

KONFERENCJA
„Od studenta do specjalisty Business Intelligence”

HURTOWNIE DANYCH I BUSINESS INTELLIGENCE

PIOTR ZASKÓRSKI
PROF. WAT & WWSI



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI



WARSZAWSKA
WYŻSZA SZKOŁA
INFORMATYKI

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



AGENDA

- OGÓLNE KOMPONENTY MODELU ZASOBÓW INFORMACYJNYCH,
- HD - SYSTEMY OLAP
- OTOCZENIE SYSTEMÓW KLASY OLAP,
- SYSTEMY ZASILAJĄCE/ PROCESY EKSPLORACJI DANYCH,
- DEKOMPOZYCJA FUNKCJONALNA PRZEDMIOTU MODELOWANIA.
- IDEA MODELOWANIA I KLASY MODELI OLAP,
- MODELE TRANSFORMACJI ZASOBÓW INFORMACYJNYCH,
- DM = ODKRYWANIE WIEDZY
- SYSTEMY BI = DSS

KLASY DANYCH I SYSTEMÓW

DANE
OPERACYJNE/
TRANSAKCYJNE
ŹRÓDŁO = OLTP
MRP/ERP/DEM/Cax
ZSIZ
SI DZIEDZINOWE

DANE
HISTORYCZNE/ANALITYCZN
E/
OCZYSZCZONE
HD

ETL

OLAP

WYMIAROWANIE
wg kryterium

- ☐ CZASU
- ☐ MIEJSCA
- ☐ PRZEDMIOTU
- ☐ STOPNIA
AGREGACJI
- ☐ OBSZARU
- ☐ PROCESÓW
- ☐ ZADAŃ
- ☐ FUNKCJI
- ☐ POSTACI
- ☐ INNEGO

GENEZA HURTOWNI DANYCH

**RELACYJNE BAZY DANYCH NIE SĄ WYSTARCZAJĄCYM
ROZWIĄZANIEM
DLA SYSTEMÓW WSPOMAGANIA DECYZJI
(DECISION SUPPORT SYSTEM - DSS).**

**SPECYFIKA SYSTEMÓW DSS
ZAWIERA SIĘ W TYM , ŻE ABY
MOGŁY SPRAWNIE FUNKCJONOWAĆ,
POTRZEBUJĄ:**

**ODPOWIEDNIO JUŻ PRZYGOTOWANYCH
(OCZYSZCZONYCH, ZAGREGOWANYCH,
PRZETRANSFORMOWANYCH) DANYCH.**

WARSTWOWA STRUKTURA HD

WARSTWA DANYCH TERAŹNIEJSZYCH

WARSTWA DANYCH HISTORYCZNYCH

WARSTWA DANYCH OCZYSZCZONYCH

WARSTWA DANYCH SUMARYZOWANYCH

WARSTWA METADANYCH

FUNKCJONALNA STRUKTURA HD

**UTRZYMYWANIE JEDNEGO
WSPÓLNEGO REPOZYTORIUM DANYCH**

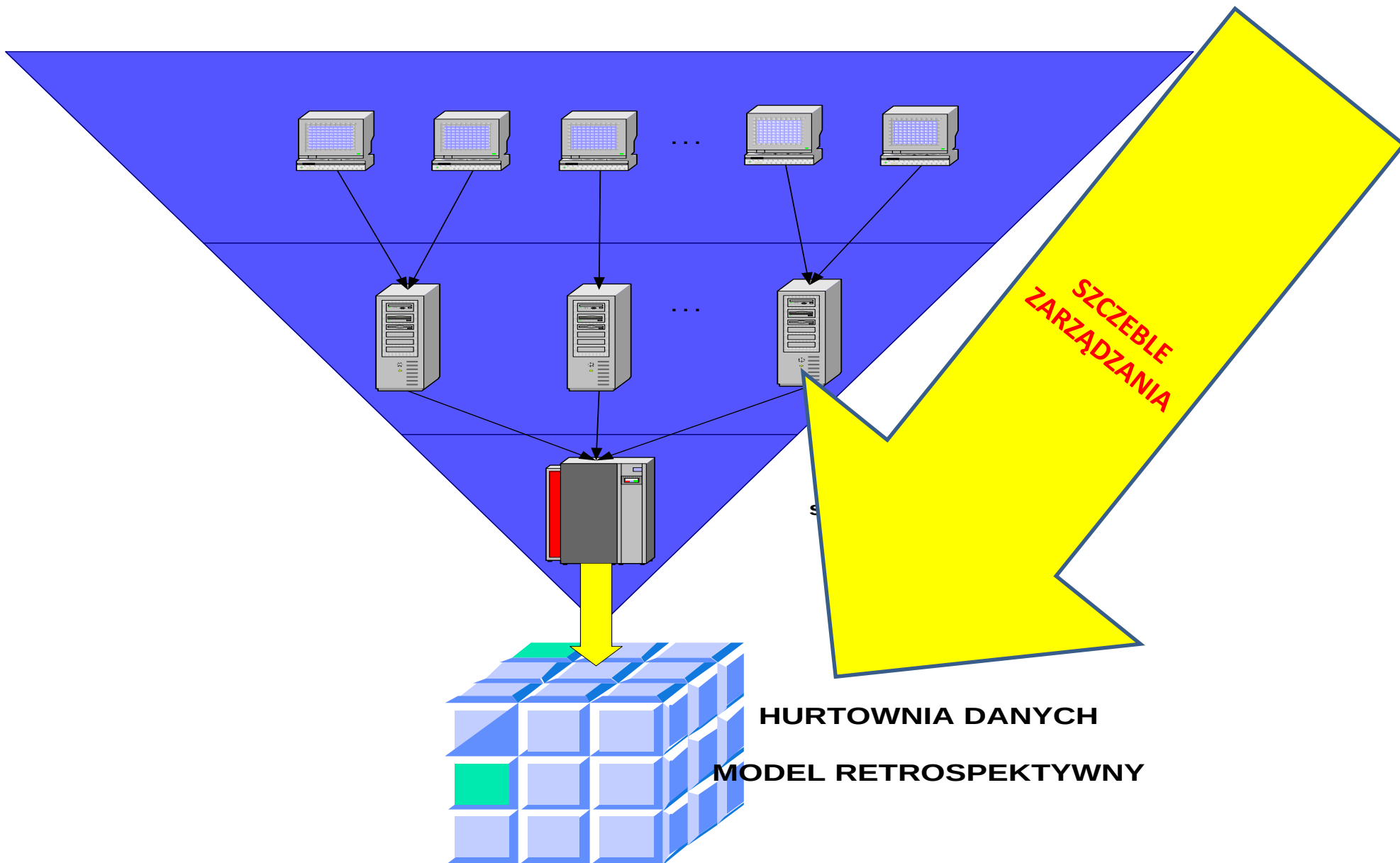
**ETL = EKSTRAKCJA DANYCH
Z HETEROGENICZNYCH ŹRÓDEŁ**

INTEGRACJA DANYCH ROZPROSZONYCH

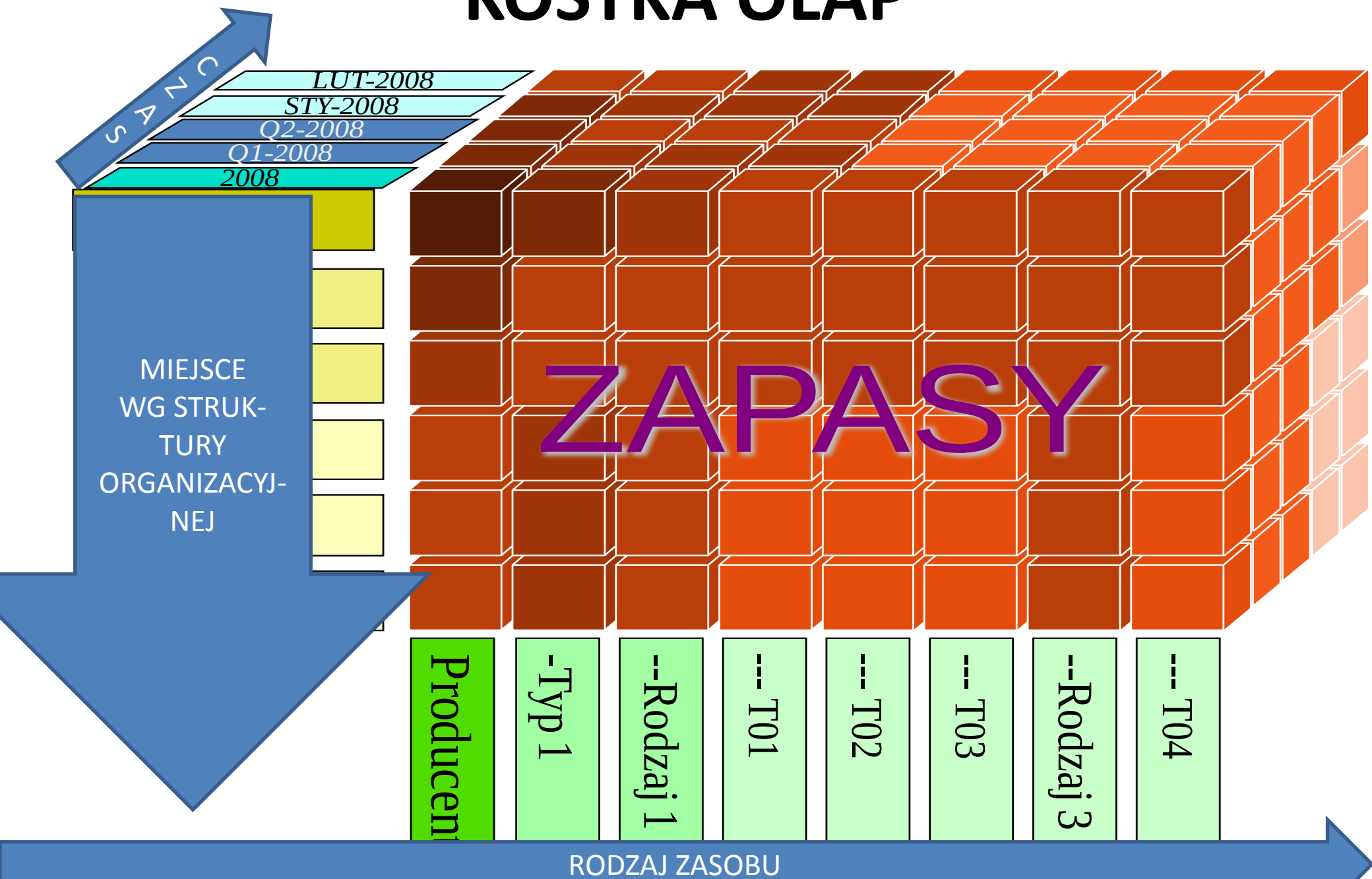
ANALIZY BIZNESOWE

**OBSŁUGA ROZPROSZONYCH
STRUKTUR DECYZYJNYCH**

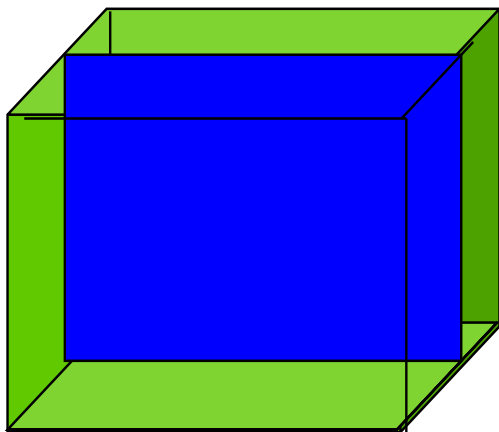
AGREGOWANIE DANYCH ANALITYCZNYCH



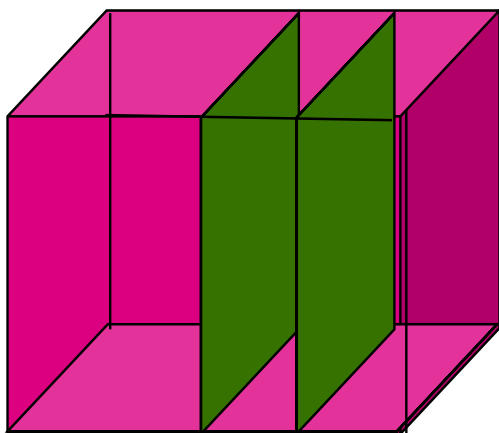
KOSTKA OLAP



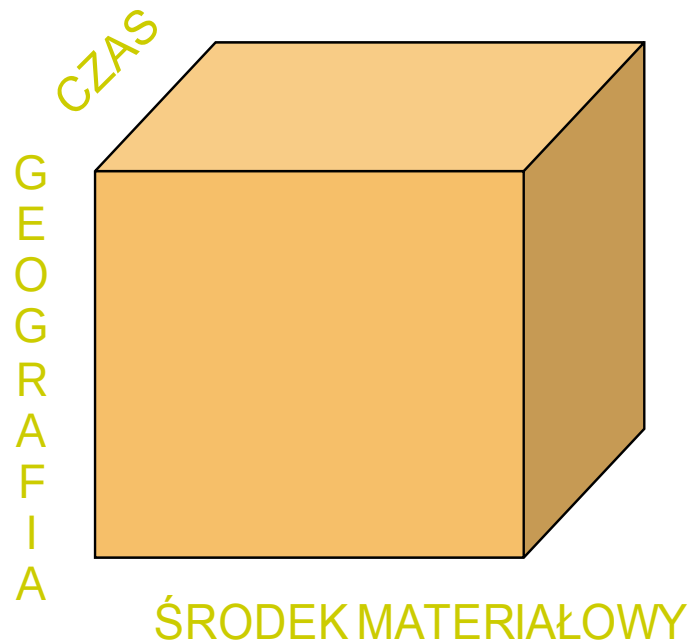
PERSPEKTYWA KOSTEK OLAP



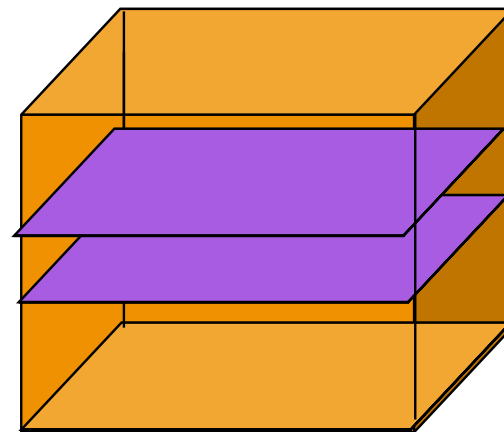
PODMIOT



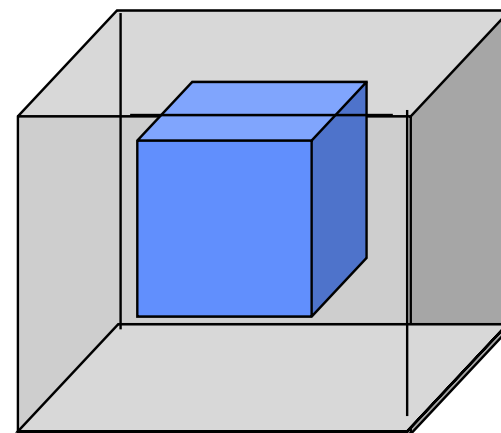
REGION



ŚRODEK MATERIAŁOWY



MAGAZYN



KOMÓRKA
ORGANIZACYJNA

FAKTY, WYMIARY, MIARY I AGREGACJE

ZWYMIAROWANE FAKTY

- PRODUKCJA = $f(t)$
- SPRZEDAŻ = $g(t)$
- USŁUGI = $u(t)$; PROJEKTY = $p(t)$

MIARY AGREGACJI

- PROSTE
 - ILOŚĆ, WARTOŚĆ, PRZYCHÓD, ZYSK, W_d , M_p , W_{kl}
- ZŁOŻONE
 - EFEKTYWNOŚĆ
 - NIEZAWODNOŚĆ
 - JAKOŚĆ
 - WYDAJNOŚĆ, PROGNOZY

MIARY

- SPRZEDAŻ
- WSKAŹNIKI EKONOMICZNE
- WSKAŹNIKI JAKOŚCI
- WSKAŹNIKI EFEKTYWNOŚCI
- WSKAŹNIKI NIEZAWODNOŚCI
- WSKAŹNIKI RYZYKA
- ANALIZA WARTOŚCI
- PROGNOZOWANIE
- PLANY

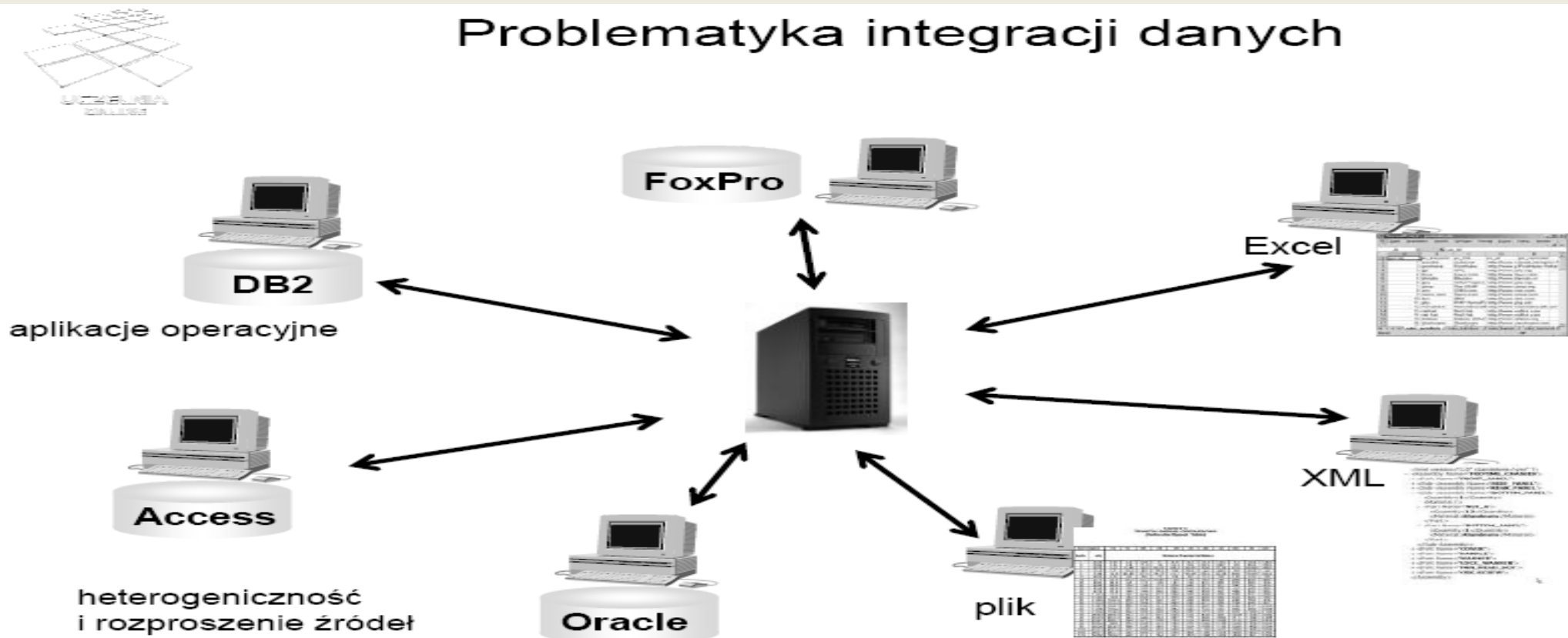
PROCEDURY AGREGACJI = MATADANE

- OBLICZENIA WSKAŹNIKOWE
- ANALIZA PORÓWNAWCZA
 - PARETO – LORENZA
 - INŻYNIERIA WARTOŚCI
- ANALIZA NIEZAWODNOŚCI
- DOM JAKOŚCI = QFD
- RISK – SCORE
- PLAN = f (PROGNOZA)

Problem integracji danych

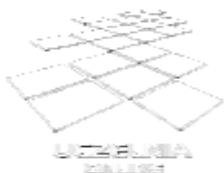
Heterogeniczność w systemach informatycznych oznacza, że systemy posiadają różne struktury, funkcjonalność i wykorzystują różne modele danych (np. hierarchiczne, relacyjne, obiektowe)

Dodatkowym problemem w dostępie do informacji jest geograficzne rozproszenie źródeł danych.

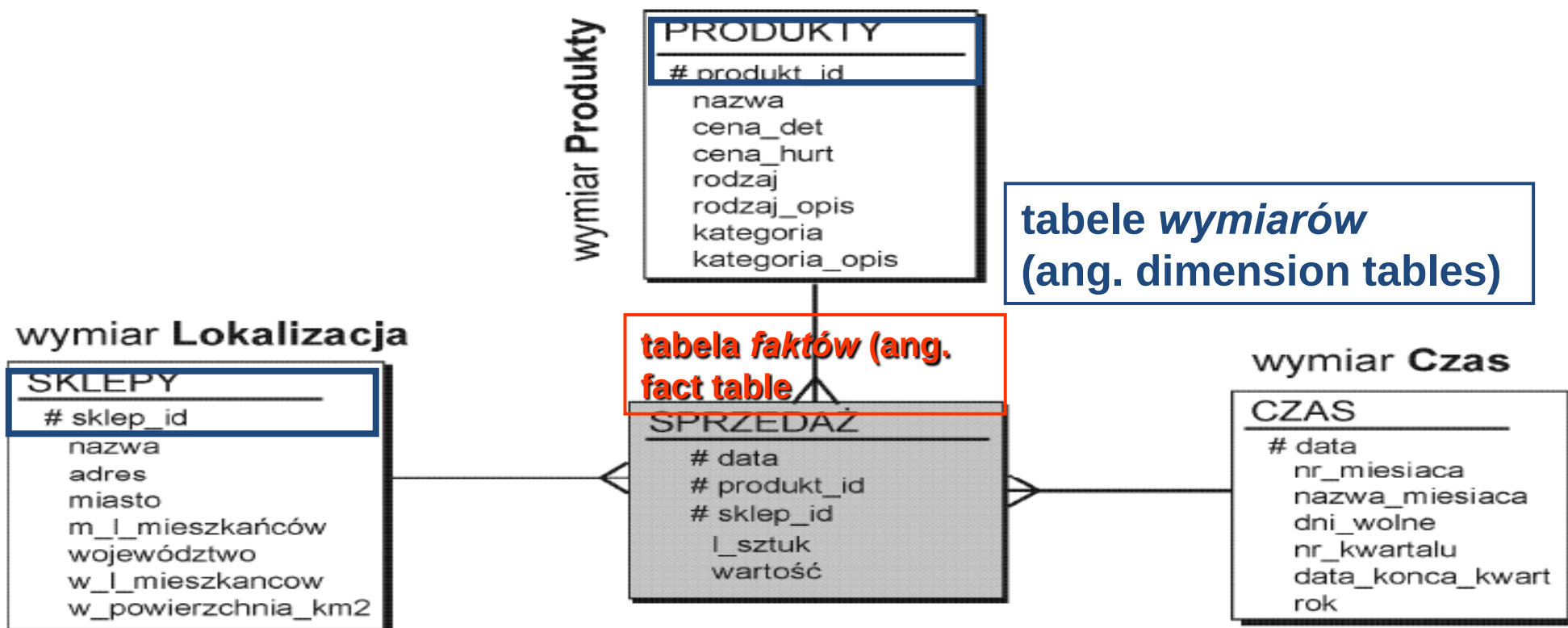


Schemat gwiazdy

Magazyn danych w technologii ROLAP o strukturze gwiazdy (ang. star schema)



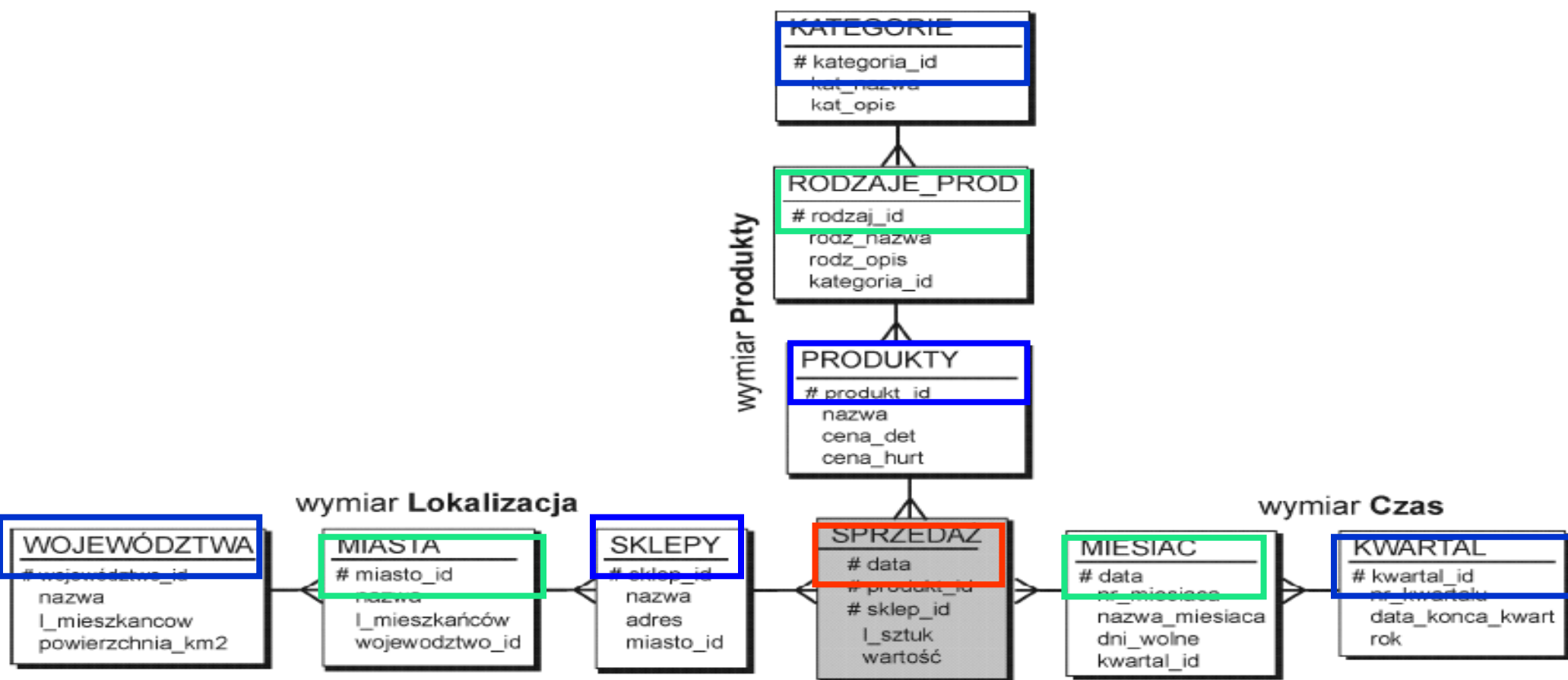
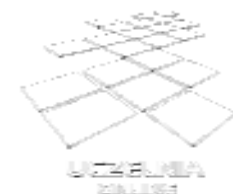
Schemat gwiazdy



Schemat płatka śniegu

Wymiary mają postać hierarchii w której produkty należą do rodzajów, a rodzaje do kategorii.

Schemat płatka śniegu



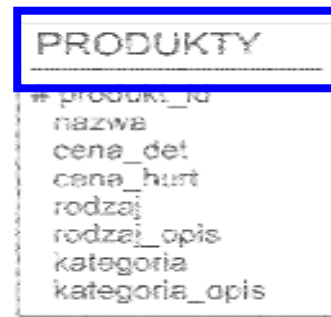
Schemat gwiazda-płatek śniegu

Jest połączeniem modelu gwiazdy i płatka śniegu.

Schemat gwiazda-płatek śniegu

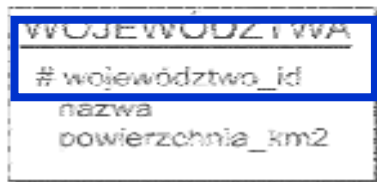
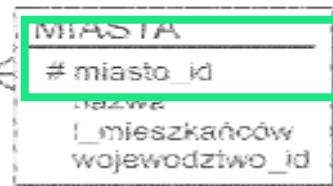
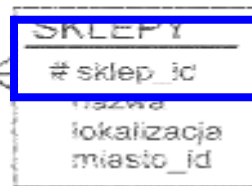
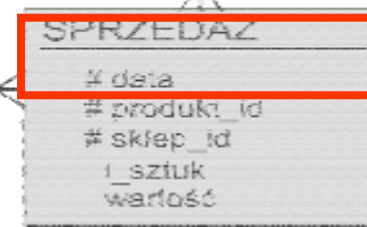
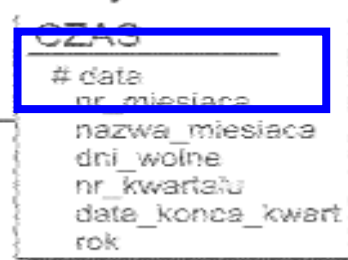
**Znormalizowany
wymiar**

wymiar Produkty



wymiar Lokalizacja

wymiar Czas



Implementacja MOLAP

Hurtownia danych w technologii MOLAP do przechowywania danych najczęściej wykorzystuje wielowymiarowe tablice (ang. multidimensional arrays, datacubes).

Tablice te zawierają wstępnie przetworzone (m.in. zagregowane) dane pochodzące z wielu źródeł.

Przykład:

Komórki tablicy zawierają np. zagregowane informacje o sprzedaży wybranych samochodów w poszczególnych latach, w wybranych miastach

Kostka

Miara sprzedaży BMW
w Warszawie w 1999 r.

L
o
k
a
l
i
z
a
c
j
a

Wa-wa

300 100 225 425 775

Kraków

600 210 345 900 650

Poznań

300 125 175 400 215

Kielce

415 690 630 325 510

1995

1996

1997

1998

1999

Czas

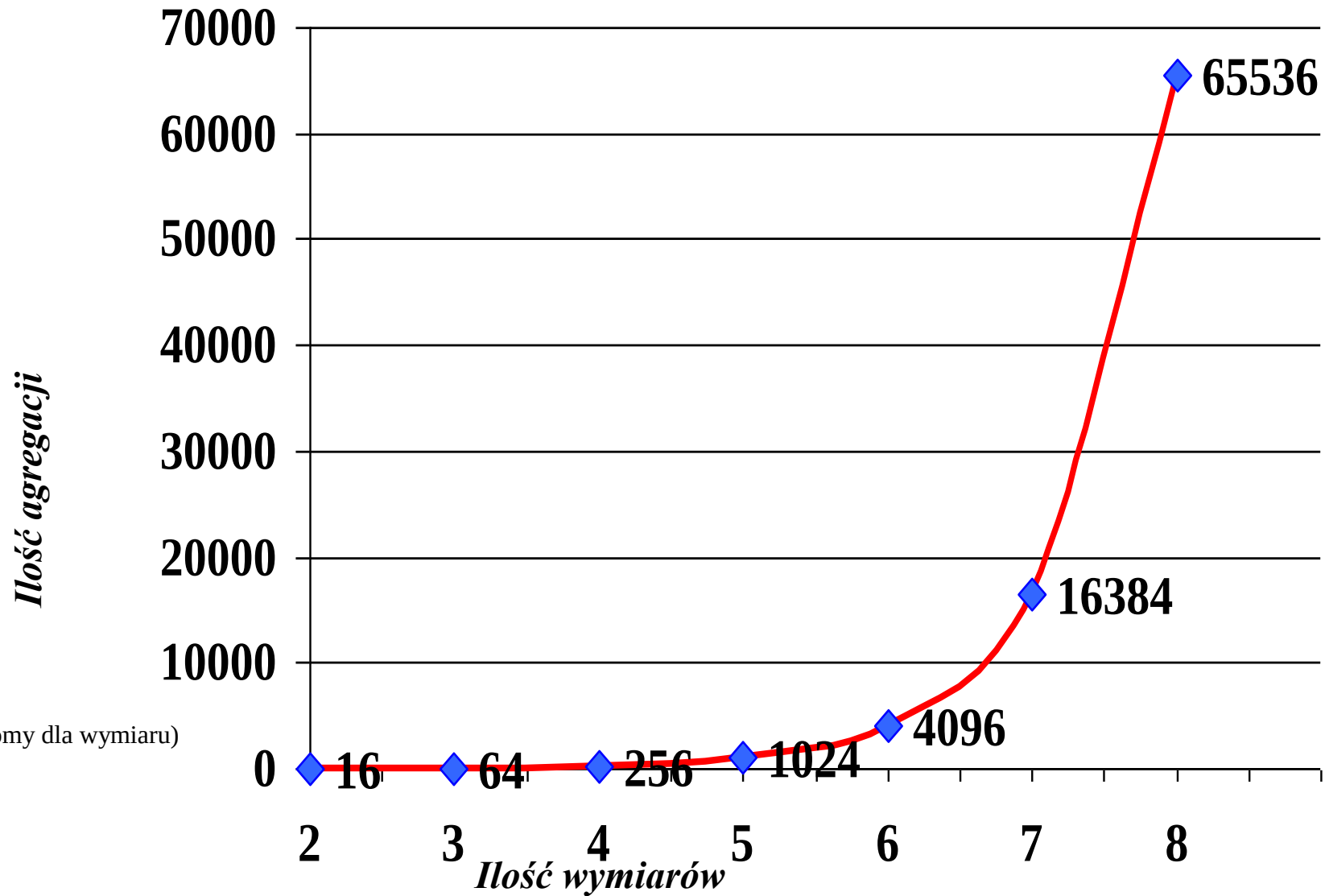
Samochód

Ford

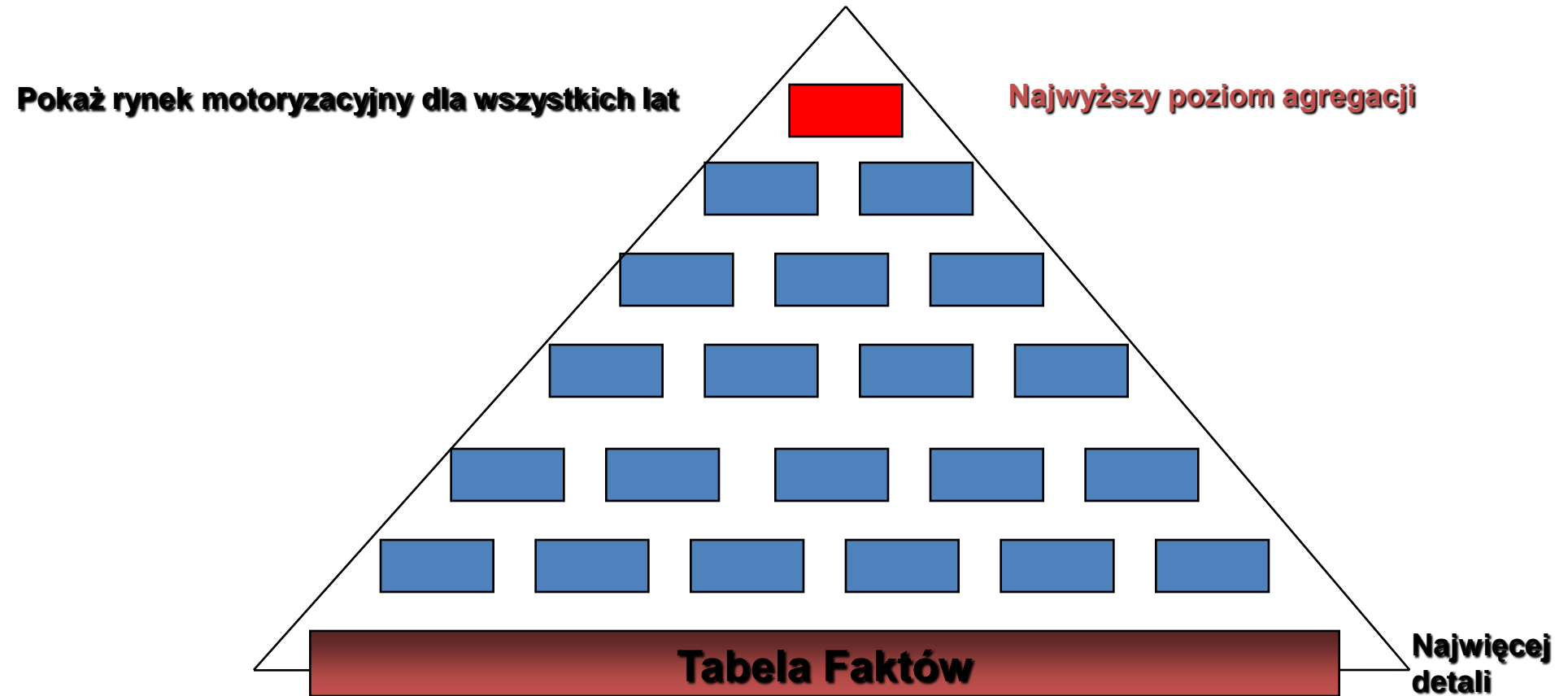
Audi

BMW

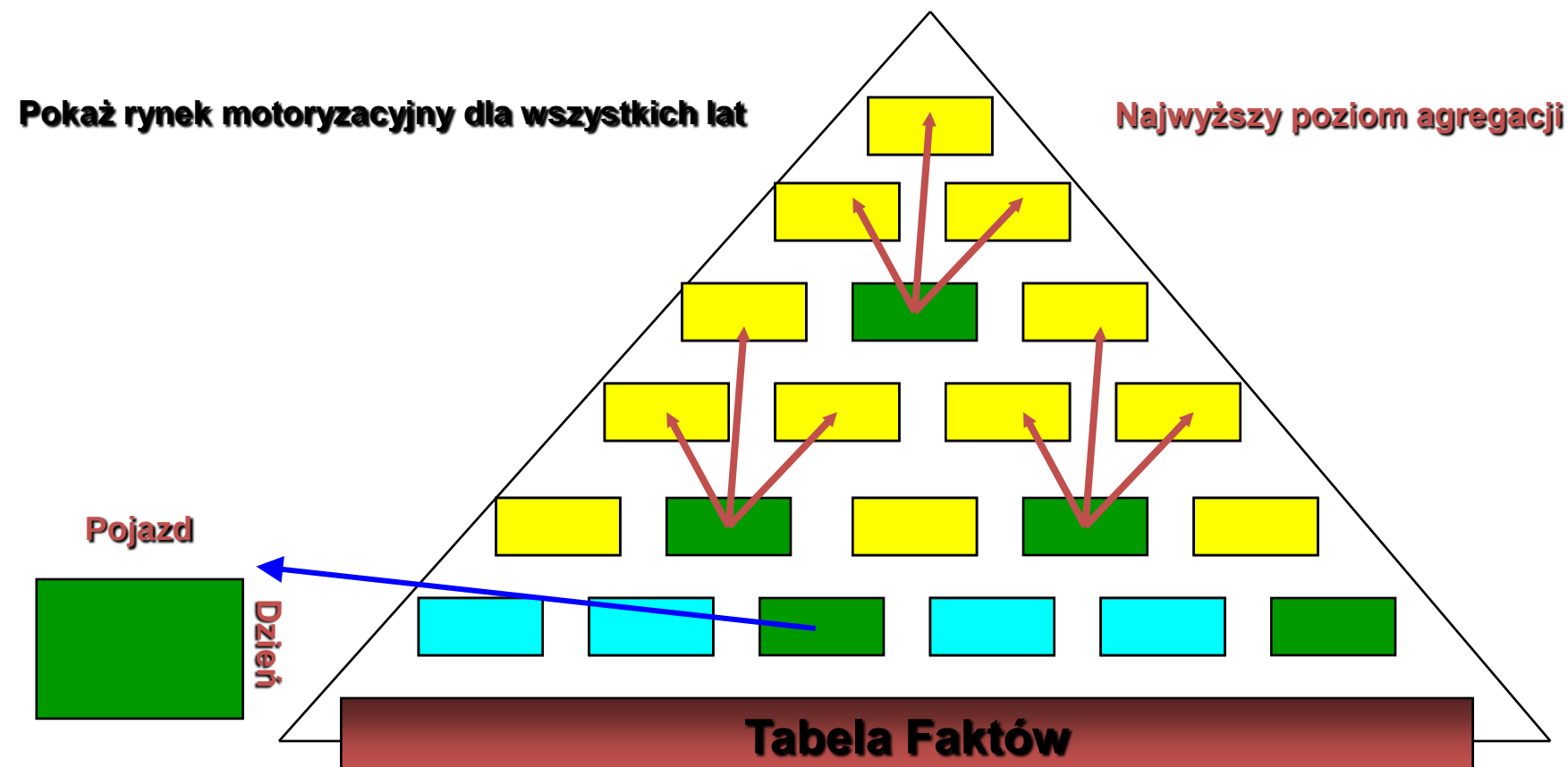
Problem eksplozji danych



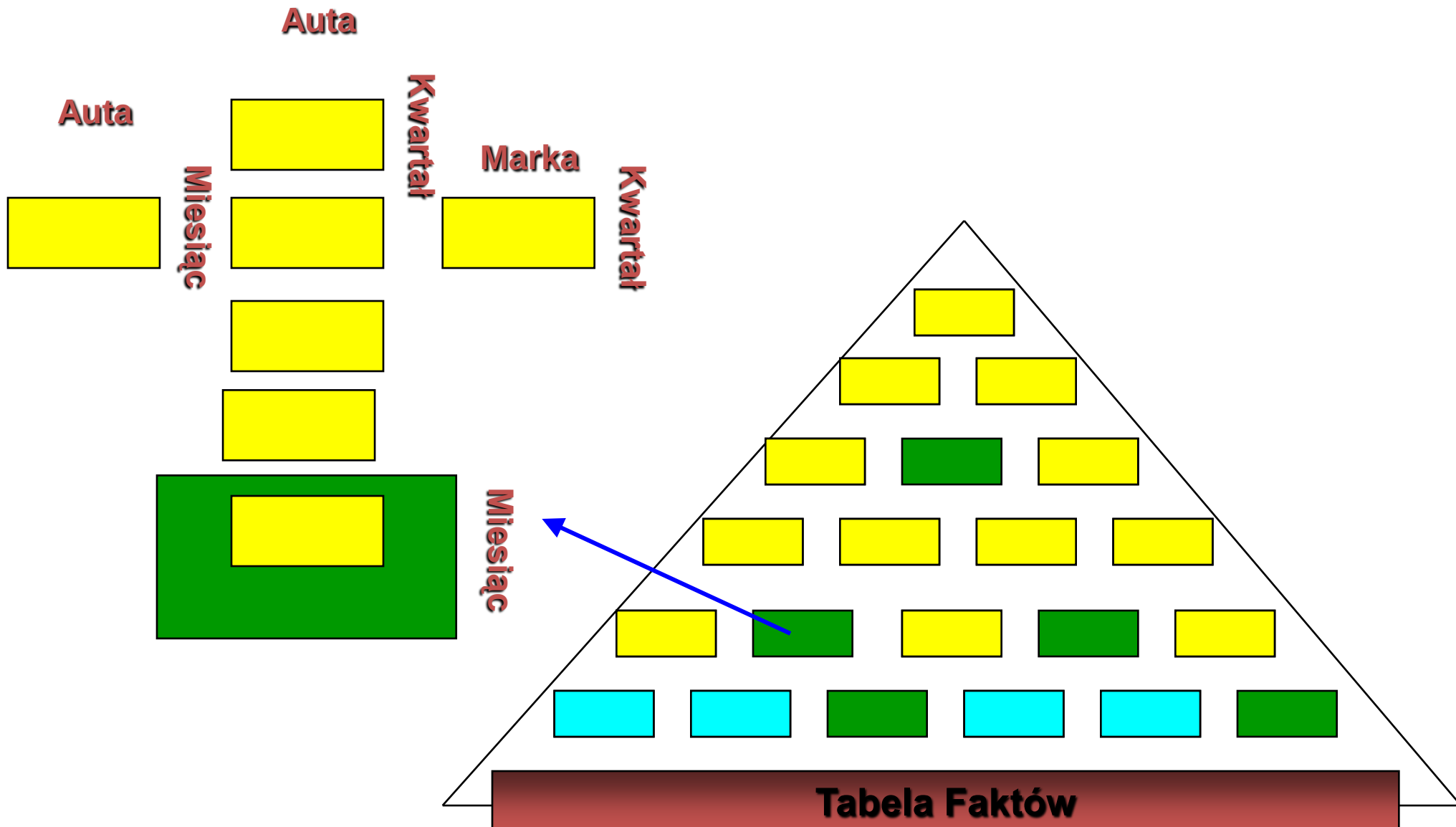
Agregacje (1)



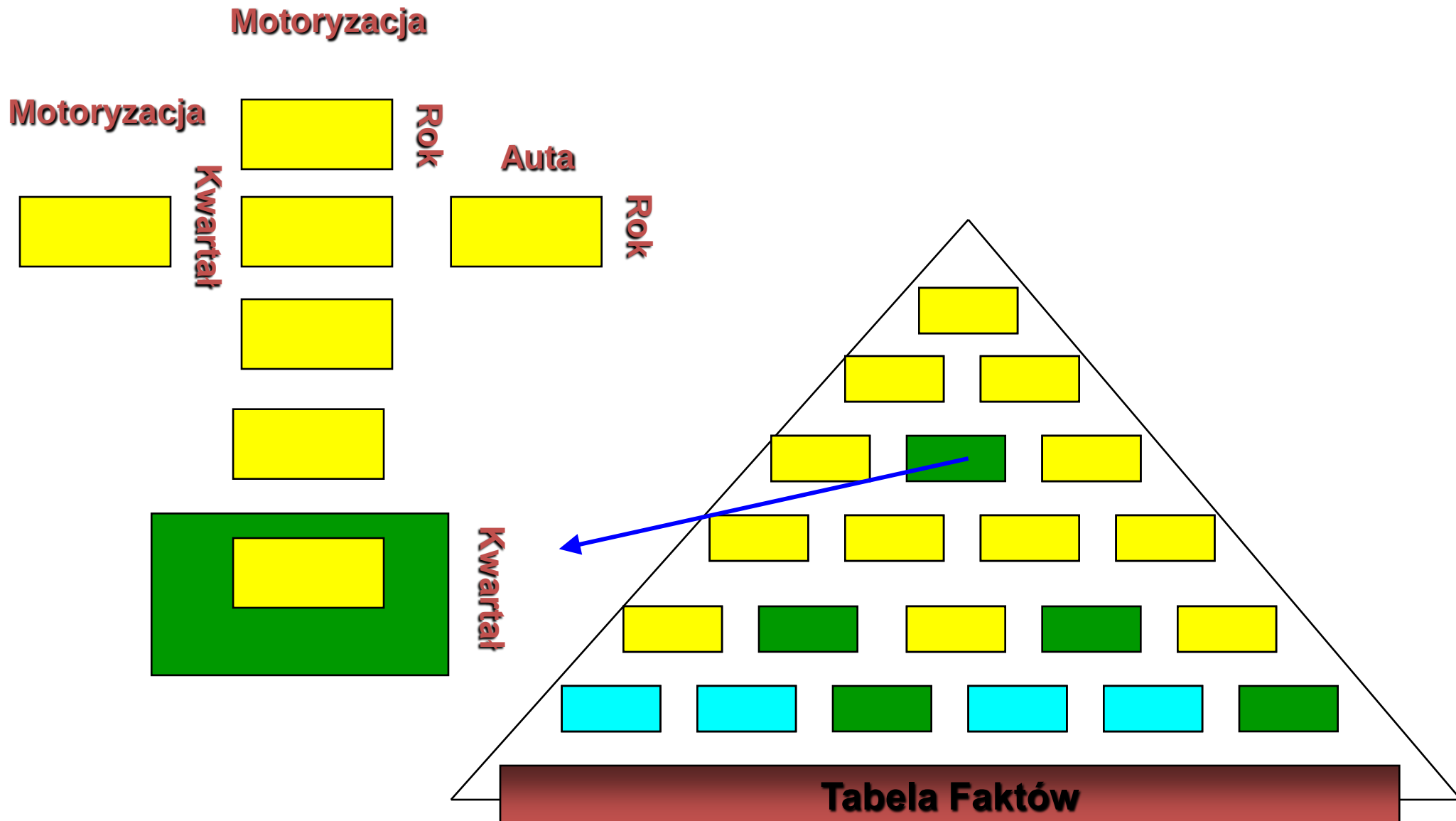
Agregacje (2)



Agregacje (3)



Agregacje (4)



ANALIZA SYSTEMU INFORMACYJNEGO

**1. IDENTYFIKACJA POTRZEB
INFORMACYJNYCH**

**FAKTY, MIARY, WYMIARY
NAZWY, ATRYBUTY,
DOMINUJĄCE WARTOŚCI CECH**

**2. IDENTYFIKACJA RELACJI
DECYDENTÓW/AGREGACJI**

**RELACJE
NAZWA I CHARAKTER
ZNACZENIE/METADANE**

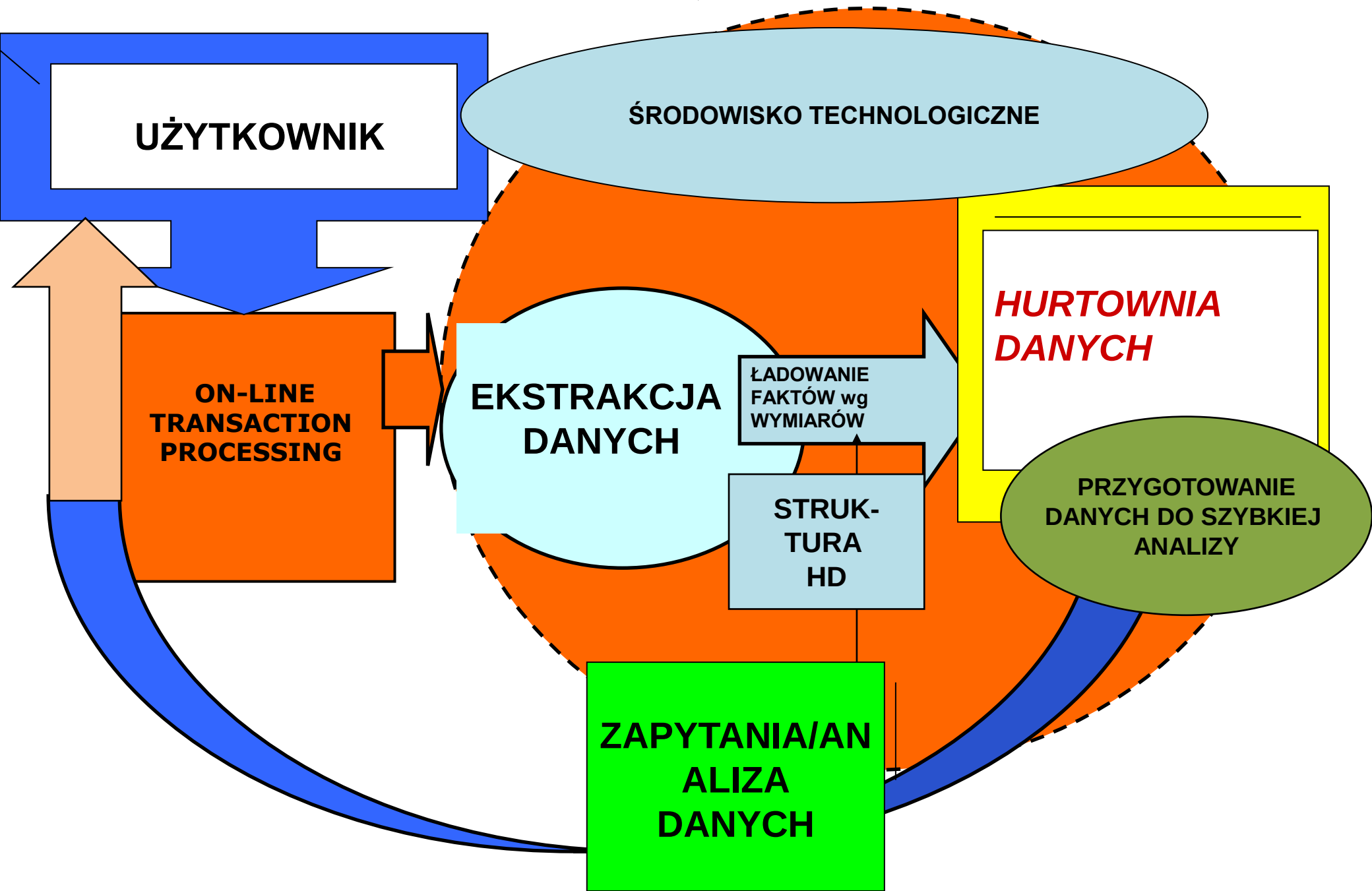
**3. IDENTYFIKACJA POTRZEB
PRZETWARZANIA**

**OPERACJE I PROCESY
CECHY I RELACJE, AGREGACJE
CZĘSTOTLIWOŚĆ
PRIORYTETY
RODZAJ DOSTĘPU
POTRZEBY UŻYTKOWNIKA
MOŻLIWOŚCI INTEGRACJI**

**4. IDENTYFIKACJA RELACJI
Z OTOCZENIEM HD-OLTP**

**ŹRÓDŁA DANYCH
PRAWA WŁASNOŚCI**

LOGIKA WYKORZYSTANIA HD



CYKL BUDOWY HD

Planowanie systemu

Zbieranie wymagań na dane
Modelowanie systemu

Projekt systemu
Projekt bazy danych

Pozyskiwanie, integracja
i odwzorowanie danych

Załadowanie hurtowni danymi

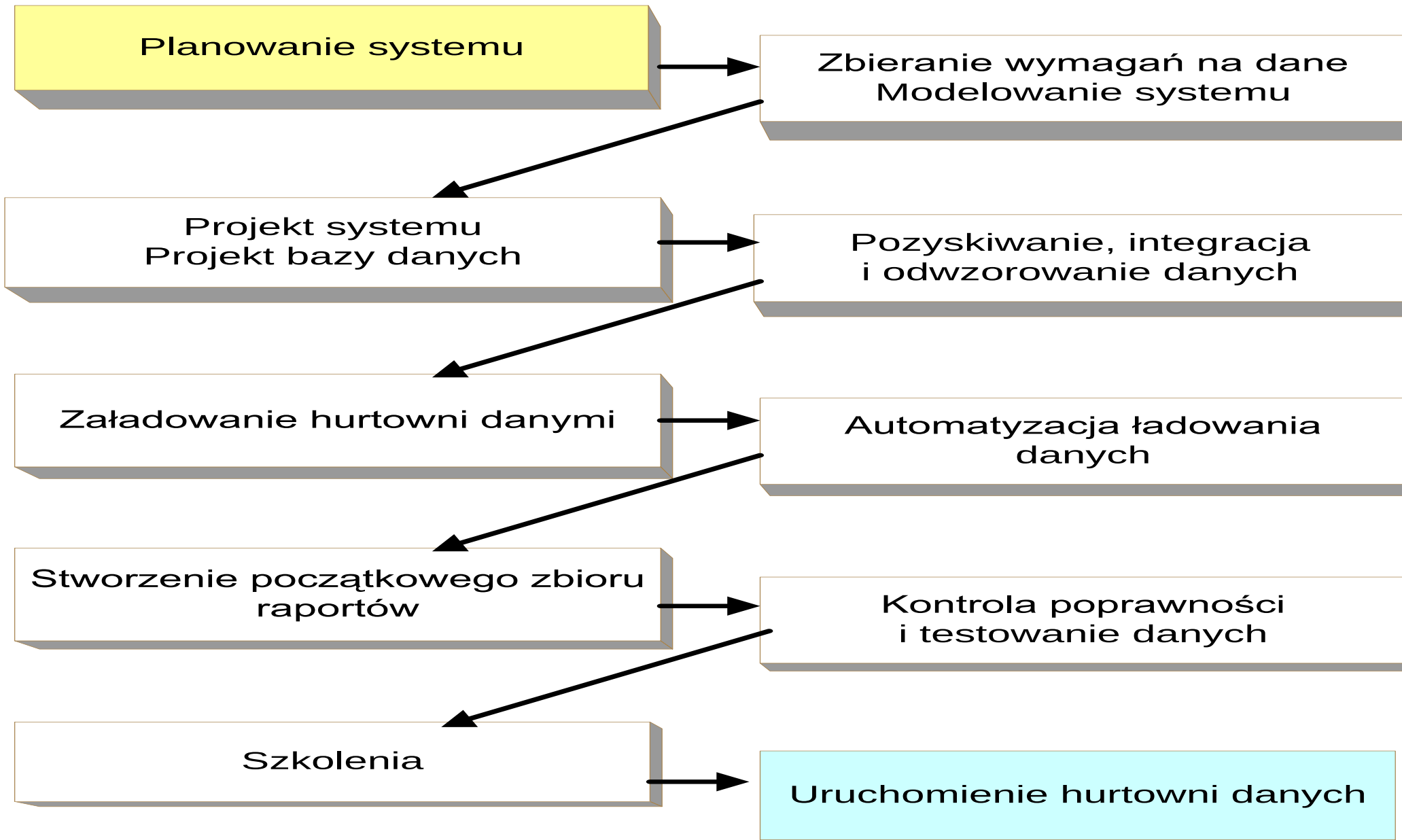
Automatyzacja ładowania
danych

Stworzenie początkowego zbioru
raportów

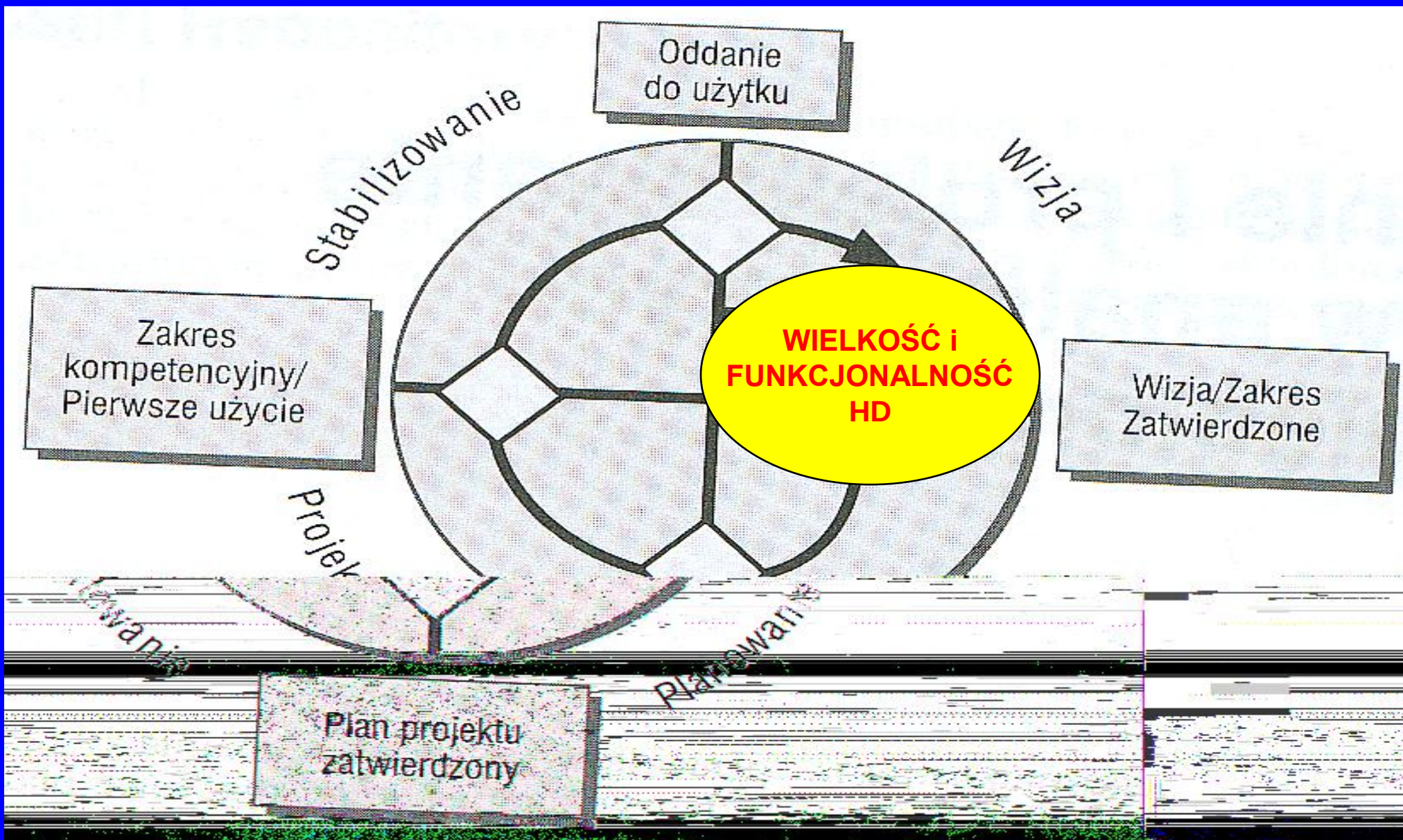
Kontrola poprawności
i testowanie danych

Szkolenia

Uruchomienie hurtowni danych



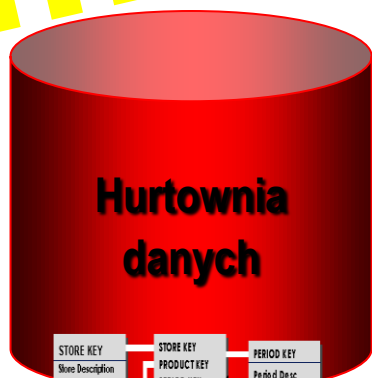
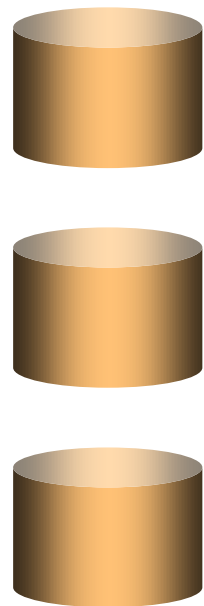
CYKL BUDOWY HD wg MICROSOFT



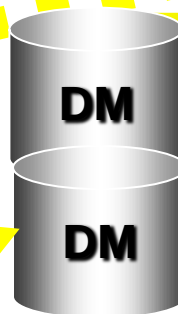
Business Intelligence - architektura

Informacja → Wiedza → Wnioski → Działanie → Wyniki

Systemy źródłowe

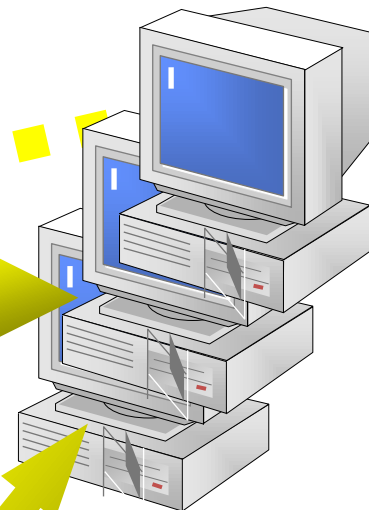


STORE KEY Store Description City State District ID District Desc. Region ID Region Desc. Regional Mgr. Level	STORE KEY Product Key PERIOD KEY Dollars Units Price	PERIOD KEY Period Desc Year Quarter Month Day Current Flag Resolution Sequence
	PRODUCT KEY Product Desc. Brand Color Size Manufacturer Level	



Kostka OLAP

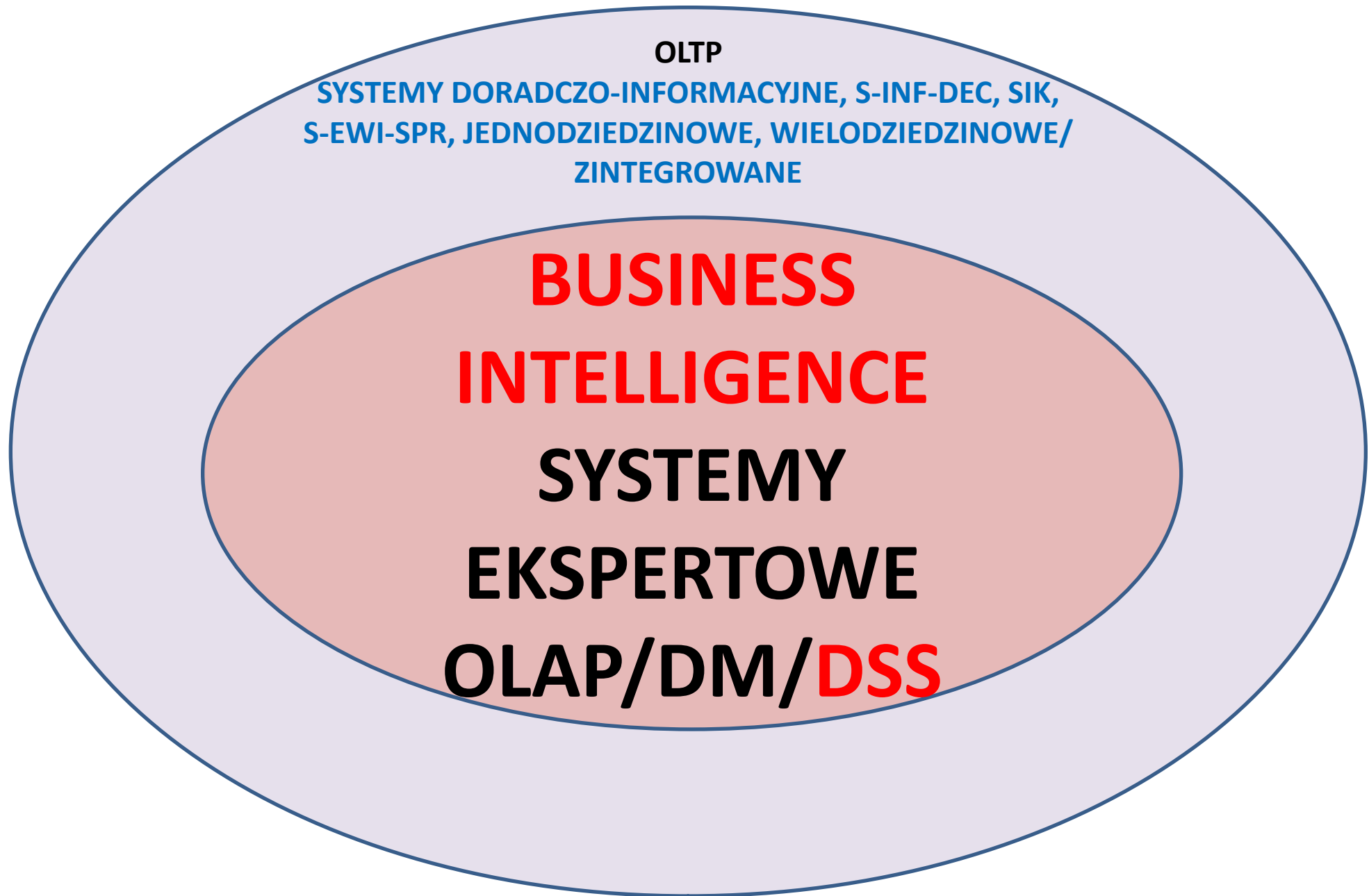
Kostka OLAP



Użytkownicy:

- Raporty
- Zapytania do bazy
- Analizy OLAP
- Budżetowanie
- Data Mining

BUSINESS INTELLIGENCE



ANALIZA I WSPOMAGANIE DECYZJI

**System
wspomagania
procesów
planistycznych**

$$\text{DSS} = p \{ \text{OLAP} = f(\text{OLTP}) \}$$

**PRZESZŁOŚĆ
MODEL
RETROSPEKTYWNY**

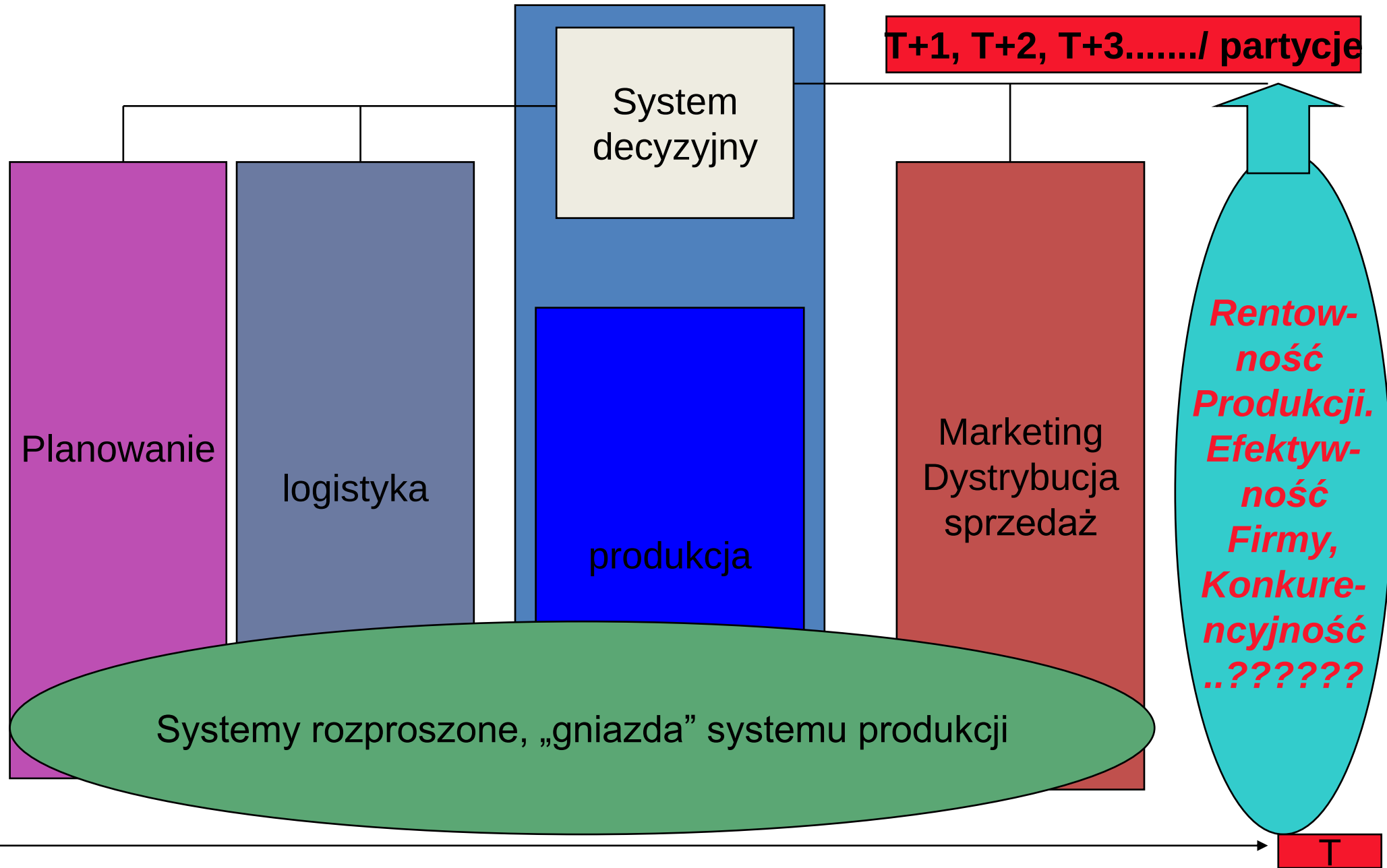
OLAP

X-ENGINEERING

**PRZYSZŁOŚĆ
MODEL
PROSPEKTYWNY
PROGNOZOWANIE**

DATA MINING

Informacyjno-decyzyjna ciągłość działania



FUNKCJONALNOŚĆ BI

NARZĘDZIA I APLIKACJE ANALITYCZNE
QUERY&REPORT/ARKUSZE/WIZUALIZACJA
DATA/TEXT MINING
APLIKACJE PROBLEMOWE/DZIEDZINOWE
PORTALE
AUTOMATYCZNA DYSTRYBUCJA

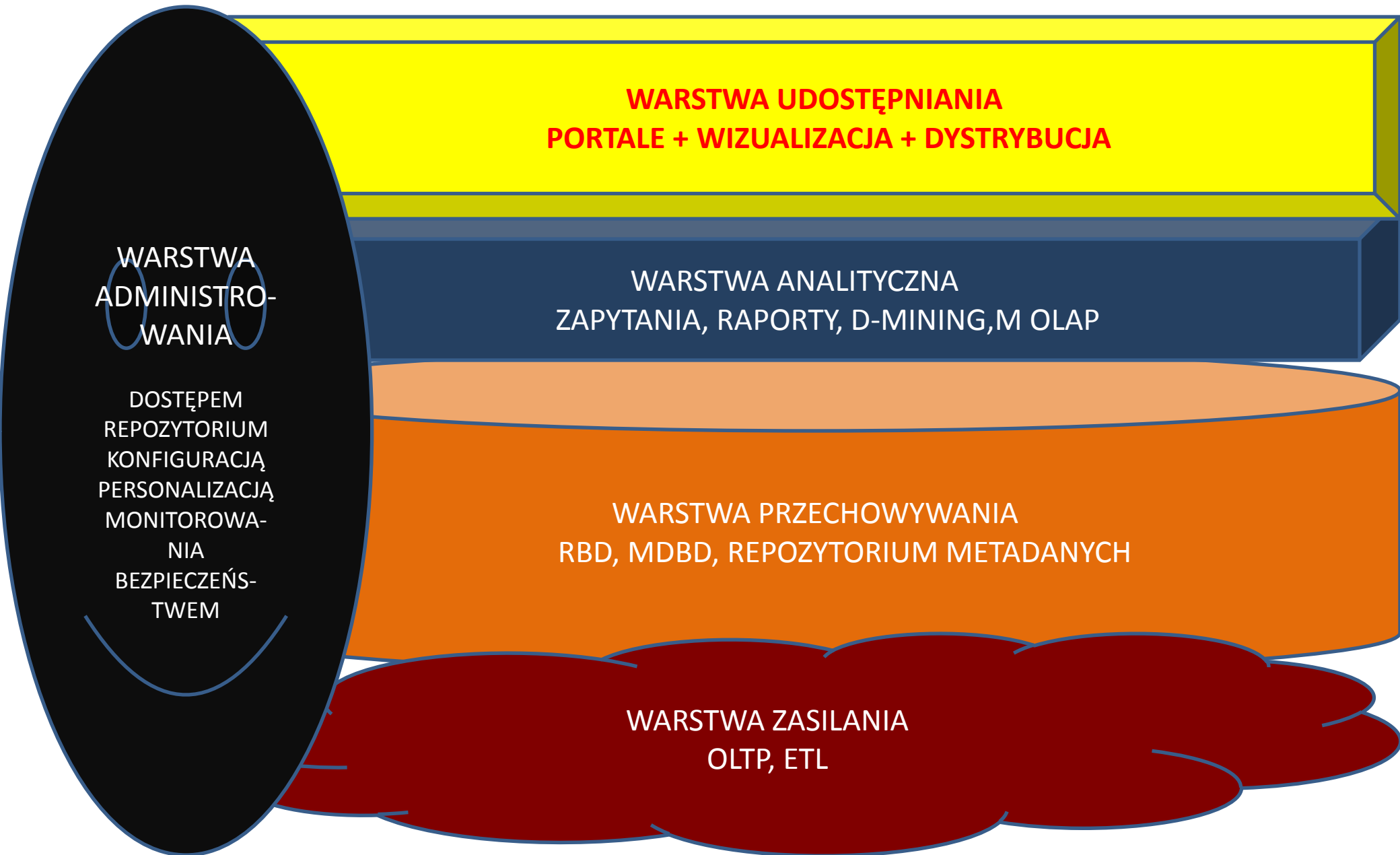
HURTOWNIA(AGREGACJE OLAP)

PROCESY I NARZĘDZIA ETL

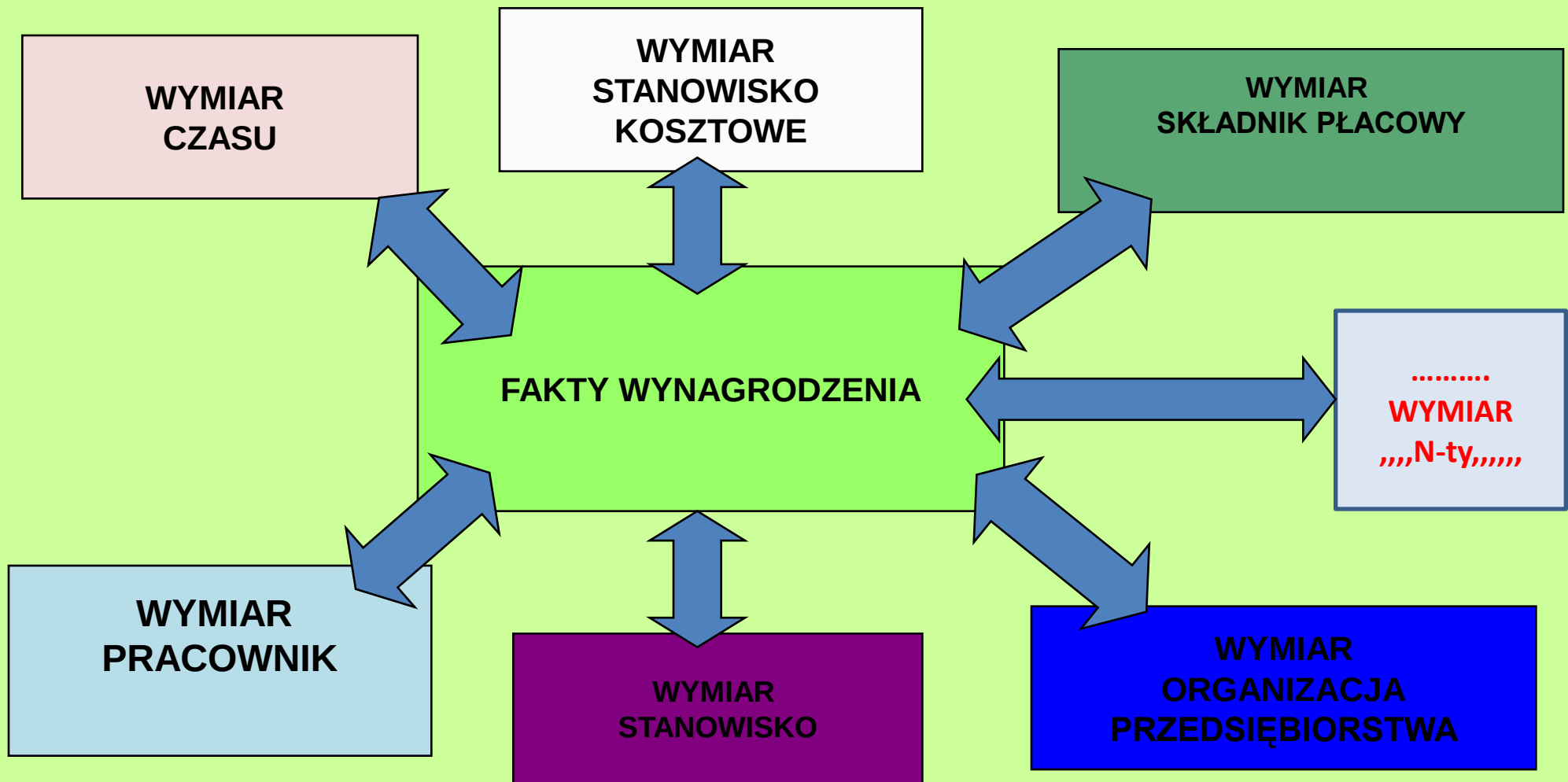
CZĘŚĆ
DO
STĘP
U

ZŁOŻONOŚĆ NARZĘDZI

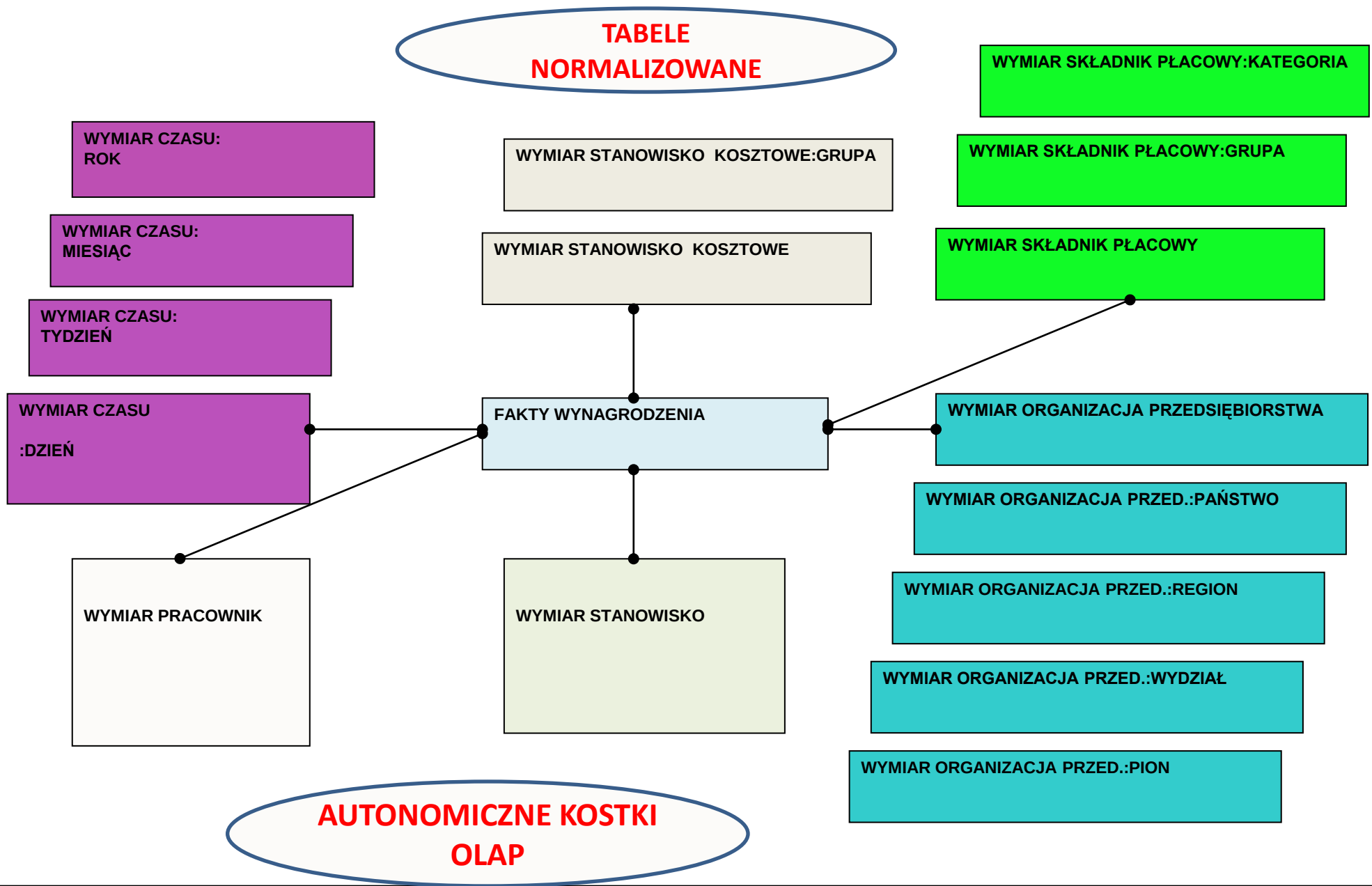
UNIWERSALNA ARCHITEKTURA BI



Generowanie informacji wielowymiarowej



Generowanie informacji wielowymiarowej - hierarchicznej



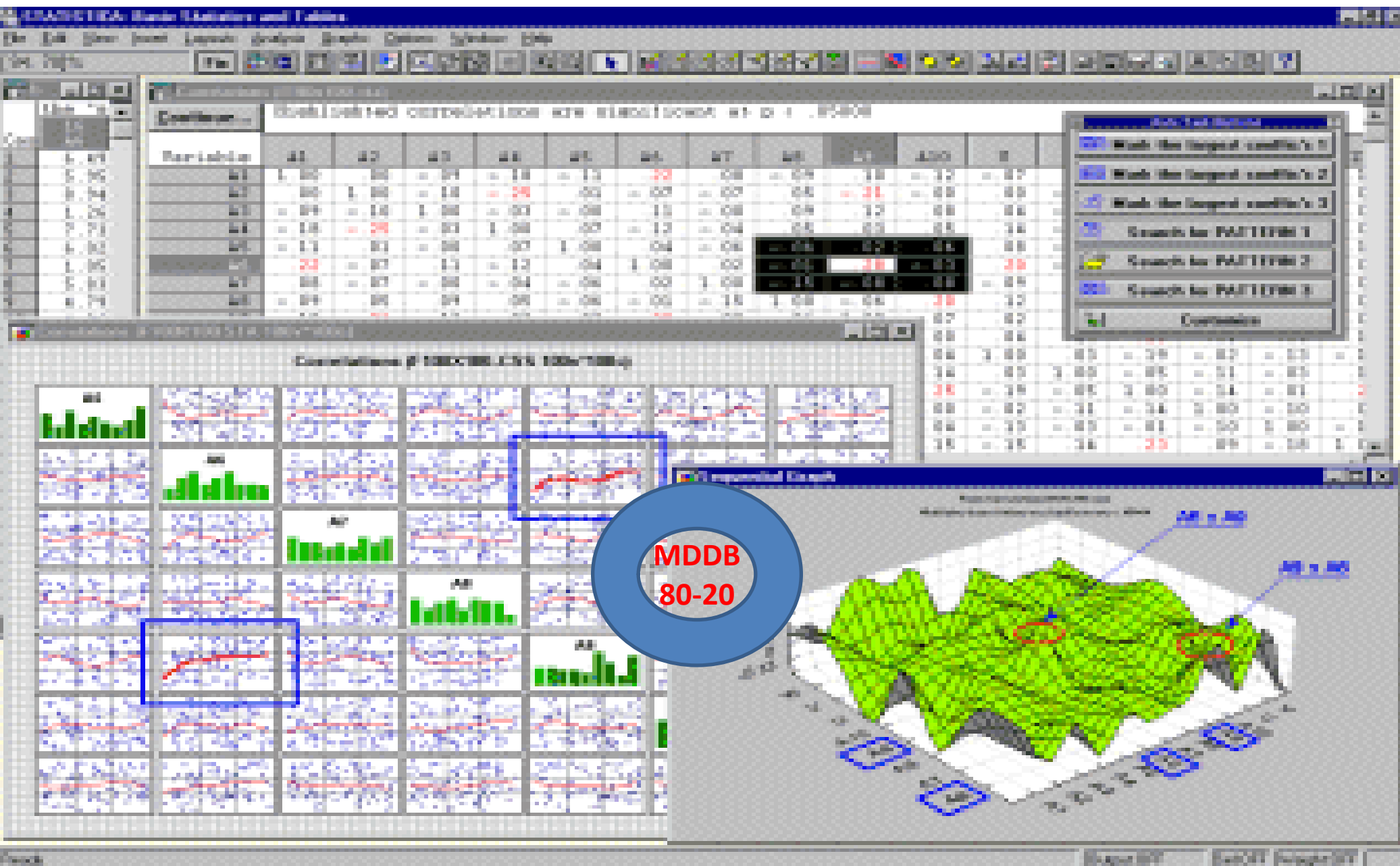
OLAP – DATA MINING

- OLAP – WIEMY, CZEGO NIE WIEMY
- DATA MINING – NIE WIEMY, CZEGO NIE WIEMY np.
 - IBM INTELLIGENT MINER FOR DATA
 - SAS ENTERPRICE MINER
 - ORACLE 9i DATA MINING

METODY DATA MINING

- EKSPLORACYJNA ANALIZA DANYCH
=
TWORZENIE HIPOTEZ NA PODSTAWIE
OGŁĄDANYCH DANYCH W CELU
POSZUKIWANIA WZORCA = WYKRESY
ROZRZUTU

WYPEŁNIENIE KOSTKI OLAP



METODY DATA MINING

- MODELOWANIE OPISOWE

=

MODEL CAŁOŚCIOWEGO ROZKŁADU
PRAWDOPODOBIEŃSTWA, ZWIĄZKI
MIĘDZY ZMIENNYMI, KLASY, SEGMENTY,
SKUPIENIA (KLASTRY)

METODY DATA MINING

- MODELOWANIE PREDYKCYJNE

=

PRZEWIDYWANIE WARTOŚCI,
MODELE REGRESJI, KLASYFIKACJI DANYCH
(OBIEKTÓW) NP. DRZEWA DECYZYJNE, b-
DRZEWA i tp.

METODY DATA MINING

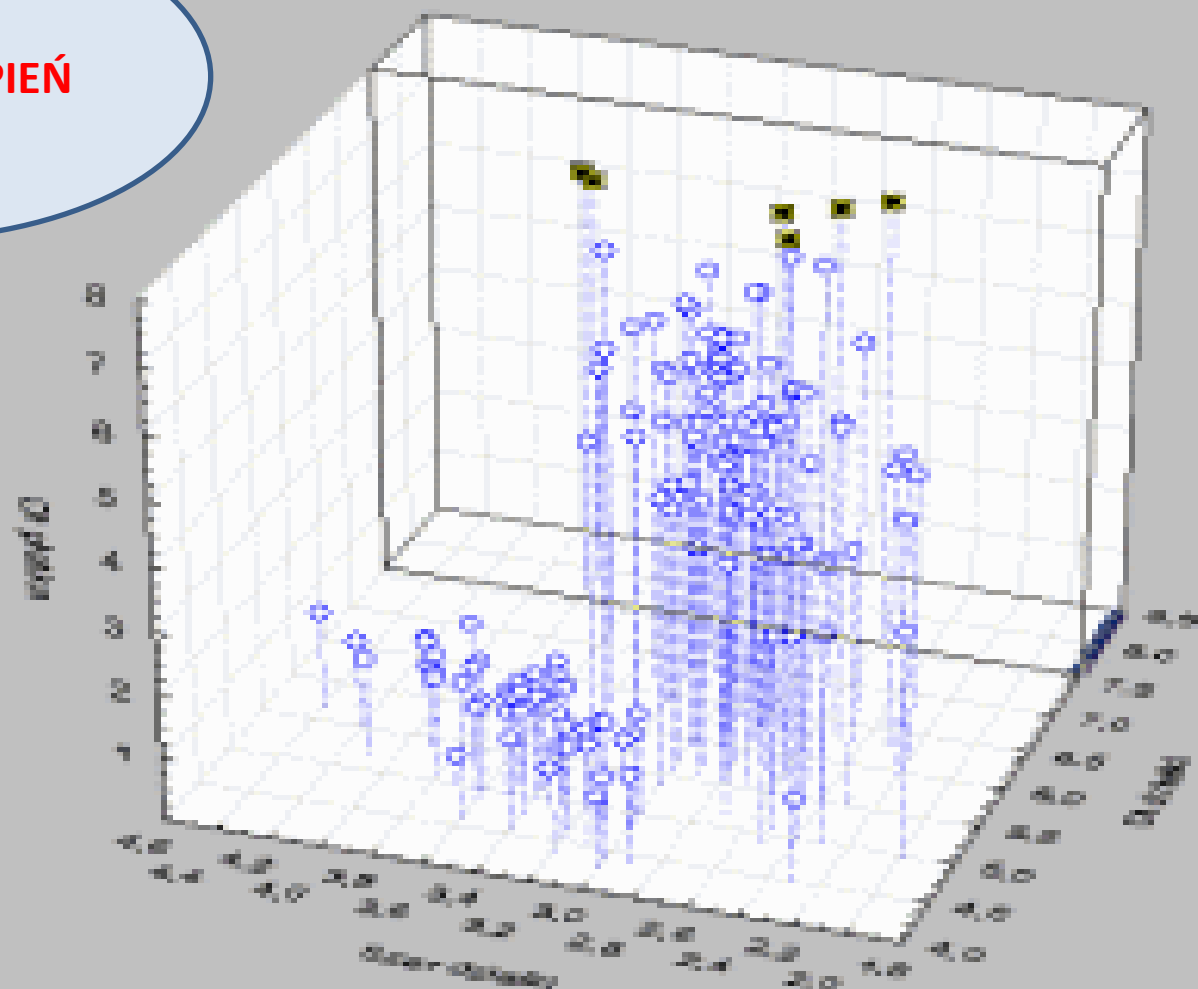
- ODKRYWANIE WZORCÓW I REGUŁ

=

WEDŁUG WYBRANEGO ASPEKTU,
PODOBIENSTWO SKOJARZEŃ

KOSTKA OLAP

ANALIZA SKUPIEŃ



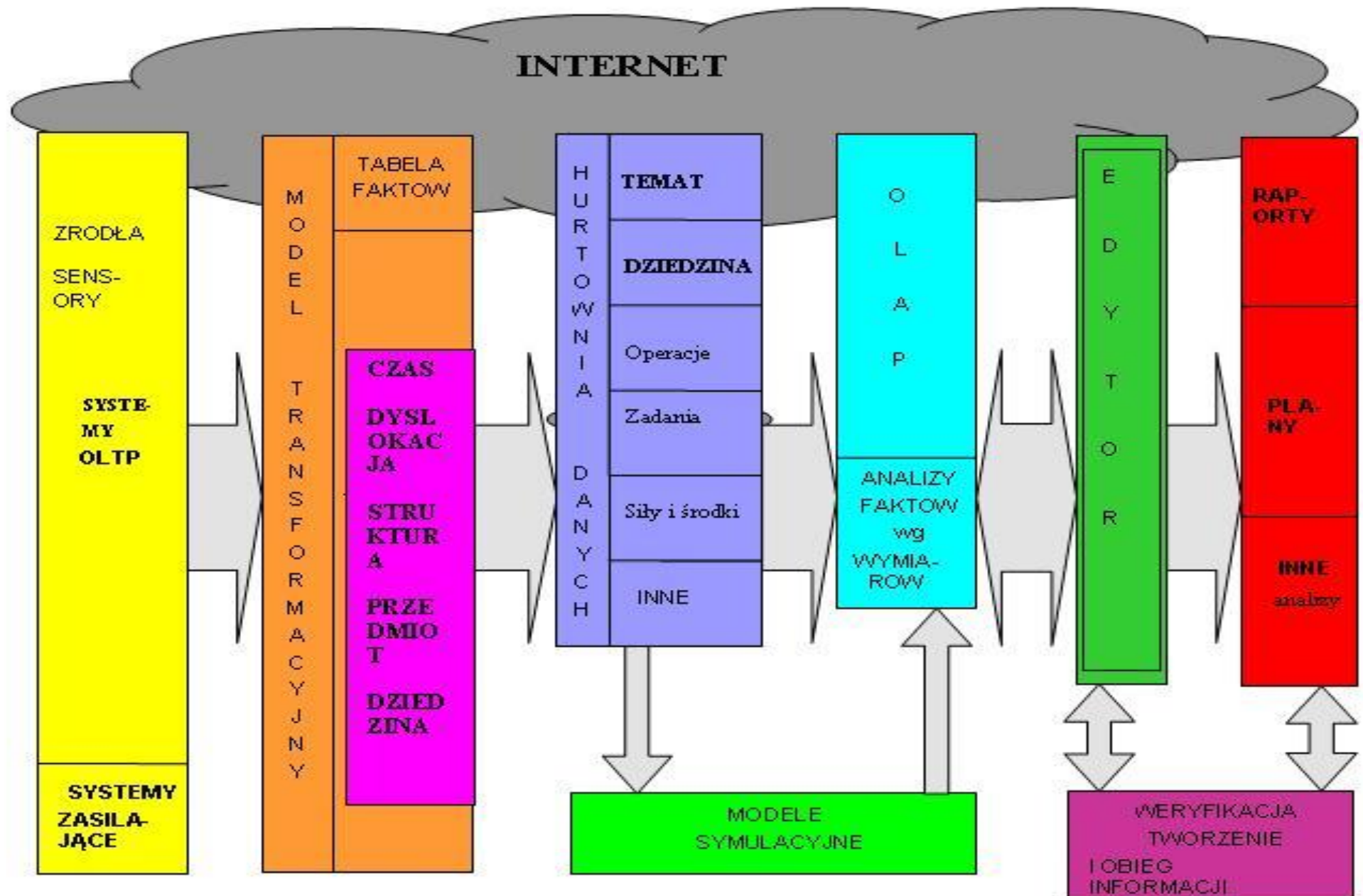
METODY DATA MINING

- WYSZUKIWANIE WG
ZAWARTOŚCI/WZORCA

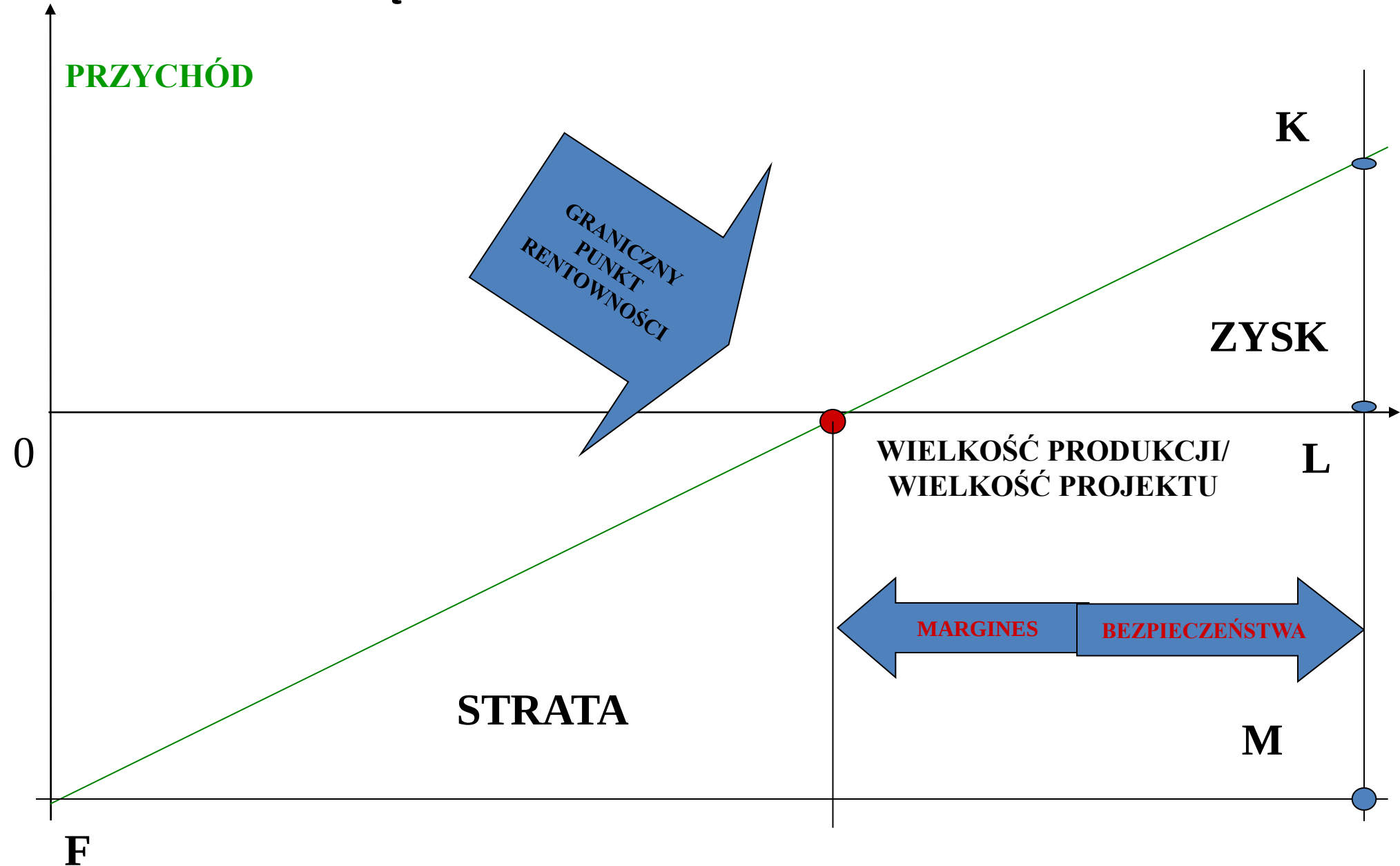
=

TEKSTY, OBRAZY, SZEREGI CZASOWE,
DOWOLNE DANE,
PODOBIENSTWO DO WZORCA

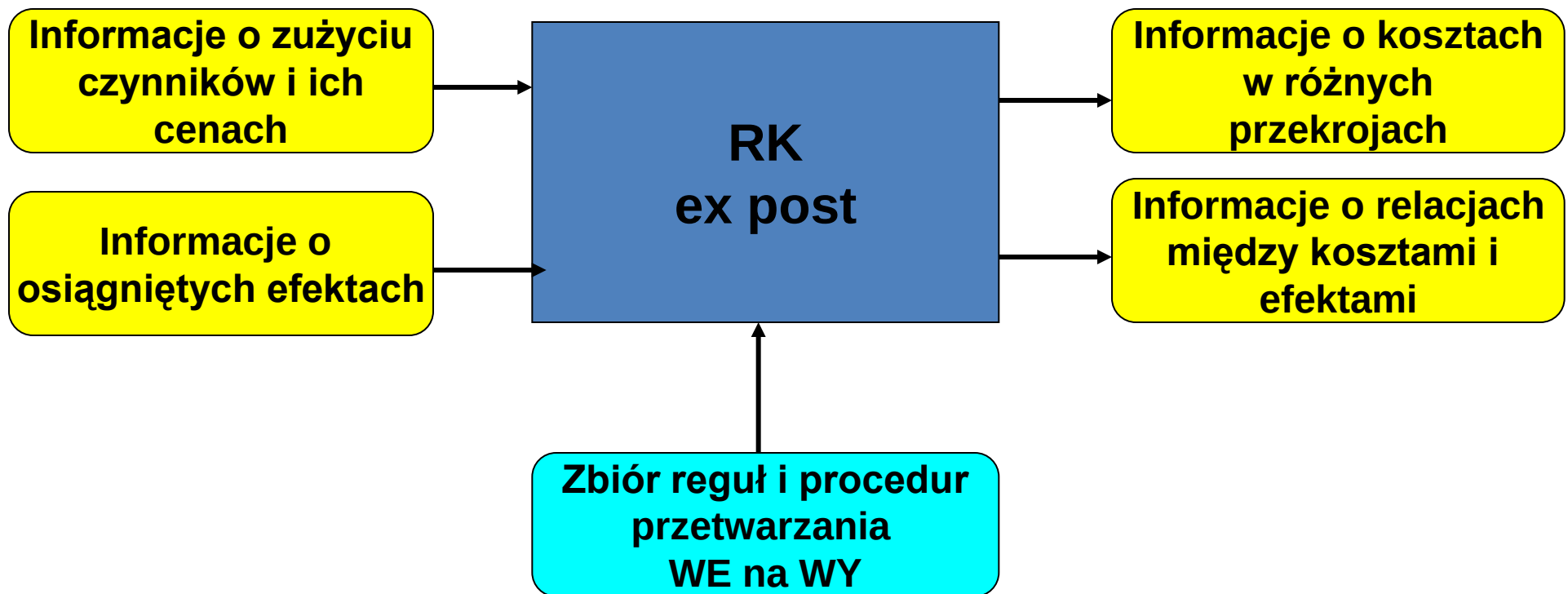
X-ENGINEERING



ZARZĄDZANIE MARGINESEM BEZPIECZEŃSTWA



Rachunek kosztów ex post



Rachunkowość zarządcza i controlling (1)

Rachunkowość zarządczą można zdefiniować jako „system gromadzenia, opracowywania (klasyfikacji, syntetyzowania, analizy) i prezentacji informacji (finansowych i operacyjnych) dotyczących przeszłych i przyszłych zjawisk gospodarczych w celu wspomagania kierownictwa jednostki gospodarczej w planowaniu, podejmowaniu decyzji i kontroli”

Posługując się tą definicją można wyodrębnić następujące funkcje obsługiwane przez rachunkowość zarządczą w przedsiębiorstwie, tj.:

- gromadzenie informacji,
- opracowywanie informacji,
- prezentacja informacji,
- orientacja na przeszłość i przyszłość,
- wspomaganie planowania,
- wspomaganie podejmowania decyzji,
- wspomaganie kontroli.

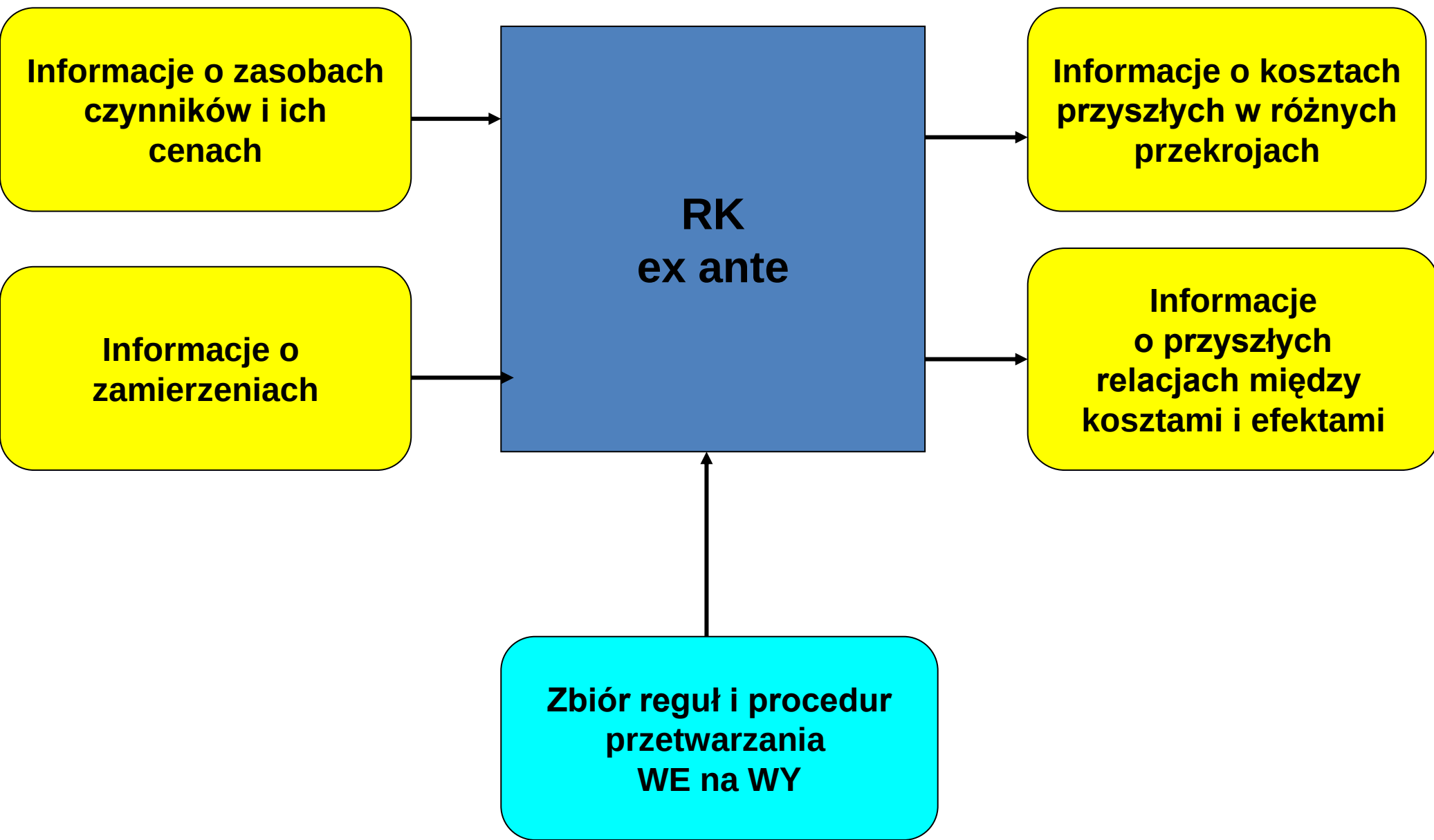
Rachunkowość zarządcza i controlling (2)

- Dzięki zastosowaniu wielowymiarowych kostek OLAP wspomagana jest funkcja prezentacji informacji. W szczególności wiąże się to z ułatwieniem dostępu Użytkownikom do informacji zawartych w kostkach poprzez końcowe aplikacje klienckie, co objawia się m.in. szybkością dostępu, intuicyjnością wielowymiarowego modelu danych, jak również wysoką elastycznością graficznych interfejsów samych aplikacji końcowych.
- Orientacja na przeszłość bezpośrednio wynika z historycznego charakteru danych źródłowych będących przedmiotem przetwarzania w architekturze OLAP. Odniesienie do przyszłości odbywa się natomiast m.in. poprzez możliwość łatwego konstruowania modeli symulacyjnych i prognostycznych na bazie danych historycznych oraz zapisania wyników przetwarzania w celu późniejszego wykorzystania np. na potrzeby planowania.

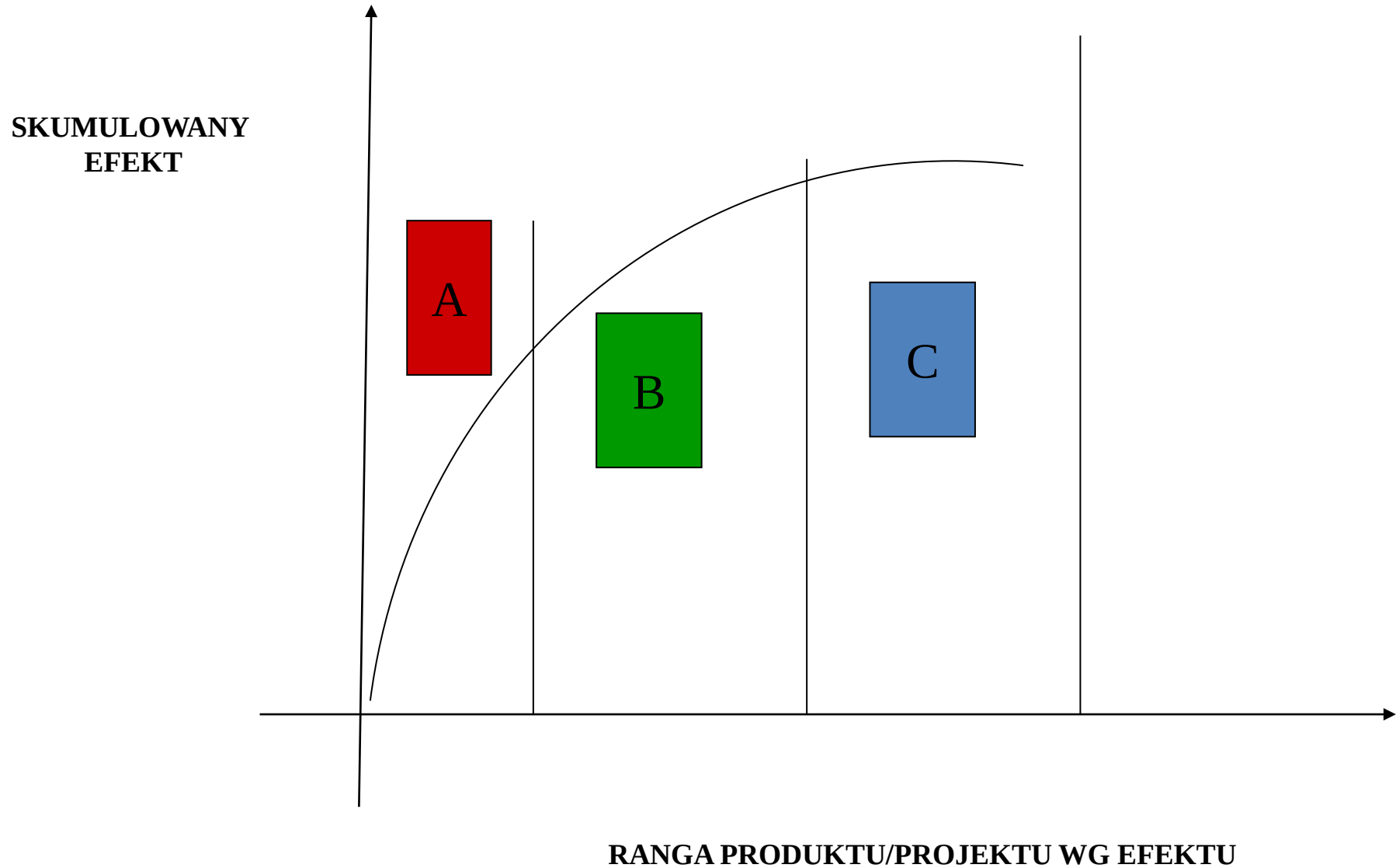
Rachunkowość zarządcza i controlling (3)

- Istotną korzyścią zastosowania technologii OLAP w ujęcia problemów rachunkowości zarządczej jest wielowymiarowe podejście w reprezentacji informacji (ang. *multidimensional view*).
- Dzięki temu modele analityczne zyskują na przejrzystości i intuicyjności, co ma bezpośredni wpływ na efektywność procesu podejmowania decyzji. Podejście to zakłada jawne wyodrębnienie miar, reprezentujących wielkości badanych zjawisk oraz wymiarów, będących kryteriami analizy tych zjawisk.

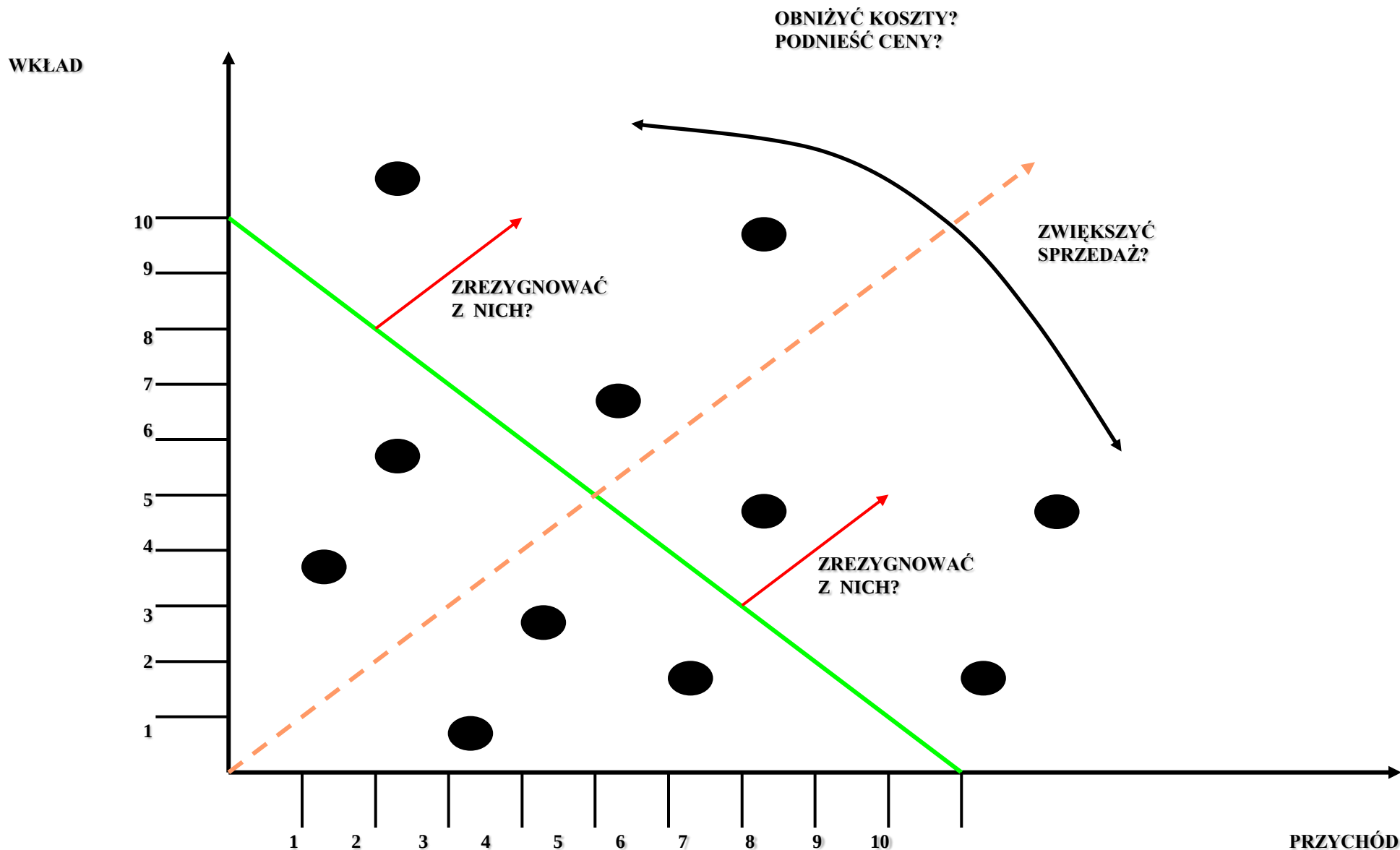
Rachunek kosztów ex ante



ANALIZA PARETO = ABC = 80-20



RÓŻNORODNOŚĆ = ANALIZA WARTOŚCI



Planowanie DUMPINGU

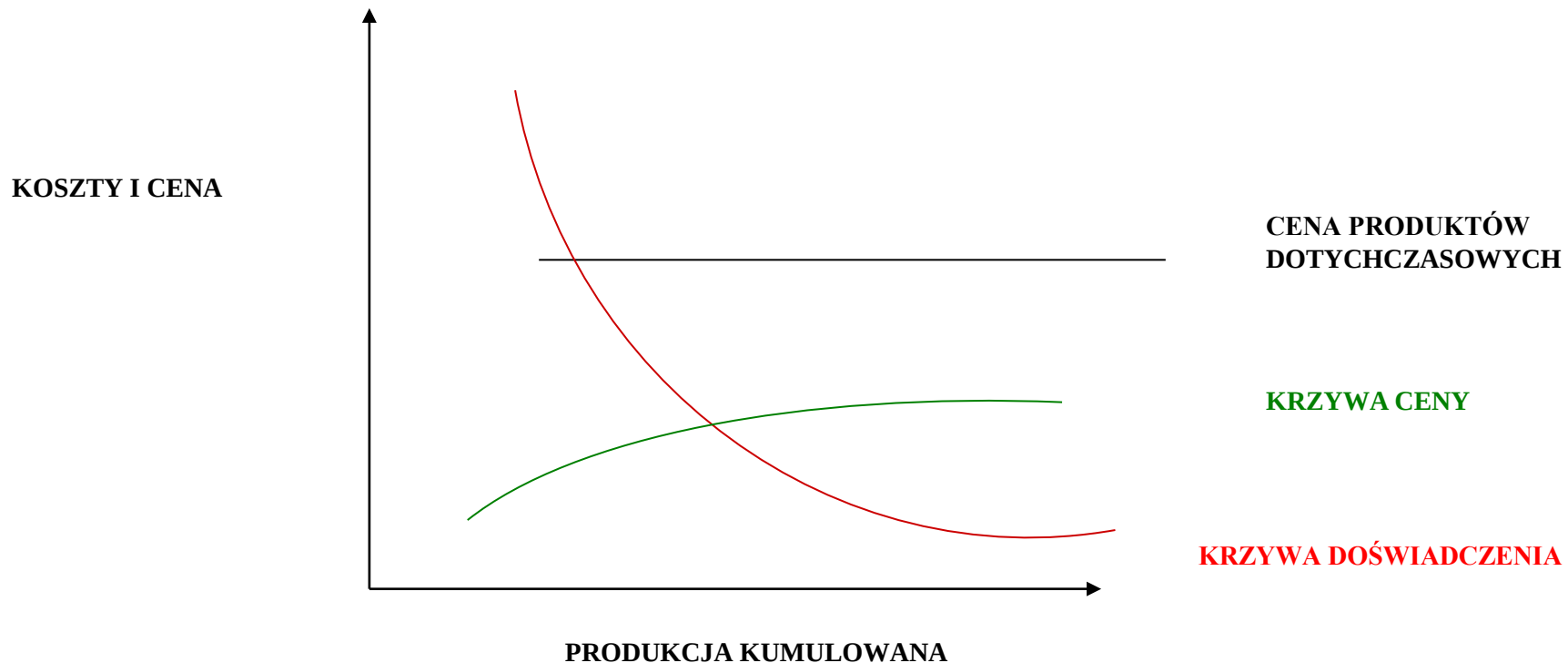
ISTOTA: AKCEPTACJA POCZĄTKOWYCH STRAT

CEL:

WPROWADZENIE NA RYNEK PRODUKTU SUBSTYTUCYJNEGO.

SZYBKI WZROST SPRZEDAŻY

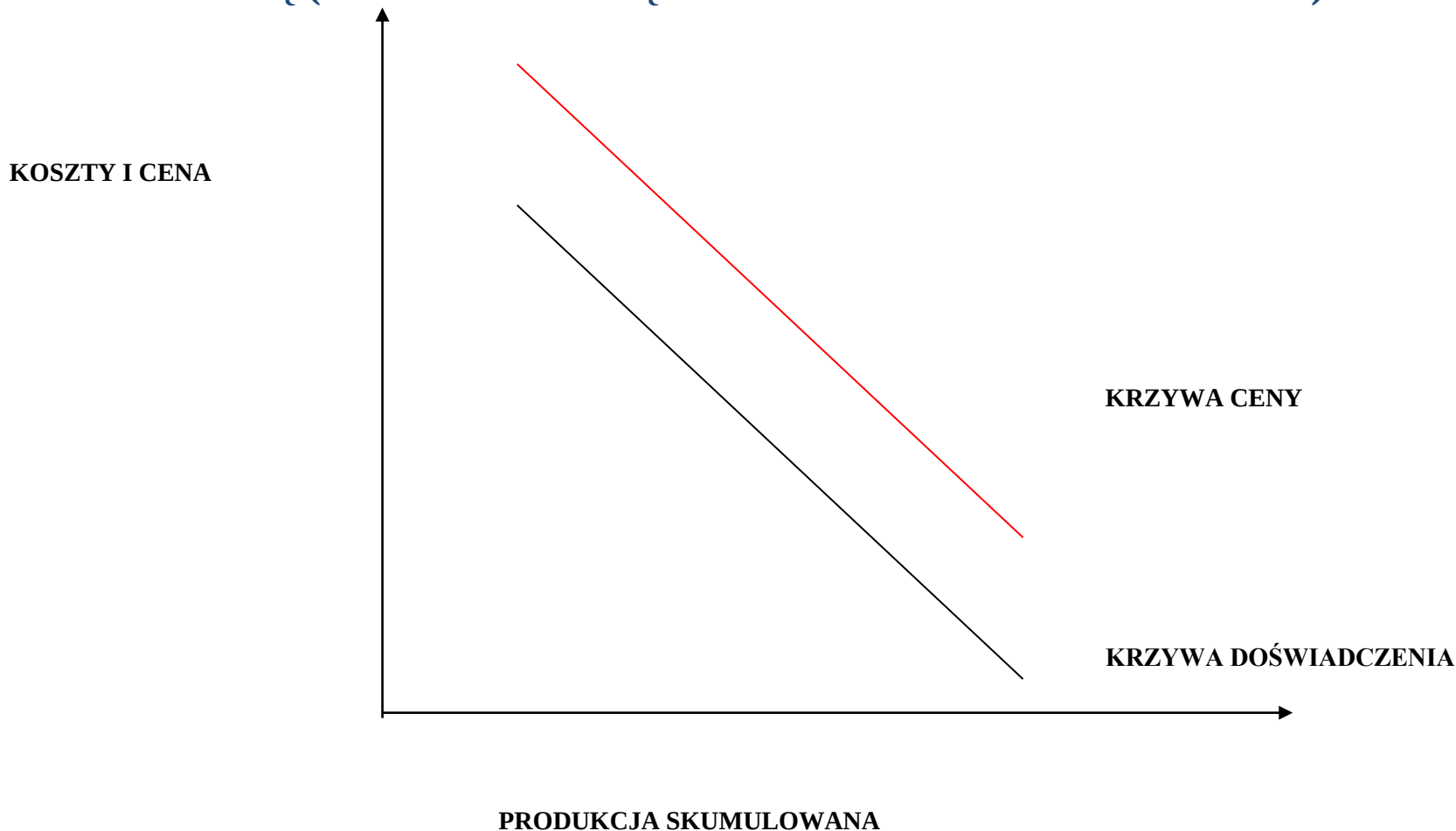
UZYSKANIE SZYBKIEGO EFEKTU DOŚWIADCZENIA.



Planowanie STRATEGII DOMINACJI

ODZWIERCIEDLENIE **KOSZTÓW W CENACH**.

SYSTEMATYCZNE **KORYGOWANIE** POZIOMU CENY ZGODNIE Z FORMUŁĄ KOSZTOWĄ (KOSZT PRZECIĘTNY PLUS OKREŚLONA MARŻA).



Planowanie STRATEGII PARASOLA

ISTOTA: CZASOWE UTRZYMYWANIE CENY NA NIEZMIENIONYM POZIOMIE

CEL: ZWIĘKSZENIE MARŻY ZYSKU.

SZYBKI WZROST ZAINWESTOWANEGO KAPITAŁU

OBNIŻENIE RYZYKA FINANSOWEGO (OCHRONA PRZED RYZYKIEM)

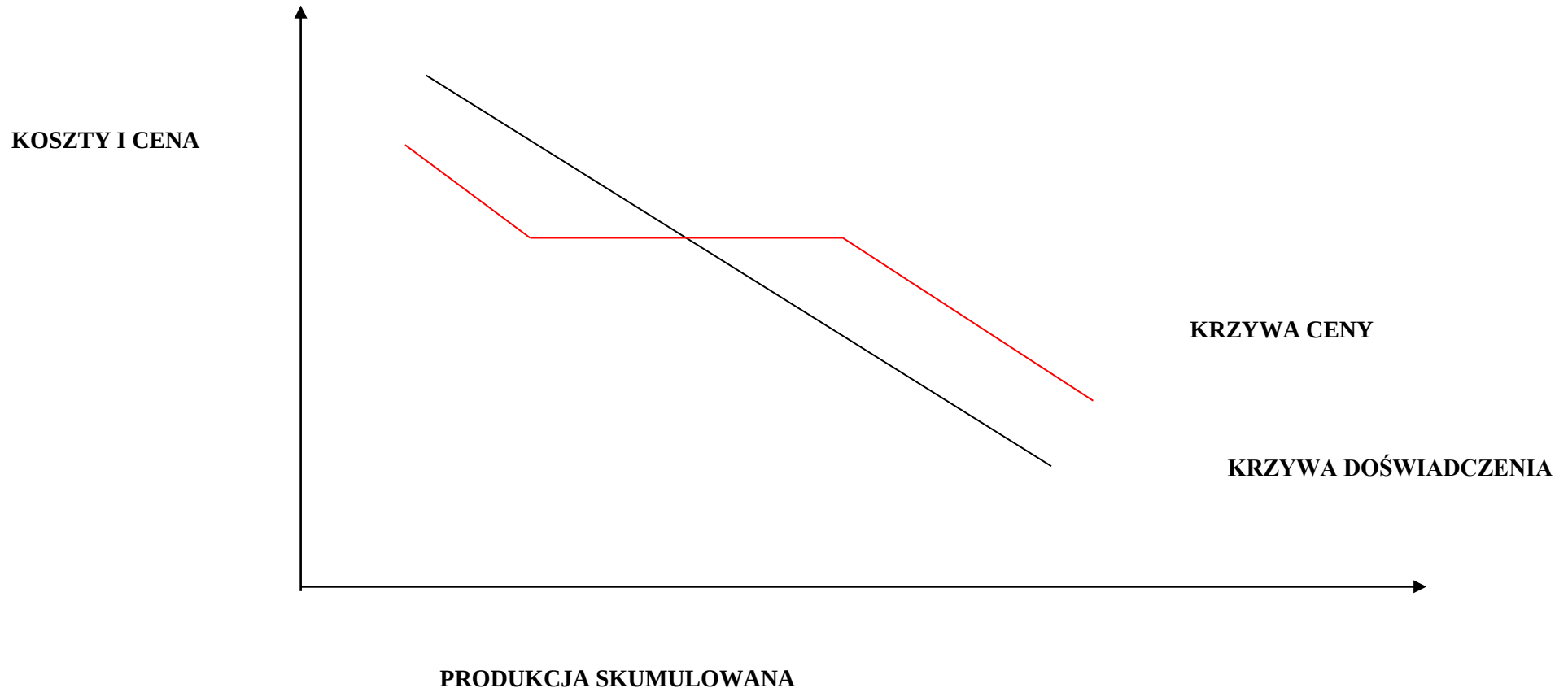
WADY: NIEBEZPIECZNA W WARUNKACH **WOJNY CENOWEJ**, WYMUSZAJĄCEJ OBNIŻENIE CENY
W ŚLAD ZA OBNIŻKĄ KOSZTÓW JEDNOSTKOWYCH.



Walidacja STRATEGII PRZECHWYCENIA

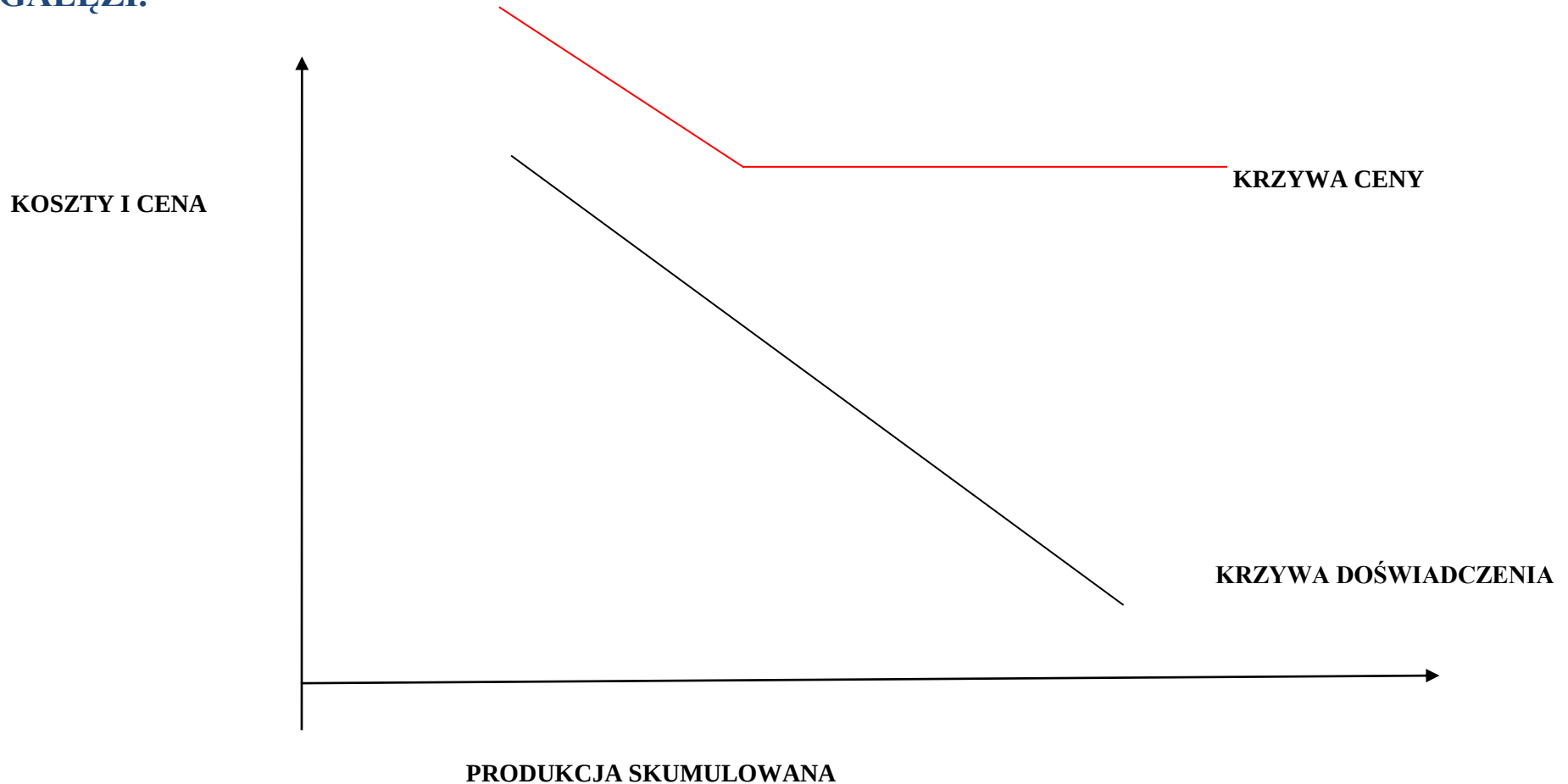
PRZEJĘCIE RYNKU.

STRATEGIA TAKA JEST CZĘSTO STOSOWANA PRZEZ FIRMY ZNAJDUJĄCE SIĘ W **NIEKORZYSTNEJ POZYCJI** WYJŚCIOWEJ I DOTYCZY NA OGÓŁ PODMIOTÓW O MARGINALNYM ZNACZENIU NA RYNKU.



Walidacja STRATEGII PORZUCENIA

WYCOFANIE SIĘ Z RYNKU Z JEDNOCZESNYM MAKSYMALIZOWANIEM RENTOWNOŚCI.
STRATEGIA DZIAŁANIA **FIRM O SŁABEJ POZYCJI RYNKOWEJ**, PRZYGOTOWUJĄCYCH SIĘ
DO TZW. "ŻNIW" W CELU SKIEROWANIA WYGENEROWANEGO KAPITAŁU DO INNYCH
GAŁĘZI.



GEOANALIZY

SWPP

Modele
analityczne

Modele decyzyjne

GEOINFOMACJA

PGO (GIS)



SERWER ANALITYCZNY

SERWER MAP CYFROWYCH

PROLOG = wnioskowanie

- PROGRAM = ALGORYTM +
DANE
- ALGORYTM = LOGIKA +
STEROWANIE

Wnioski

- HD = ORGANIZACJA DANYCH + analizy
- OLAP = EKSPLOZJA DANYCH
- DM = ODKRYWANIE WIEDZY - BI
- OLAP/DM wspomaga procesy decyzyjne
- Umożliwia prognozowanie i planowanie wielowymiarowe
- Integruje funkcjonowanie organizacji
- Umożliwia współdziałanie organizacji sieciowych i wirtualnych

LITERATURA PODSTAWOWA

1. **JANUSZEWSKI A.: FUNKCJONALNOŚĆ INFORMATYCZNYCH SYSTEMÓW ZARZĄDZANIA. TOM I i II. WN PWN, Warszawa 2008.**
2. **POE V., LAUER P. , BROBST S.: TWORZENIE HURTOWNI DANYCH. WSPOMAGANIE PODEJMOWANIA DECYZJI. WNT, Warszawa 2000.**
3. **STURM J. : HURTOWNIE DANYCH. SQL SERVER 7.0, PRZEWODNIK TECHNICZNY. MICROSOFT. 2000.**
4. **ZASKÓRSKI P.: STRATEGIE INFORMACYJNE W ZARZĄDZANIU ORGANIZACJAMI GOSPODARCZYMI. WAT, Warszawa 2005.**



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI



WARSZAWSKA
WYŻSZA SZKOŁA
INFORMATYKI

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY

