

Wprowadzenie do technologii Business Intelligence i hurtowni danych

Plan rozdziału

- Wprowadzenie do Business Intelligence
- Hurtownie danych
- Produkty Oracle dla Business Intelligence

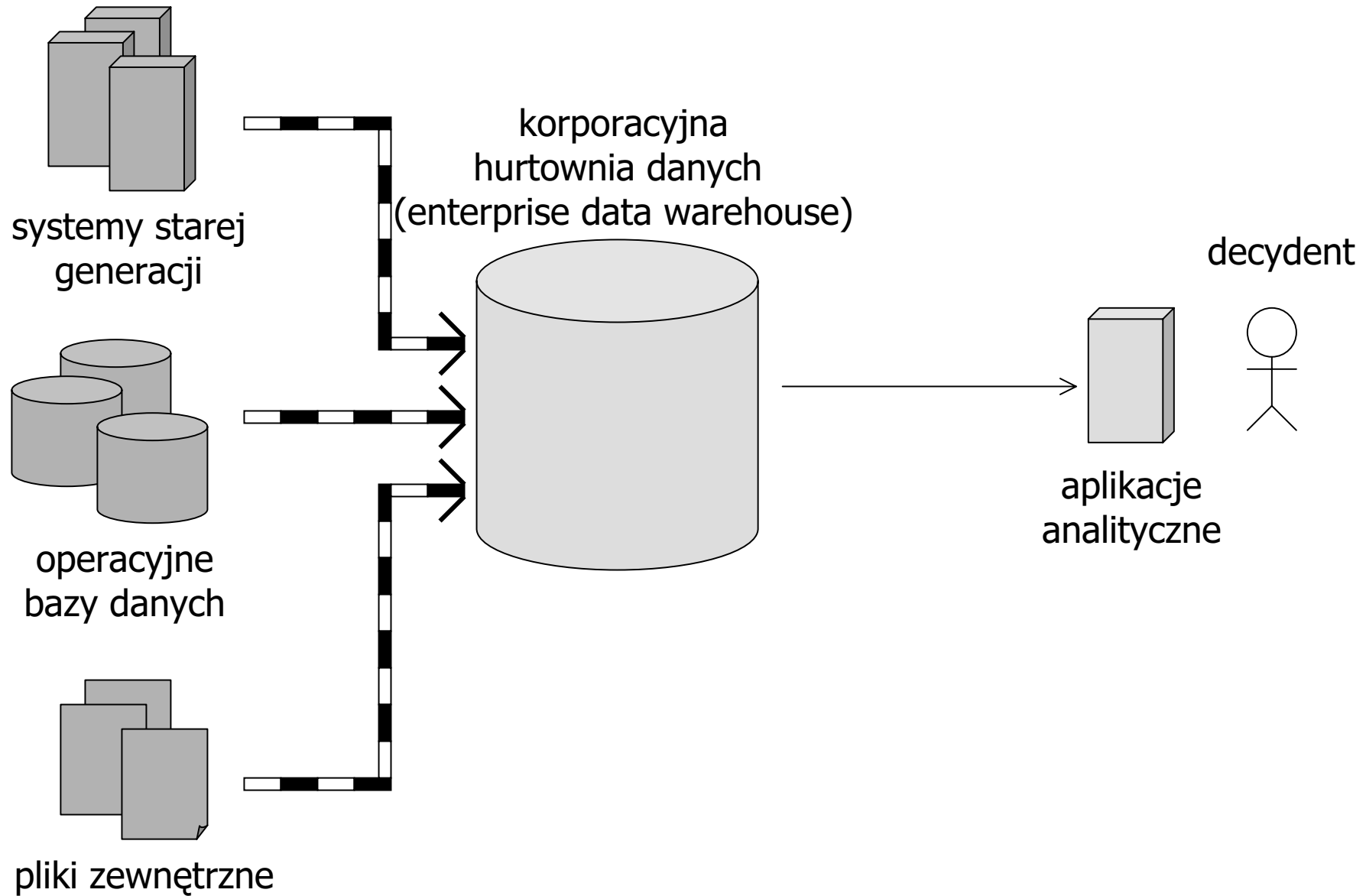
Business Intelligence

- Technologia informatyczna służąca przekształcaniu dużych wolumenów danych w informacje, a następnie przekształcaniu tych informacji w wiedzę
- Adresowana do pracowników szczebla kierowniczego, wspomagająca podejmowanie ich decyzji
- Stawiająca drastyczne wymagania wydajnościowe, przede wszystkim z powodu ogromnych rozmiarów danych, które podlegają przetwarzaniu
- Skupiona wokół technologii hurtowni danych, będących tematycznymi bazami danych, gromadzącymi historyczne dane o działalności przedsiębiorstwa

Środowisko hurtowni danych

Model podstawowy

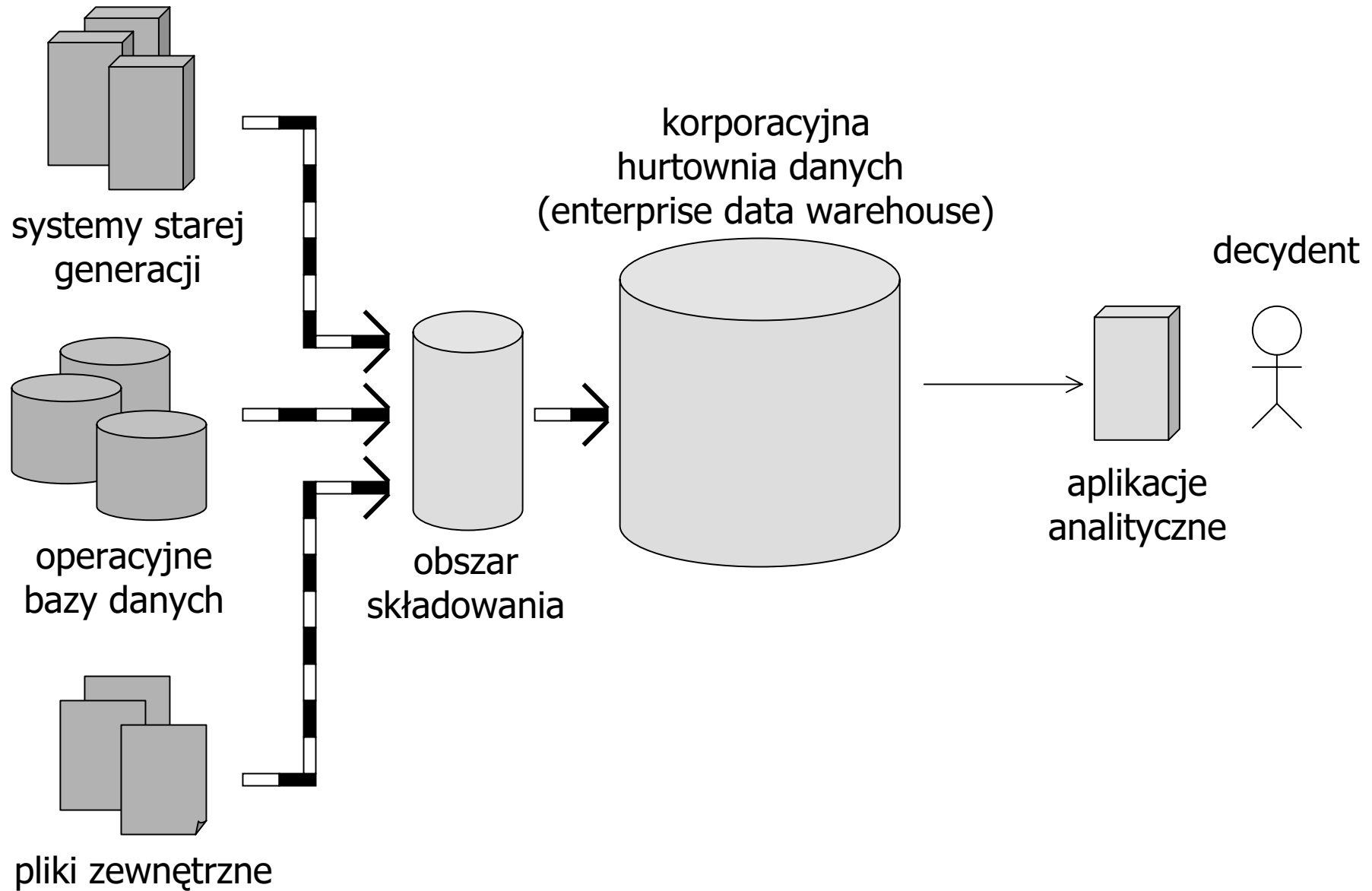
4



Środowisko hurtowni danych

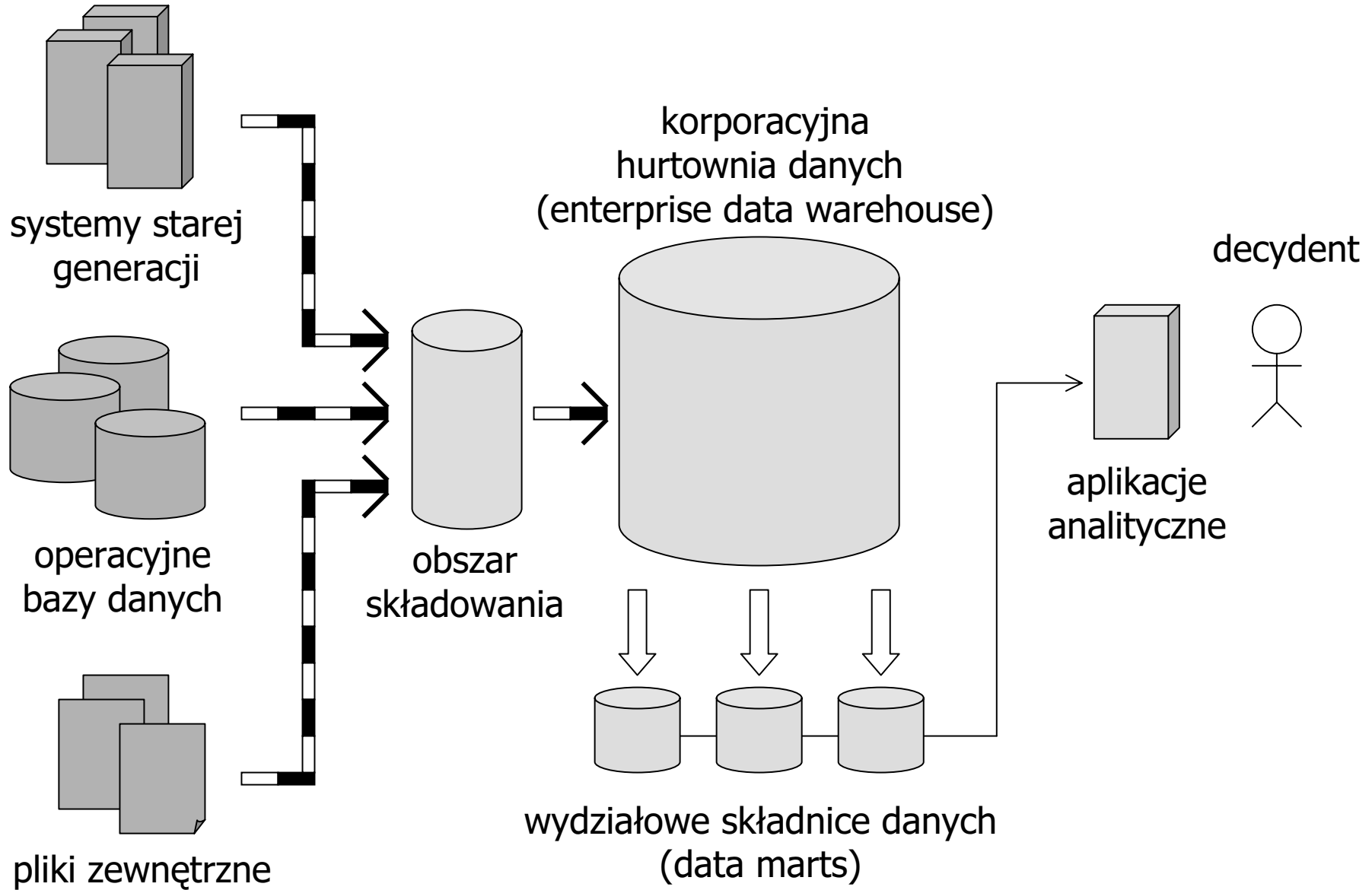
Architektura z obszarem składowania

5



Środowisko hurtowni danych

Architektura z obszarem składowania i składnicami danych



Hurtownia danych - definicja

- „Hurtownia danych to tematyczna baza danych, która trwale przechowuje zintegrowane dane opisane wymiarem czasu” [Inmon96]
- „Tematyczna baza danych” – dane dotyczą głównych obszarów działalności przedsiębiorstwa
- „trwale przechowuje” – dane nie są zmieniane ani usuwane; hurtownia danych ma charakter przyrostowy
- „zintegrowane dane” – dane dotyczące tego samego podmiotu stanowią całość
- „opisane wymiarem czasu” – dane opisują zdarzenia historyczne, a nie tylko stan aktualny

Porównanie hurtowni danych z systemami OLTP

Cecha	OLTP	Hurtownia danych
czas odpowiedzi aplikacji	ułamki sekundy – sekundy	sekundy – godziny
wykonywane operacje	DML	select
czasowy zakres danych	30-60 dni	2-10 lat
organizacja danych	według aplikacji	tematyczna
rozmiar	małe – duże	duże – wielkie
intensywność operacji dyskowych	mała – średnia	wielka

Porównanie hurtowni danych ze składnicami danych

Cecha	Hurtownia danych	Składnica danych
zasięg wykorzystywania	przedsiębiorstwo	wydział
zakres tematyczny	wielotematyczna	jednotematyczna
liczba źródeł danych	wiele	1 – kilka
czas implementacji i wdrożenia	miesiące – lata	miesiące

Metodyki projektowania i wdrażania hurtowni danych

10

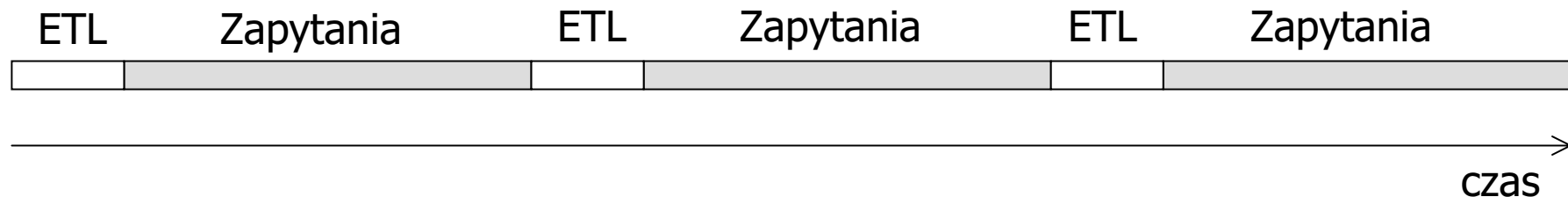
- „Wielki wybuch”
 - budowa korporacyjnej hurtowni danych w ramach jednego projektu
 - globalna analiza wymagań, implementacja całościowej hurtowni danych, implementacja aplikacji analitycznych
 - długi czas realizacji, początkowo wybrane technologie przestają być wspierane przez dostawców
- Od ogółu do szczegółu
 - iteracyjna realizacja hurtowni danych, po kolei dla każdego obszaru tematycznego, konstrukcja zależnych składnic danych
- Od szczegółu do ogółu
 - realizacja niezależnych składnic danych, z wizją ich przyszłej integracji w jedną hurtownię danych

ETL: Extraction, Transformation, Loading

- Ekstrakcja: odczyt źródłowych danych z operacyjnych baz danych, systemów starej generacji, plików zewnętrznych
- Transformacja: łączenie danych, ich weryfikacja, walidacja, czyszczenie i znakowanie czasowe
- Wczytywanie: wprowadzanie danych do docelowej hurtowni danych
- Realizacja ETL jest najtrudniejszym zadaniem implementacji hurtowni danych (pochłania nawet 70% czasu)

Dwa tryby pracy hurtowni danych

- Ładowanie danych
 - zwykle wykonywane w regularnych odstępach czasu, w porze niskiej aktywności użytkowników
- Realizacja zapytań analitycznych
 - podstawowy rodzaj obciążenia systemu hurtowni danych



Powyższa zmienność charakterystyki obciążenia systemu hurtowni danych stanowi poważny problem konfiguracyjny dla administratora bazy danych.

Kroki technicznej implementacji systemu ¹³ Business Intelligence

- Analiza wymagań – zgromadzenie wiedzy o wymaganiach biznesowych w zakresie przetwarzania analitycznego
- Projekt logiczny hurtowni danych – pojęciowa definicja wymaganych struktur danych
- Implementacja struktur fizycznych hurtowni danych – tworzenie bazy danych, tabel, indeksów, materializowanych perspektyw
- Implementacja oprogramowania ETL – konstrukcja modułów programowych służących do zasilania hurtowni danych nowymi danymi
- Realizacja aplikacji analitycznych – implementacja programów dla użytkowników końcowych
- Strojenie hurtowni danych – rekonfiguracja serwera bazy danych, tworzenie dodatkowych indeksów i materializowanych perspektyw

Technologie Oracle dla hurtowni danych

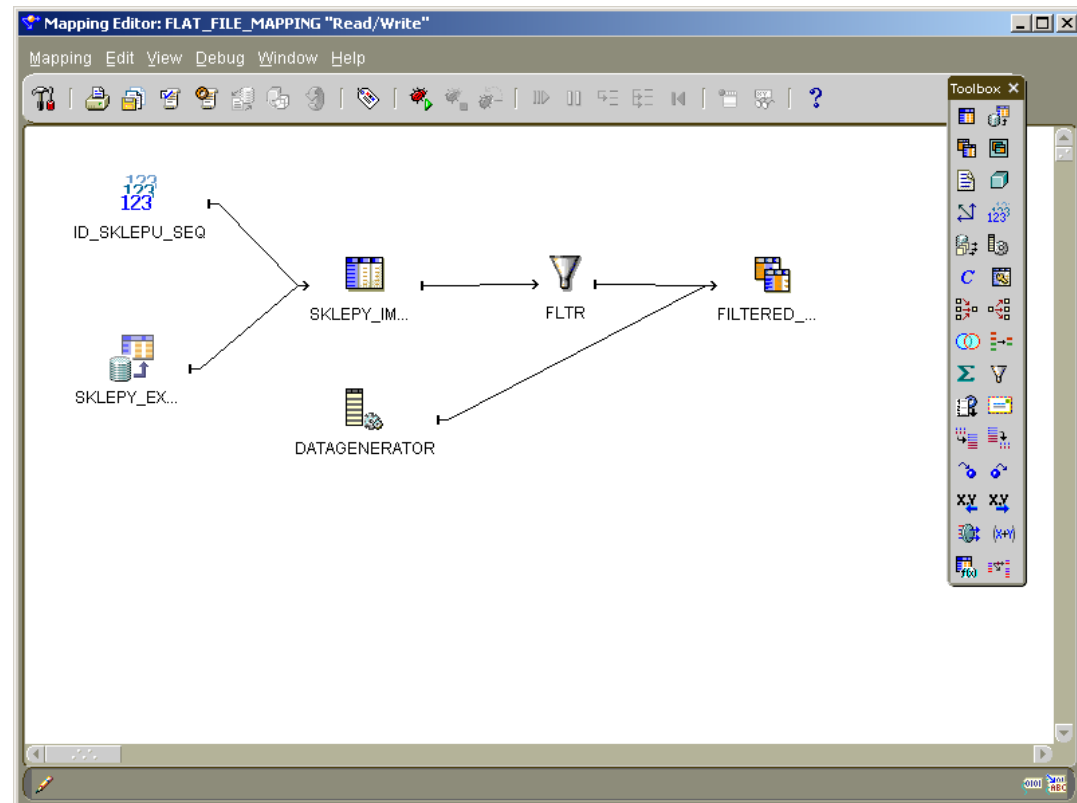
- Implementacja hurtowni danych
 - Oracle Database
 - Oracle OLAP
 - Oracle Warehouse Builder
- Implementacja ładowania danych
 - SQL*Loader
 - Replikacja
 - Tabele zewnętrzne
 - ODBC/JDBC
 - Oracle Gateways
- Implementacja aplikacji analitycznych
 - Oracle Discoverer
 - Oracle Business Intelligence Beans
 - Oracle Reports
 - Oracle Data Mining

Oracle Database 10g

- Serwer bazy danych
- Analityczne funkcje SQL
- Struktury fizyczne dla hurtowni danych
- Wielowymiarowy model danych
- Repozytorium metadanych OLAP
- Narzędzia ETL

Oracle Warehouse Builder

- Modelowanie i implementacja schematu hurtowni danych
- Modelowanie i implementacja procesów ETL
- Definiowanie metadanych OLAP



Oracle Enterprise Manager 10g

- Administrowanie bazą danych Oracle 10g
- Definiowanie metadanych OLAP

Create Dimension

General Levels Hierarchies Attributes **OLAP Options**

Plural Name SKLEPY

Default Display Hierarchy H_SKLEPY

Short Description SKLEPY

Long Description

Primary Order Hierarchical Level

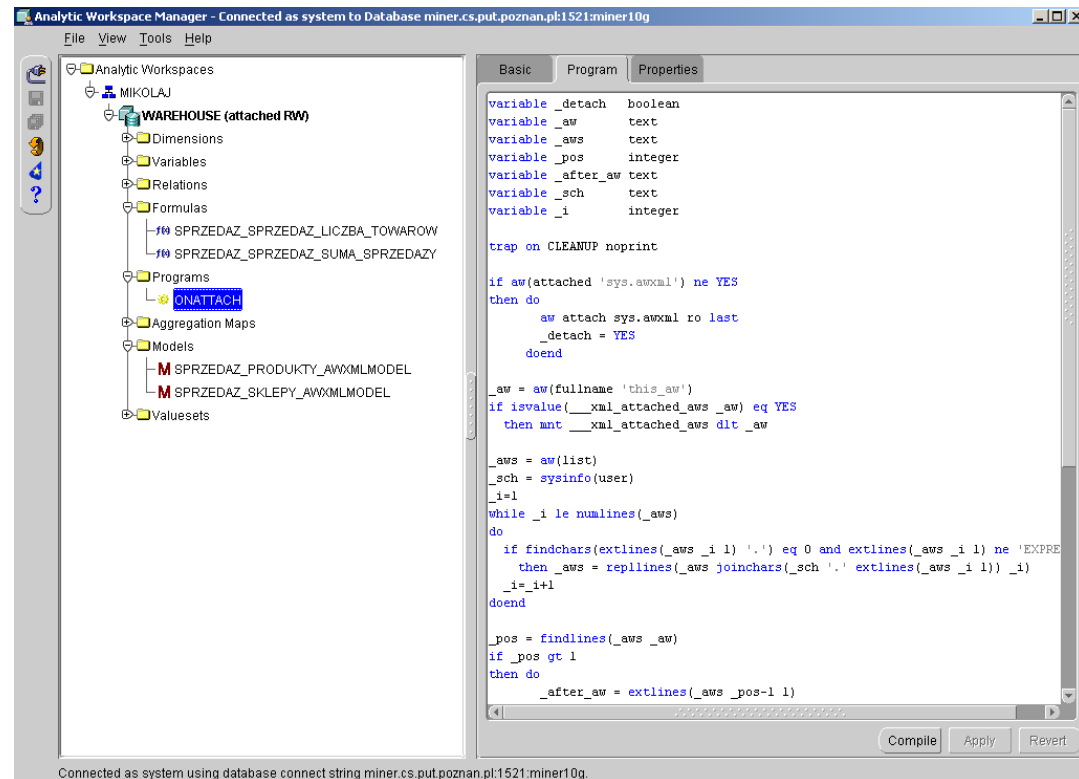
Secondary Order Ascending Descending

Display Names and Descriptions

Component	Type	Display Name
SKLEP	Level	SKLEP
MIEJSCOWOSC	Level	MIEJSCOWOSC
REGION	Level	REGION
H_SKLEPY	Hierarchy	H_SKLEPY
Long_Description	Attribute	Long_Description
Short_Description	Attribute	Short_Description
NAZWA_SKLEPU	Attribute	NAZWA_SKLEPU

Analytic Workspace Manager

- Przeglądanie OLAP Catalog
- Automatyczne tworzenie analitycznych przestrzeni roboczych
- Tworzenie i edycja obiektów wielowymiarowych
 - wymiary
 - zmienne
 - formuły
 - programy
 - zbiory wartości



OLAP Worksheet

- Wywoływanie komend SQL
- Podstawowy interfejs dla OLAP DML

The screenshot displays the Oracle OLAP Worksheet interface. On the left, a SQL query window shows the command `select * from scott.sklepy`. The result is a table with 20 rows and 5 columns: ID_SKL..., NAZWA_SKL..., MIEJSCOWOSC, and REGION.

ID_SKL...	NAZWA_SKL...	MIEJSCOWOSC	REGION	
0	1	Store No. 1	New York	East
1	3	Store No. 3	Atlanta	East
2	4	Store No. 4	Los Angeles	West
3	5	Store No. 5	San Francisco	West
4	7	Store No. 7	Pittsburgh	East
5	8	Store No. 8	New Orleans	East
6	9	Store No. 9	Seattle	West
7	10	Store No. 10	Dallas	Central
8	11	Store No. 11	Cincinnati	Central
9	12	Store No. 12	Minneapolis	Central
10	13	Store No. 13	Louisville	Central
11	14	Store No. 14	Phoenix	West
12	15	Store No. 15	Denver	West
13	16	Store No. 16	St. Louis	Central
14	17	Store No. 17	Washington	East
15	18	Store No. 18	Miami	East
16	19	Store No. 19	Boston	East
17	20	Store No. 20	Nashville	Central
18	2	Store No. 2	Chicago	Central
19	6	Store No. 6	Philadelphia	East

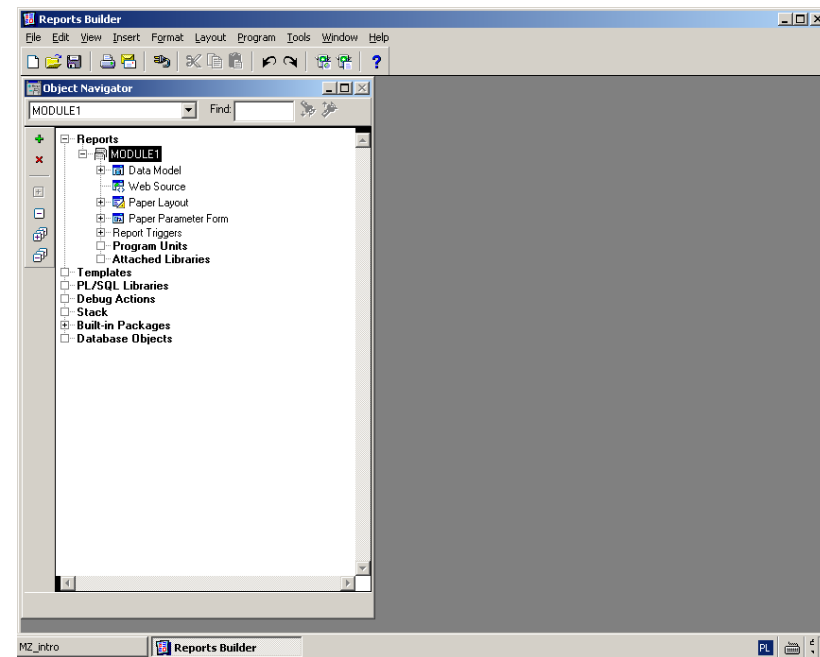
The main window shows a pivot table titled "LICZBA_KLIENTOW" (Number of Customers) with "SKLEPY" (Stores) on the rows and "PRODUKTY" (Products) on the columns. The pivot table includes a "TOTAL" column. The data is as follows:

SKLEPY	PRODUKT_1	PRODUKT_2	PRODUKT_3	PRODUKT_4	PRODUKT_5	PRODUKT_6	PRODUKT_7	PRODUKT_8	PRODUKT_9	PRODUKT_10	TOTAL
MIASTO_Atlanta	54	29	70	8	3	74	68	86	89	72	55
MIASTO_Boston	74	45	93	54	89	58	57	93	30	17	61
MIASTO_Chicago	66	56	82	29	89	77	74	14	26	98	61
MIASTO_Cincinnati	95	88	35	54	39	40	3	61	19	88	52
MIASTO_Dallas	83	45	58	68	62	47	98	91	100	89	74
MIASTO_Denver	83	49	73	37	7	45	98	14	23	43	47
MIASTO_Los Angeles	71	2	28	56	82	90	78	2	85	65	55
MIASTO_Louisville	10	79	80	99	79	84	90	26	85	58	69
MIASTO_Miami	66	86	80	85	2	23	25	50	22	42	48
MIASTO_Minneapolis	45	67	31	86	46	46	69	44	99	91	62
TOTAL	647	546	630	576	498	584	660	481	578	663	5,86

At the bottom of the main window, the SQL command `report grandtotals subtotals down sklepy rowtotals across produkty: liczba_klientow` is visible.


Oracle Reports

- Środowisko budowy aplikacji raportujących w oparciu o zawartość bazy danych Oracle
- Przeznaczone do tworzenia aplikacji w środowiskach OLTP



Oracle Discoverer

- Aplikacja wspomagająca procesy podejmowania decyzji
- Wykonywanie zapytań ad hoc
- Analizowanie oraz formatowanie wyników zapytań
- Przygotowywanie danych do ich prezentacji
- Zarządzanie danymi



Oracle Discoverer Desktop - [Workbo...]

Plik Edycja Widok Arkusz Format Narzędzia

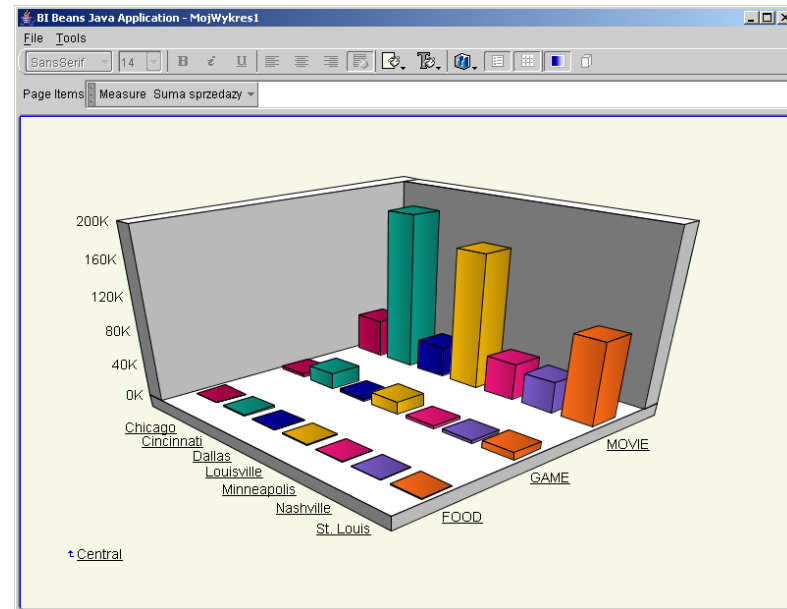
Wykres Okno Pomoc

	Suma Zysku SUM		
	▶ 1998	▶ 1999	▶ 2000
▶ FOOD	4390,14	3674,39	1943,80
Beverage	2665,73	2528,88	1135,71
Snacks	1724,41	1145,51	808,09
▶ GAME	73952,70	56764,49	35499,87
Game Rental	73952,70	56764,49	35499,87
▶ MOVIE	432946,44	520755,15	347341,16
Laser Disc Rent	43207,23	41360,66	27039,69
Video Rental	158074,10	148838,70	89079,87
Video Sale	231665,11	330555,79	231221,60

◀ ▶ Tabela Tabela ze stroną Macierz

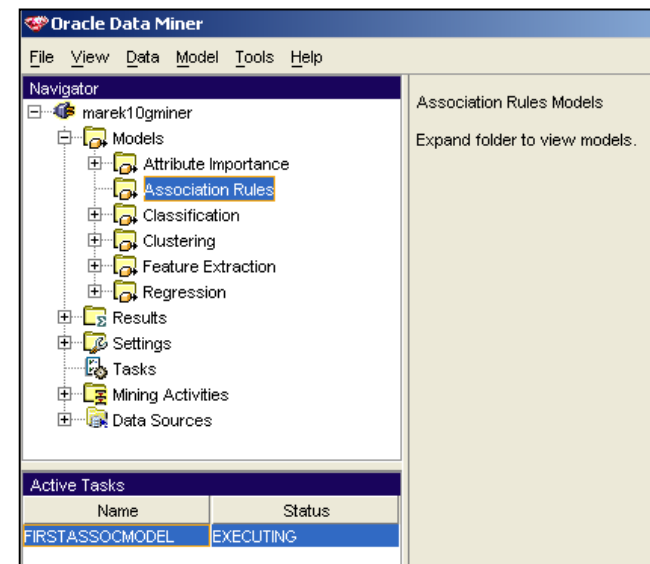
BI Beans

- Komponenty Java umożliwiające budowę aplikacji analitycznych opartych na bazie danych Oracle
- Wykorzystywane w środowisku Oracle Jdeveloper
- Komponenty do wykorzystania w aplikacjach samodzielnych Java i w aplikacjach JSP

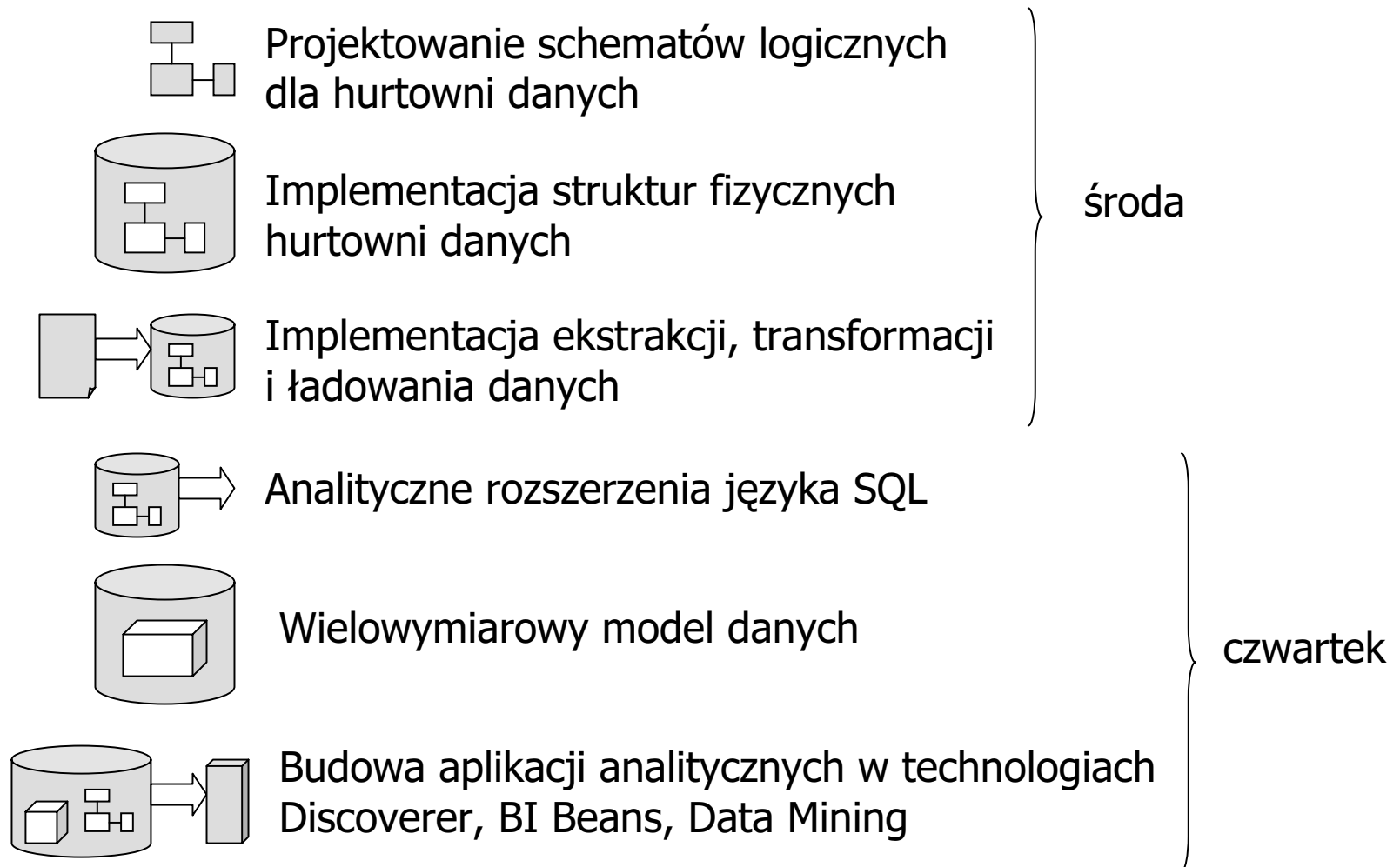


Oracle Data Miner

- Biblioteka Java dla środowiska Oracle Jdeveloper umożliwiającą budowę aplikacji eksploracji danych
- Narzędzie klienckie służące do wydawania zapytań eksploracyjnych ad-hoc
- Odkrywanie asocjacji
- Klasyfikacja



Struktura materiału IV Szkoły PLOUG



Struktury danych wykorzystywane w przykładach

25

