozwalają na tworzenie aplikacji OLAP w architekturze klient/serwer metodą „drag and drop”. Express Objects posiada przy tym dodatkowe możliwości zastosowania bardziej złożonych narzędzi programistycznych. W oparciu o doświadczenia z pracy z tymi narzędziami można powiedzieć, że proces analizy i raportowania jest łatwiejszy przy użyciu Discoverer'a. Użytkownik nie musi znać szczegółów organizacji bazy danych ani języka SQL. Może bowiem korzystać z repozytorium semantycznego, zawierającego definicje obiektów stworzonych przez administratora. Dzięki temu końcowy użytkownik po krótkim przeszkoleniu może samodzielnie tworzyć raporty.

W przypadku narzędzi Express Objects i Express Analyzer sytuacja jest nieco inna. Ich wykorzystywanie jest bardziej złożone i wymaga specjalistycznej wiedzy. Tworzeniem aplikacji dla końcowych użytkowników powinien się więc zajmować profesjonalny programista.

LITERATURA

1. Kimball R.: The Data Warehouse Lifecycle Toolkit. J. Willey, 1998.
2. Codd E. F., Codd S. B., Salley C. T.: Providing OLAP (On-Line Analytical Processing) to User-Analyst.
3. Gorawski M., Koziątek A.: Data Warehouse: Analiza porównawcza środowisk. Software 2.0 10/99, s. 32.
4. Poe V., Klauer P., Brobst S.: Tworzenie hurtowni danych. WNT, Warszawa 2000.

5. Matysiak M.: Technologia OLAP. III Konferencja PLOUG, Zakopane 1997.
6. Hareźlak K., Kozielski S., Smerd S.: Projekt i implementacja hurtowni danych dziekanatu (w opracowaniu).
7. <http://www.oracle.com.pl/dokumenty/olapexprservtoenterprise.pdf>.
8. <http://www.oracle.com.pl/dokumenty/olapexprobjds.pdf>.
9. <http://www.oracle.com.pl/dokumenty/olapexpranalds.pdf>.

Recenzent: Dr hab. inż. Stanisław Wołek, prof. Pol. Rzeszowskiej

Wpłynęło do Redakcji 23 lutego 2001 r.

Abstract

OLAP applications, built for the university data warehouse are presented in the paper. Oracle programming tools used for building OLAP applications are discussed at the beginning: Discoverer used in ROLAP architecture and Oracle Objects together with Express Analyzer, used in MOLAP architecture. Two applications were built accordingly to the two-level data warehouse architecture: ROLAP and MOLAP. These applications allow perform histograms relative to numbers of students, exams' results and sums of scholarships. Finally both programming tools and both applications were compared with regard to their functionality, scalability, facility of application building and speed of task performing.