

VIII Konferencja PLOUG
Kościelisko
Październik 2002

Zastosowanie systemu Oracle do budowy systemu Magazynu Danych wspomagającego przeprowadzanie analiz giełdowych

Gerard Głowacki

Biuro Usług Komputerowych

e-mail: gerard_glowacki@poczta.onet.pl

Abstrakt

W referacie omawia się możliwość wykorzystania systemu Oracle do budowy systemu Magazynu Danych wspomagającego przeprowadzanie analiz giełdowych. Przedstawia się istotne elementy schematu pojęciowego bazy danych, jak również procedur przetwarzających dane. Wskazuje się na mechanizmy bazy danych Oracle wspomagające implementację założeń projektowych.

1. Wstęp

Rozważany system można zaliczyć do systemów klasy OLAP (ang. On Line Analytical Processing), a więc systemów wspomagających przetwarzanie danych w celu wsparcie procesu podejmowania decyzji wykorzystujących w architekturze nowy typ bazy danych zwany magazynem danych (ang. Data Warehouse).

W tym przypadku zadaniem systemu jest wspomaganie analiz giełdowych, a w szczególności umożliwienie łatwego dostępu do danych dotyczących wyceny akcji, danych o charakterze fundamentalnym, oraz danych dotyczących struktury akcjonariatu, a także przygotowanie danych o charakterze zagregowanym na potrzeby analizy portfelowej.

Najczęściej definicja magazynu danych jest przytaczana za publikacjami W. H.Immon: "Building the Data Warehouse, QED Tech. Pub. Group, 1992" gdzie przez magazyn danych określa się zorientowaną tematycznie, zintegrowaną, zmienną w czasie i trwałą kolekcję (bazę) danych zaprojektowaną i zaimplementowaną dla potrzeb wspomaganie podejmowania decyzji, w której dane odnoszą się do określonej chwili czasowej".

W literaturze przez Magazyn (Hurtownie) Danych, rozumie się system, których schemat pojęciowy bazy danych opiera się o strukturę gwiazdy, strukturę płątka śniegu, lub strukturę konstelacji faktów. W centrum takiej struktury znajdują się przechowywane dane numeryczne tabele faktów, które są powiązane kluczami obcymi z tabelami wymiarów.

W strukturze gwiazdy centralna encja opisuje miarę, która jest połączona z encjami wymiarów. W strukturze płątka śniegu mamy do czynienia z hierarchią wymiarów, natomiast w przypadku konstelacji faktów występuje zbiór encji faktów, który współdzieli zbiór encji wymiarów, najczęściej na różnym poziomie hierarchii tych wymiarów.

Rozważany w niniejszym artykule schemat można zakwalifikować do struktury konstelacji faktów, w której fakty dotyczące wyceny akcji, danych ze sprawozdań finansowych, emisji akcji, oraz zmian struktury akcjonariatu współdzielią zbiór encji wymiarów - przede wszystkim encje związane z czasem - na różnych poziomach hierarchii np. dnia, tygodnia, miesiąca, kwartału, oraz encje spółki (branży).

W przypadku zastosowania technologii Hurtowni Danych dla celów wspomaganie analiz giełdowych, a w szczególności analizy portfelowej istnieje konieczność wprowadzenia dodatkowych tabel przechowywujących dane zaagregowane(m.in. stopa zwrotu, średnia stopa zwrotu, kowariancja).

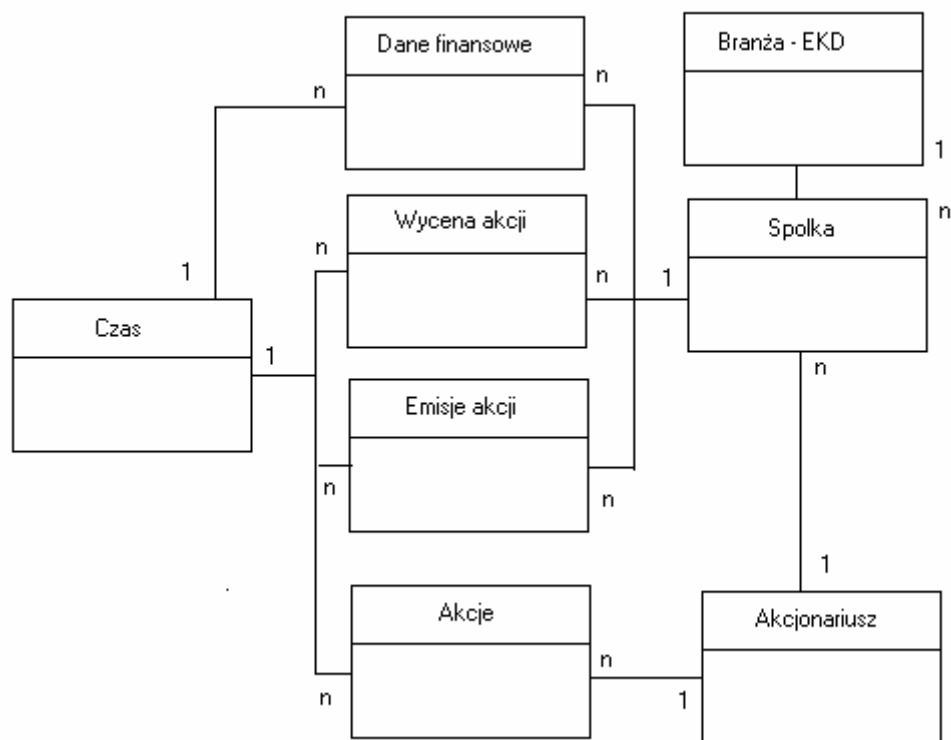
2. Ogólny schemat pojęciowy systemu wspomagającego przeprowadzanie analiz giełdowych

Jak już zasygnalizowano systemy Hurtowni Danych posiadają w centrum schematu pojęciowego tabele faktów. W rozważanym przypadku do tabel faktów zaliczyć można tabele :

- przechowujące dane odnośnie wyceny papierów wartościowych na GPW,
- dane ze sprawozdań finansowych - potrzebne do analizy fundamentalnej,
- dane o strukturze akcjonariatu, oraz emisjach akcji.

Istotnym elementem rozważanego systemu są omawiane w dalszej części artykułu, tabele przechowujące dane zaagregowane na potrzeby analizy portfelowej.

Ogólny schemat pojęciowy wygląda następująco :



Rys. 1. Ogólny schemat pojęciowy

Encja faktów <Dane Finansowe> przechowuje informacje ze sprawozdań finansowych, a więc atrybuty typowe dla bilansu, rachunku wyników, oraz rachunku przepływów pieniężnych z uwzględnieniem różnic i podobieństw występujących w przypadku banków, zakładów ubezpieczeń oraz innych przedsiębiorstw.

Encja faktów <Wycena Akcji> przechowuje informacje o kursach akcji oraz obrotach poszczególnych papierów wartościowych na giełdzie, jak również pewne podstawowe dane statystyczne np. kurs maksymalny i minimalny z poprzedzających 52 tygodni, kurs średni oraz średnia wartość obrotu z ostatnich 10 sesji.

Encja faktów <Emisje Akcji> przechowuje informacje o kolejnych emisjach, sumarycznej ilości akcji, prawach do głosu na walnym zgromadzeniu, oraz ilości akcji wprowadzonych do obrotu giełdowego.

Encja faktów <Akcje> przechowuje informacje o strukturze akcjonariatu, a więc dane o ilości posiadanych akcji, ilości głosów na walnym zgromadzeniu, oraz procentowym udziale w kapitale akcyjnym, oraz procentowej ilości głosów na walnym zgromadzeniu.

Encja wymiarów <Czas> przechowuje informacje potrzebne do określenia momentów czasowych dla zajścia poszczególnych zdarzeń (dzień, tydzień, miesiąc, kwartał, rok).

Encja wymiarów <Spółka> przechowuje podstawowe informacje o spółce notowanej na giełdzie, a więc nazwa, adres, dane rejestracyjne, dane o zarządzie.

Encja wymiarów <Akcjonariusz> przechowuje podstawowe informacje o akcjonariuszu, w niektórych przypadkach może być powiązana z encją <Spółka>.

Podczas implementacji takiego schematu pojęciowego można wykorzystać dostępne w systemie Oracle mechanizmy klucza obcego i referencji. Przykładowo komenda do utworzenia tabeli akcji :

```
CREATE TABLE Akcje
( SpolkaID          NUMBER,
  AkcjonariuszID   NUMBER,
  CzasID            NUMBER,
  Ilosc_akcji       NUMBER,
  Ilosc_glosow      NUMBER,
  Ilosc_akcji_proc   NUMBER,
  Ilosc_glosow_proc NUMBER,

  CONSTRAINT fk_SpolkaID FOREIGN KEY ( SpolkaID )
  REFERENCES Spolka ( Spolka_ID )

  CONSTRAINT fk_AkcjonariuszID FOREIGN KEY ( SpolkaID )
  REFERENCES Akcjonariusz ( Spolka_ID )

  CONSTRAINT fk_CzasID FOREIGN KEY ( CzasID )
  REFERENCES Czas ( Czas_ID ) );
```

W praktyce podczas projektowania takiego systemu należałoby rozważyć kwestie wprowadzania danych do systemu.

W przypadku danych odnośnie notowań i sprawozdań finansowych dane byłyby pobierane z plików tekstowych. Wymaga to utworzenia dodatkowych tabel na potrzeby weryfikacji poprawności i ewentualnej obróbki wprowadzanych danych. W przypadku danych dotyczących struktury akcjonariatu oraz emisji akcji konieczne byłoby utworzenie formatek umożliwiających wprowadzenie odpowiednich modyfikacji. Z punktu widzenia struktur bazy danych mogłoby się to wiązać z koniecznością utworzenia dodatkowych tabel lub perspektyw.

Istotnym problemem związanym z wprowadzaniem informacji do systemu jest przeliczanie danych w tabelach agregacyjnych. W dalszej części niniejszego artykułu prezentuje się przykładowe tabele agregacyjne przechowujące dane na potrzeby analizy portfelowej. W tym przypadku przeliczanie agregatów powinno być zsynchronizowane z wprowadzaniem danych dotyczących notowań.

Tabele agregacyjne - wraz z rozwojem systemu - mogą dotyczyć również danych wprowadzanych w sposób asynchroniczny, (a więc w szczególności zmian w strukturze akcjonariatu lub nowych emisji akcji), a także powiązania danych wprowadzanych asynchronicznie z danymi wprowadzanymi zgodnie ze znanym harmonogramem czasowym (np. zmiany w strukturze akcjonariatu w powiązaniu ze zmianami wyników finansowych na poziomie branży lub grupy kapitałowej). Przedstawienie sposobu rozwiązania problemu asynchronicznego uaktualniania tabel agregacyjnych przekracza objętościowo ramy tego opracowania.

3. Przykładowe analizy możliwe do wykonania na podstawie stworzonego ogólnego schematu pojęciowego

Powyżej opisany ogólny schemat pojęciowy stwarza możliwości do sporządzenia analiz kojarzących dane ze sprawozdań finansowych, wyceny akcji na giełdzie, oraz zmian w strukturze akcjonariatu. Stwarza to możliwości połączenia klasycznej analizy fundamentalnej, z analizą powiązań kapitałowych. Analiza powiązań kapitałowych wraz z predykcją zachowań dużych inwestorów może być komplementarnym - obok analizy technicznej i fundamentalnej - instrumentem pracy analityków giełdowych.

Poniżej zostaną przedstawione przykładowe zapytania :

- Jaka była zmienność struktury akcjonariatu Spółki X w okresie od T1 do T2 (np. w okresie dużych zmian w wycenie akcji na giełdzie) .

```
SELECT Spolka.Nazwa, Akcjonariusz.Nazwa, Akcje.Ilosc_akcji, Ak-
cje.Ilosc_akcji_proc, Akcje.Ilosc_glosow, Akcje.Ilosc_glosow_proc
FROM Spolka, Akcje, Akcjonariusz
WHERE Spolka.SpolkaID = Akcjonariusz.AkcjonariuszID
AND Spolka.Nazwa = 'Spółka X'
AND Akcje.Czas_ID = T1
AND Akcje.Czas_ID = T2
ORDER BY Czas_ID ;
```

- W jakich spółkach akcjonariuszem była spółka X w okresie T1, T2 (np. w okresie w którym ze sprawozdań finansowych wynika wysoki przychód z działalności finansowej) .

```
SELECT Spolka.Nazwa, Akcjonariusz.Nazwa, Akcje.Ilosc_akcji, Ak-
cje.Ilosc_akcji_proc, Akcje.Ilosc_glosow, Akcje.Ilosc_glosow_proc
FROM Spolka, Akcje, Akcjonariusz
WHERE Spolka.SpolkaID = Akcjonariusz.AkcjonariuszID
AND Akcjonariusz.Nazwa = 'Spółka X'
AND Akcje.Czas_ID = T1
AND Akcje.Czas_ID = T2
ORDER BY Czas_ID ;
```

- Porównanie wartości akcji spółek notowanych na giełdzie w momentach czasowych T1,T2 posiadanych przez spółkę X z wartością Finansowego Majątku Trwałego w sprawozdaniach finansowych w momencie T1. (Na tej podstawie można estymować przychód z działalności finansowej).

```
SELECT Spolka.Nazwa, Akcjonariusz.Nazwa, Akcje.Ilosc_akcji * WycenaAkcji.Cena
FROM Spolka, Akcje, WycenaAkcji, Akcjonariusz
WHERE Spolka.SpolkaID = Akcje.SpolkaID
AND WycenaAkcji.SpolkaID = Spolka.SpolkaID
AND WycenaAkcji.CzasID = Akcje.CzasID
AND Akcje.AkcjonariuszID = Akcjonariusz.AkcjonariuszID
AND Akcjonariusz.Nazwa = 'Spółka X'
AND Akcje.Czas_ID = T1
AND Akcje.Czas_ID = T2
ORDER BY Czas_ID ;
```

```
SELECT Spolka.Nazwa, Dane_Finansowe.Finansowy_Majatek_Trwaly
FROM Spolka, Dane_Finansowe
WHERE Spolka.SpolkaID = Dane_Finansowe.SpolkaID
AND Spolka.Nazwa = 'Spółka X'
AND Akcje.Czas_ID >= T1
ORDER BY Czas_ID ;
```

- Określenie średniego obrotu na sesji giełdowej akcjami spółki X w okresie od T1 do T2.

```
SELECT Spolka.Nazwa, Czas.Tydzien, AVG(WycenaAkcji.Obrot)
FROM Spolka, Czas, WycenaAkcji
WHERE Spolka.SpolkaID = WycenaAkcji.SpolkaID
AND Spolka.Nazwa = 'Spółka X'
AND Akcje.Czas_ID >= T1 AND Akcje.Czas_ID <= T2 ;
```

- Uzyskanie informacji o strukturze pasywów spółki giełdowej, w chwili T1.

```
SELECT Spolka.Nazwa, Dane_Finansowe.KapitalWlasny, Dane_Finansowe.Rezerwy,
Dane_Finansowe.Zobowiazania
FROM Spolka, Dane_Finansowe
WHERE Spolka.Nazwa = 'Spółka X'
AND Spolka.SpolkaID = Dane_Finansowe.SpolkaID
AND Dane_Finansowe.Czas_ID = T1 ;
```

Przedstawiony schemat pojęciowy można wykorzystać dla celów przeprowadzenia badań nad możliwością zastosowania algorytmów eksploracji danych (w szczególności w zakresie odkrywania wzorców sekwencji i podobieństw w szeregach czasowych) dla konstrukcji długoterminowej strategii inwestycyjnej. W dalszej części niniejszego artykułu zostaną przedstawione możliwości rozbudowy schematu pojęciowego na potrzeby analizy portfelowej.

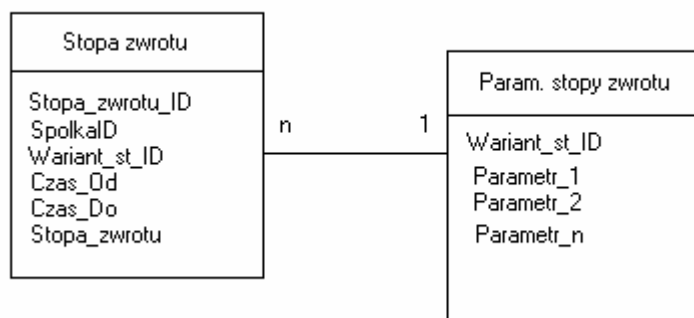
4. Rozbudowa schematu pojęciowego na potrzeby analizy portfelowej

Z punktu widzenia analizy portfelowej wyznaczenie portfela jest dwukryterialnym zadaniem decyzyjnym, w którym maksymalizujemy oczekiwaną stopę zwrotu oraz minimalizujemy ryzyko. Najczęściej oczekiwaną stopę zwrotu określa się jako średnią ze stóp zwrotu z przeszłości, natomiast ryzyko wyraża się za pomocą wariancji. Poniżej przedstawia się propozycję rozbudowy tworzonych modeli, tak aby istniała możliwość łatwego i parametryzowalnego wyznaczenia wartości niezbędnych dla analizy portfelowej, z wykorzystaniem dostępnych w systemie Oracle mechanizmów analitycznych - w szczególności funkcji analitycznych z partycjami obliczeniowymi, oraz ruchomym oknem obliczeniowym oraz funkcji statystycznych.

Aby umożliwić elastyczne wyznaczenie stóp zwrotu dla poszczególnych akcji konieczne będzie wprowadzenie oprócz tabeli przechowującej wyznaczone stopy zwrotu, wprowadzenie tabeli definiującej parametry obliczeniowe dla kolejnych wariantów. Do parametrów obliczeniowych należeć mogą w szczególności parametry związane okresem oraz sposobem przeprowadzanych obliczeń. W najprostszym ujęciu można rozważać dwa parametry :

- P1 - okres przeprowadzanych obliczeń np. tydzień, miesiąc, kwartał ,
- P2 - moment przeliczenia np. jeżeli parametr P1 przyjmuje wartość oznaczającą tydzień wówczas parametr P2 będzie przyjmował wartości 1..5, oznaczające dni tygodnia pomiędzy którymi przeprowadza się obliczenia .

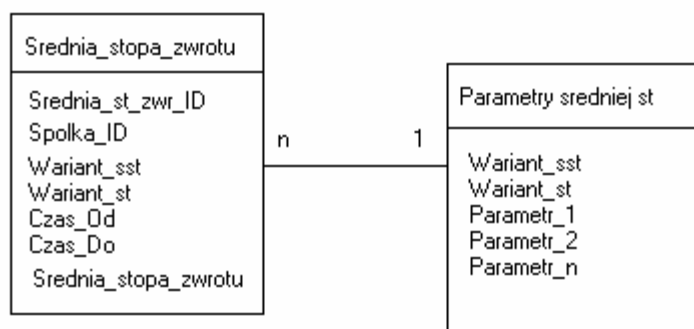
Encje przechowujące dane dotyczące stopy zwrotu oraz parametry obliczeniowe w najprostszym ujęciu można zaprojektować w sposób następujący :



Rys. 2. Encje wykorzystywane do wyznaczenia stóp zwrotu.

Encja <Stopa_zwrotu> zawiera klucze obce umożliwiające identyfikację spółki, okresu, oraz wariantu dla którego stopa zwrotu jest przeliczana. Do przeliczenia stóp zwrotu można wykorzystać funkcje wykorzystujące ruchome okno obliczeniowe, oraz partycje obliczeniowe.

Kolejnym etapem w procesie przygotowywania danych na potrzeby analizy portfelowej jest wyznaczenie średnich stóp zwrotu oraz macierzy kowariancji (w przypadku modelu Markowitza) oraz współczynników Beta (w przypadku modelu Sharpe'a). Encje potrzebne do wyznaczenia średniej stopy zwrotu mogą zostać zamodelowane następująco :



Rys. 3. Encje wykorzystywane do wyznaczenia średnich stóp zwrotu.

Encja <Średnia stopa zwrotu> zawiera klucze obce umożliwiające identyfikację spółki, wariantu obliczeniowego dla wyznaczenia stopy zwrotu, wariantu obliczeniowego dla wyznaczenia średniej stopy zwrotu, oraz okresu obliczeniowego.

Encja <Parametry średniej stopy> zawiera klucz obcy umożliwiający identyfikację wariantu obliczeniowego dla wyznaczenia stopy zwrotu oraz parametry określające wariant obliczeniowy dla wyznaczenia średniej. Do parametrów tych można zaliczyć sposób wyznaczenia średniej (arytmetyczna, geometryczna), a także inne parametry jak na przykład sposób określenia wag przywiązanych do stóp zwrotu z różnych okresów - tak aby w łatwy sposób istniała możliwość przywiązania większej wagi do stóp zwrotu z okresów bezpośrednio poprzedzających okres rozpoczęcia inwestycji, niż do stóp zwrotu uzyskiwanych w dawniejszych okresach.

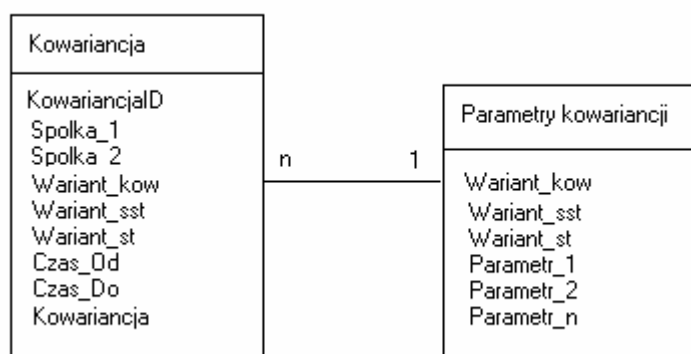
Do wyznaczenia średniej stopy zwrotu można wykorzystać funkcje z ruchomym oknem obliczeniowym, oraz partycje obliczeniowe. Przykładowo dla wyznaczenia średniej z ostatnich n notowań można posłużyć się następującą komendą:

```

INSERT INTO Srednia_Stopa_zwrotu
SELECT Spolka_ID, Czas_Od, Czas_Do,
AVG(Stopa_Zwrotu) OVER (partition by Spolka_ID order by Czas_do range between
n preceding and current row ) as Srednia_Stopa_n
FROM Stopa_zwrotu
ORDER BY Czas_Do

```

Wykorzystanie funkcji analitycznych i statystycznych dostępnych w systemie Oracle ułatwia pisanie kodu i umożliwia łatwą jego parametryzację, czyniąc go bardziej elastycznym. Analogicznie można zamodelować encje potrzebne do wyznaczenia kowariancji:



Rys.4. Encje wykorzystywane do wyznaczenia kowariancji

Posiadanie zintegrowanego źródła danych umożliwia wprowadzenie dodatkowych kryteriów optymalizacyjnych np. ograniczenie na procentową ilość udziału spółek w portfelu należących do tej samej grupy kapitałowej, jak również sformułowanie i rozwiązanie (oczywiście z wykorzystaniem odpowiednich narzędzi informatycznych) wielu wariantów zadań optymalizacyjnych.

Wszelkie potrzebne parametry do zadań optymalizacyjnych można w łatwy sposób uzyskać wykonując odpowiednie zapytanie do bazy danych. Przykładowo podaje się zapytanie umożliwiające uzyskanie wartości kowariancji oraz średniej stopy zwrotu.

- Wyznaczenie sumy kowariancji dla spółek pochodzących z WIG20, dla 2000 roku i odpowiednich wariantów obliczeniowych stopy zwrotu, średniej stopy zwrotu i kowariancji:

```

SELECT CZAS_DO, SPOLKA_1, SPOLKA_2, SUM(KOWARIANCJA) FROM KOWARIANCJA
WHERE SPOLKA_1 IN ( SELECT SPOLKA FROM WIG20) AND
SPOLKA_2 IN ( SELECT SPOLKA FROM WIG20 ) AND
CZAS_OD>200001 AND CZAS_DO <= 200012
AND WARIANT_KOW=1 AND WARIANT_SST=1 AND WARIANT_ST=2
GROUP BY CZAS_DO;

```

- Zestawienie średniej stopy zwrotu dla analogicznego okresu:

```

SELECT CZAS_DO, SPOLKA, SREDNIA FROM SREDNIA_STOPA_ZWROTU
WHERE SPOLKA IN (SELECT SPOLKA FROM WIG20)
AND CZAS_OD > 200001 AND CZAS_DO <= 200012
AND WARIANT_SST=1 AND WARIANT_ST=1;

```

5. Podsumowanie

Zastosowanie architektury Magazynu Danych z wykorzystaniem systemu Oracle do przeprowadzania analiz giełdowych może umożliwić przeprowadzanie bardziej złożonych i wielowariantowych analiz. Przedstawiony ogólny schemat pojęciowy wraz tabelami przechowującymi dane zagregowane, umożliwia łączenie różnych informacji (o charakterze fundamentalnym, technicznym, strukturze akcjonariatu) oraz elastyczne generowanie wielu możliwych wariantów obliczeniowych na potrzeby analizy portfelowej. Wykorzystanie systemu Oracle jest uzasadnione ze względu na możliwość wykorzystania dostępnych w tym systemie funkcji analitycznych, jak również stabilność systemu podczas przetwarzania dużych wolumenów informacji.

Bibliografia

1. Barker R. : Modelowanie związków encji, Wydawnictwa Naukowo Techniczne, Warszawa, 1996
2. Barker R., Longman C. : Modelowanie funkcji i procesów, Warszawa, 1996
3. Yourdon E. : Współczesna analiza strukturalna, Wydawnictwa Naukowo-Techniczne, Warszawa 1996
4. Dokumentacja systemowa w szczególności do DBMS Oracle.
5. Wrembel R., Jezierski J., Zakrzewicz M. : System zarządzania bazą danych Oracle7 i Oracle8, Wydawnictwo Nakom, Poznań 2000
6. Morzy T. : Przetwarzanie danych w magazynach danych. Stowarzyszenie Polskiej Grupy Użytkowników Systemu Oracle
7. Zakrzewicz M. : Implementacje przetwarzania analitycznego przy wykorzystaniu Oracle 9i OLAP. Stowarzyszenie Polskiej Grupy Użytkowników Systemu Oracle
8. Głowacki G. : Zastosowanie technologii Hurtowni Danych do badania powiązań kapitałowych, struktur akcjonariatu, oraz zmian cen akcji na giełdzie papierów wartościowych. Konferencja Modelowanie preferencji a ryzyko, Praca zbiorowa pod redakcją naukową Tadeusza Trzaskalika