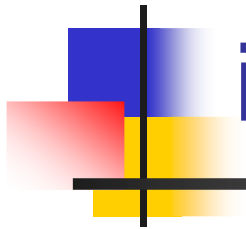
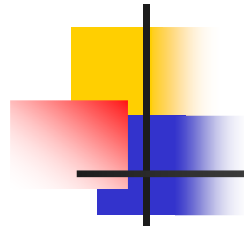


O systemach wspomagania decyzji, bazach danych i hurtowniach danych





Klasyfikacja decyzji

- Operacyjne i strategiczne - ze względu na ich wpływ na działalność organizacji.
- Definiowalne i trudno definiowalne - ze względu na ich strukturę.
- Zależne i niezależne - ze względu na ich odniesienie do pozostałych decyzji w organizacji.



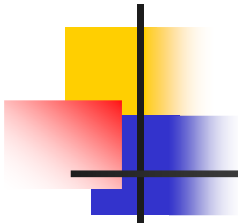
Rodzaje systemów informatycznych w obiekcie gospodarczym (1)

- Systemy ewidencyjne - tworzą zbiory elementarnych danych o funkcjonowaniu obiektu.
- Systemy informowania kierownictwa - informują o stanie obiektu i jego poszczególnych parametrach.
- Systemy wspomaganie decyzji - analiza strategicznych problemów decyzyjnych.
- Systemy ekspertowe - „samouczące się”, dokonują inteligentnej analizy danych, tworzone na podstawie koncepcji sztucznej inteligencji.

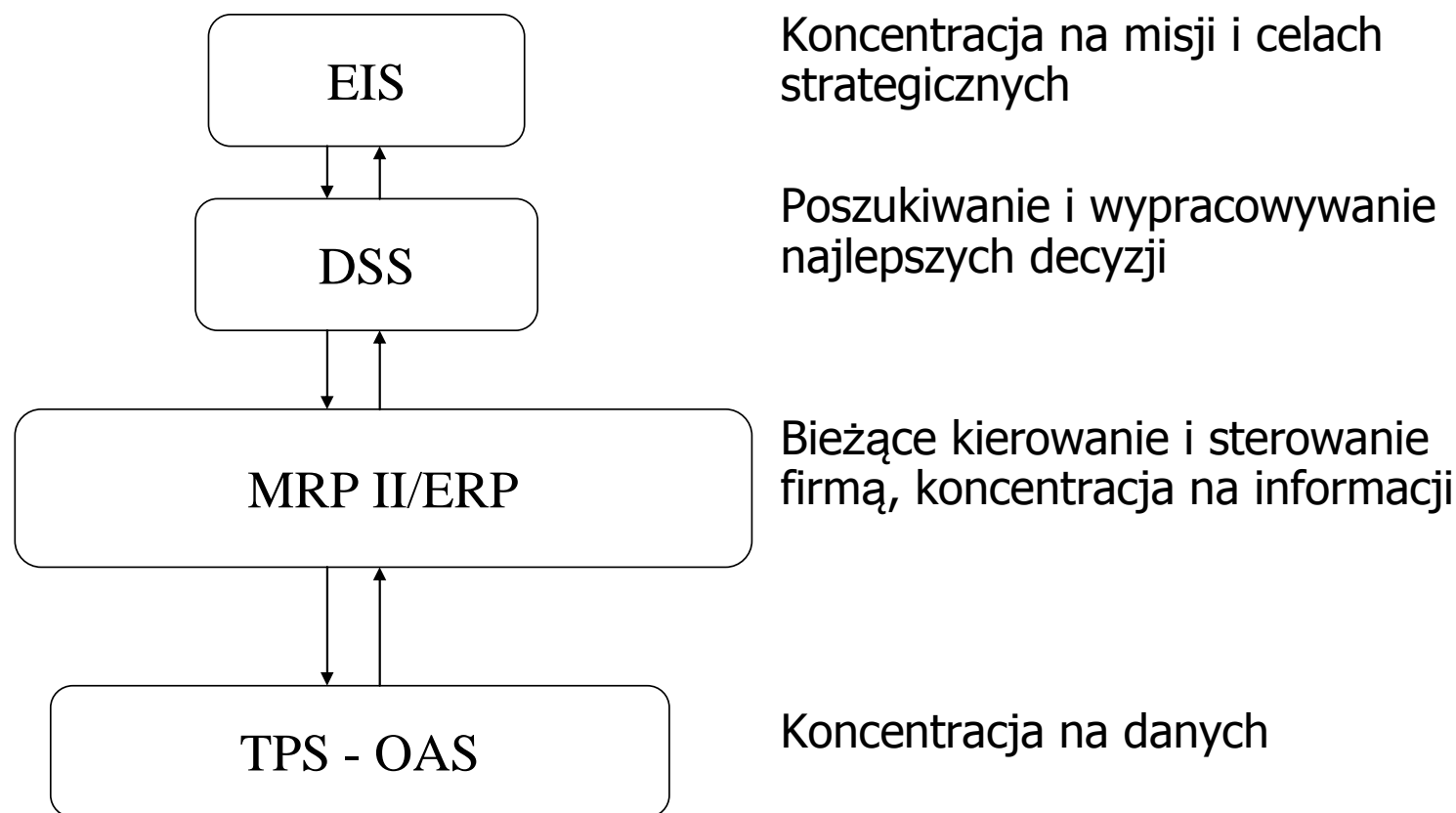


Rodzaje systemów informatycznych w obiekcie gospodarczym (2)

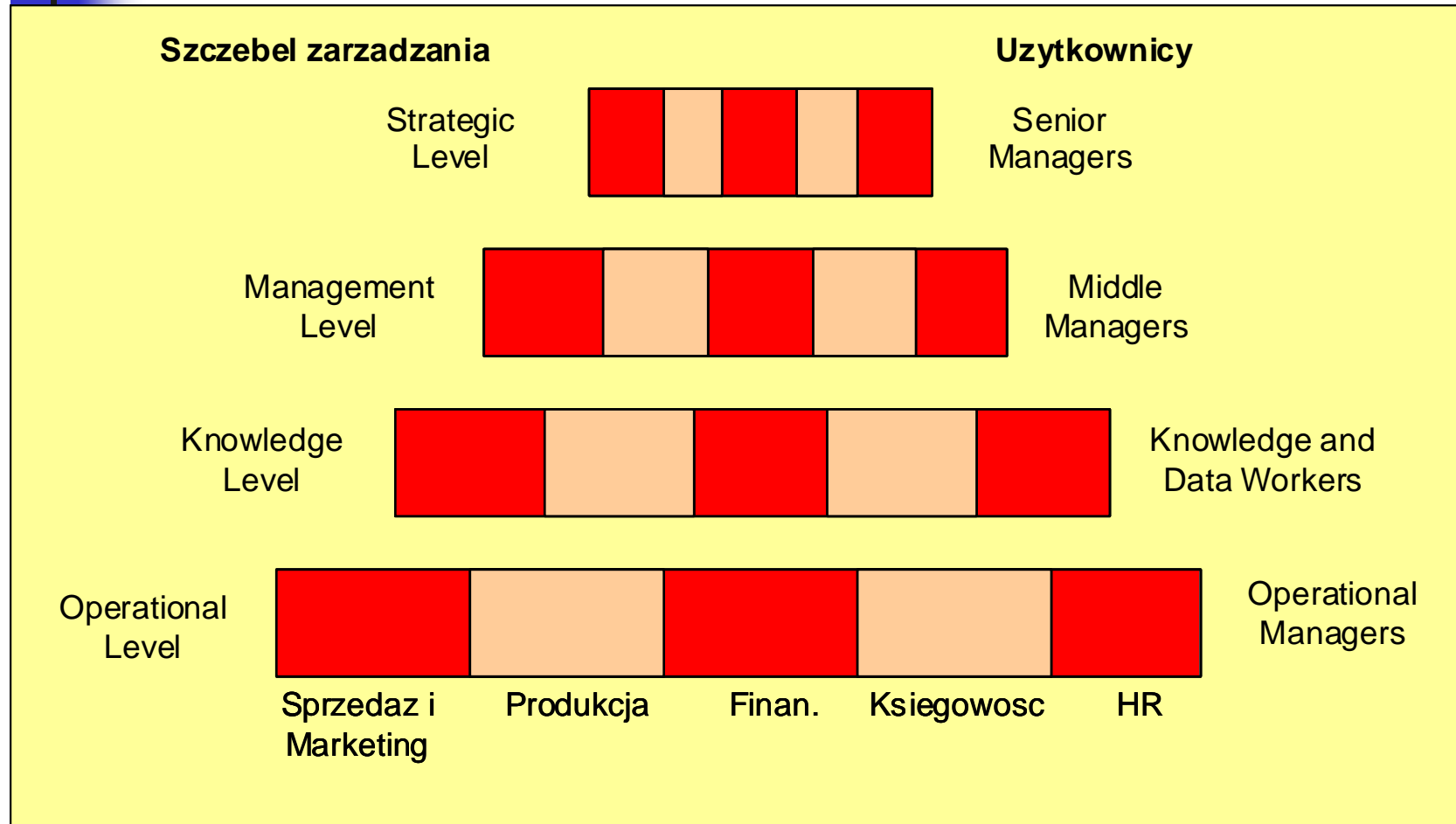
- Systemy transakcyjne (Transaction Processing Systems - TPS).
- Systemy nowoczesnego biura (Office Automation Systems - OAS).
- Systemy wspomagania zarządzania produkcją (Manufacturing Resource Planning - MRP II, Enterprise Resource Planning - ERP).
- Systemy informacyjne zarządzania (Management Information Systems - MIS).
 - Systemy wspomagania decyzji (Decision Support Systems - DSS).
 - Systemy informowania kierownictwa (Executive Information Systems - EIS).



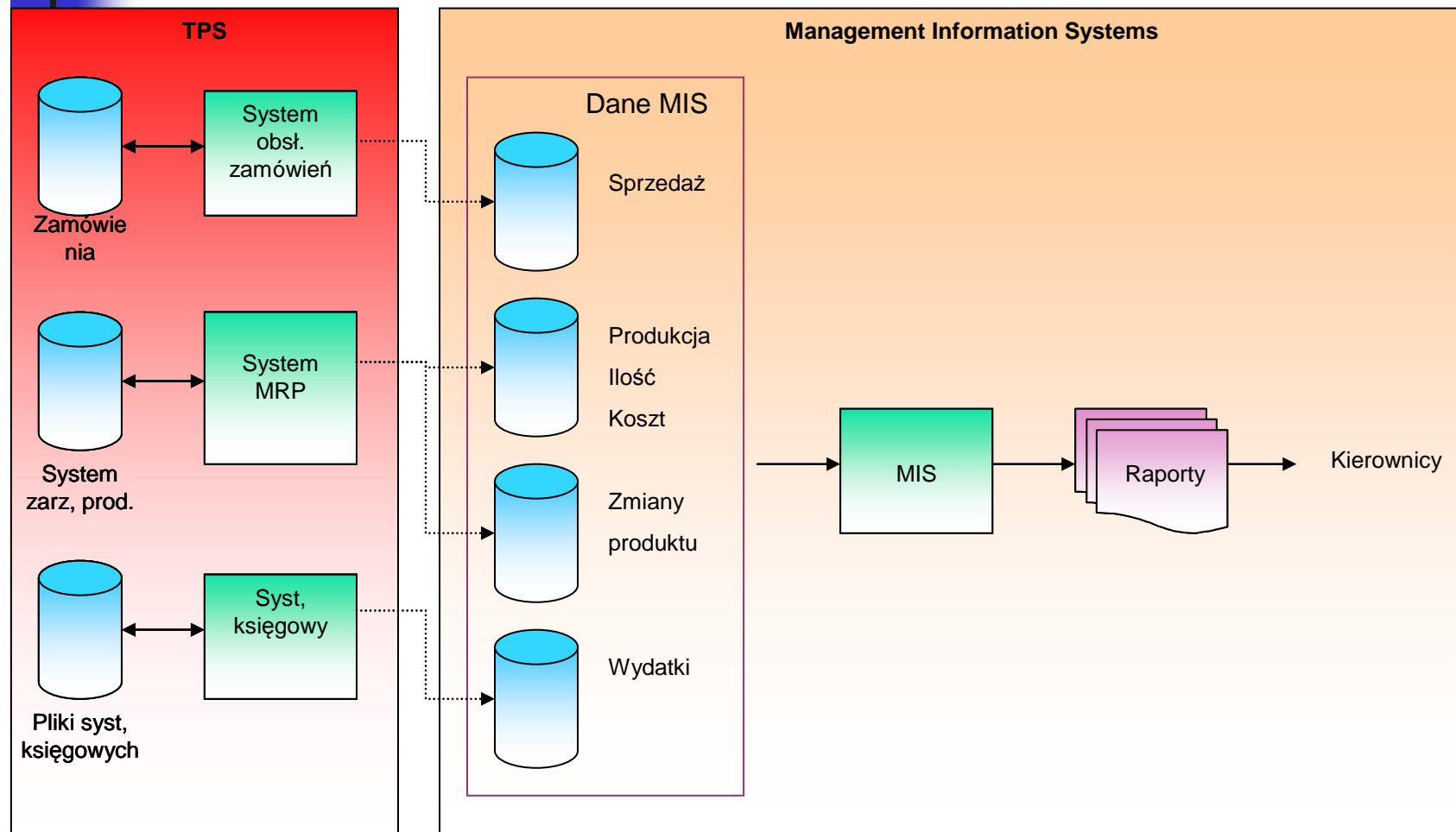
Relacje między systemami informatycznymi w organizacji (1)



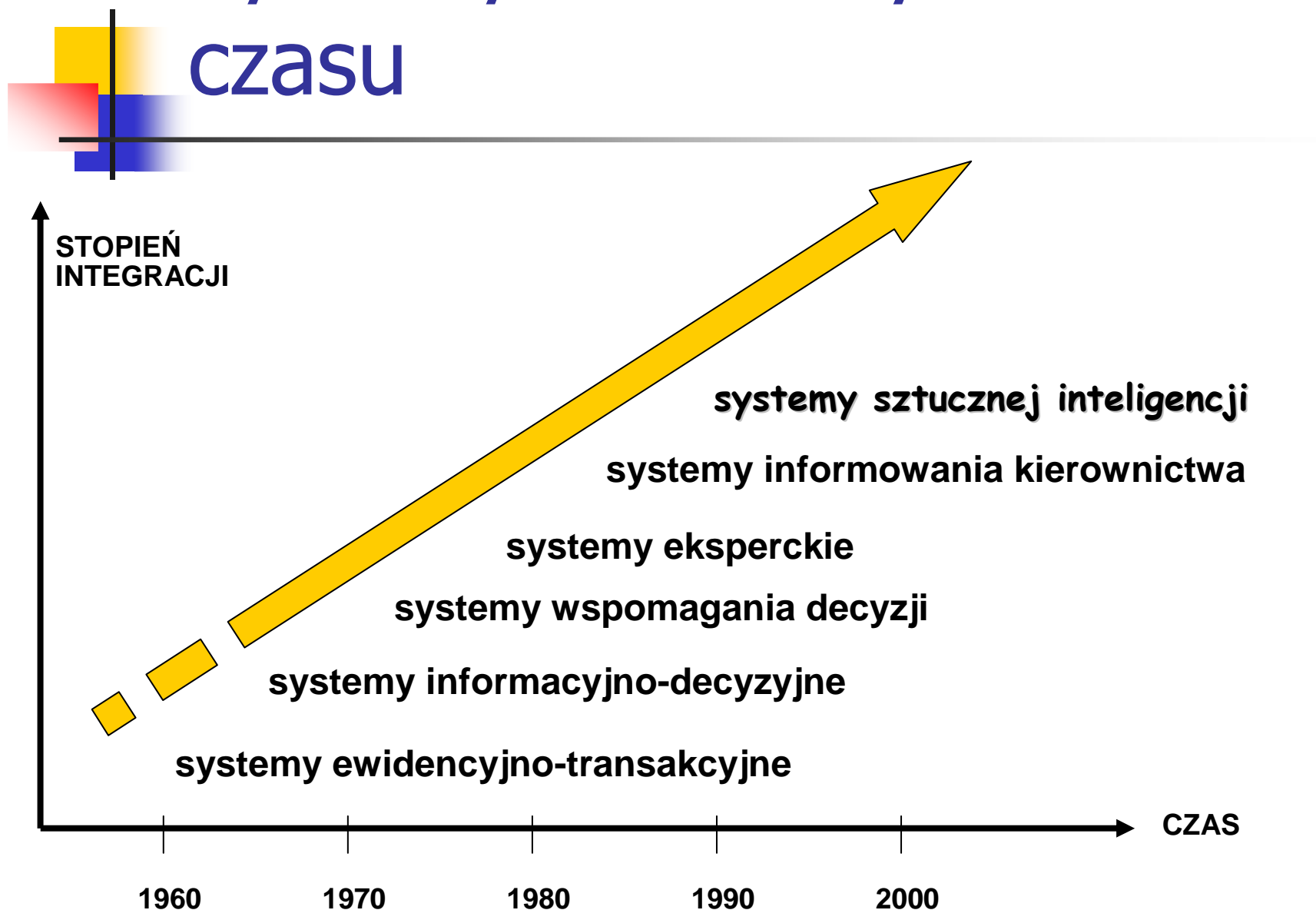
Podział systemów wg szczebla zarządzania



Relacje między systemami informatycznymi w organizacji (2)




Systemy informatyczne na osi czasu



Kryteria oceny jakości informacji (1)

- Dyspozycyjność - możliwość dostępu do każdej informacji w systemie w dowolnym momencie.
- Aktualność - wiarygodność informacji w danym momencie.
- Rzetelność - dokładność i poprawność.
- Porównywalność - zdolność przeprowadzenia analiz porównawczych.



Kryteria oceny jakości informacji (2)

- Informacja powinna być:
 - wiarygodna (określony stopień pewności, że jest ona prawdziwa)
 - relewatywna (istotna dla odbiorcy)
 - przyswajalna (nie wymagająca dalszych przekształceń)
 - dostępna w odpowiednio krótkim czasie
 - zapewniająca poufność (kodowanie, ograniczony dostęp, autoryzacja dostępu, ...)
 - gwarantująca bezpieczeństwo w przypadku awarii (bardziej chodzi tutaj o aspekt techniczny niż o samą informację)



Dobór systemów informatycznych (1)

- Realizacja potrzeb użytkownika
- Wymagania sprzętowe
- Spełnianie norm
- Serwis i gwarancja
- Referencje



Dobór systemów informatycznych (2a)

- Model rozwoju zastosowań informatyki w przedsiębiorstwie zakłada, że technologia informatyczna stanowi infrastrukturę:
 - wspomagającą istniejącą organizację,
 - przyczyniającą się do kształtowania strategii organizacji.



Dobór systemów informatycznych (2b)

- Stąd trzy typy organizacji infrastruktury technologii informatycznej:
 - Niezależny - tradycyjna rola wspomagania informatycznego (brak kontekstu strategicznego); nie ma wpływu na realizację i kształtowanie strategii organizacji; wspomaganie informatyczne traktuje się jako bieżące zabiegi administracyjne.
 - Reaktywny - infrastrukturę informatyczną projektuje się w celu wspomagania strategii działania, ale nie wpływa ona na kształtowanie strategii; rozwój informatyki jest pochodną planów działalności organizacji.
 - Współzależny - infrastruktura informatyczna jest powiązana z długofalowym planem strategicznym a nie w reakcji na plan; rozwój informatyki to inwestycja gospodarcza.



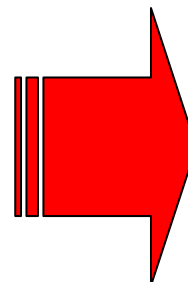
Samo życie...

- Stan zastosowań informatyki w Polsce:
 - nie zmienił istotnie organizacji pracy,
 - nie umożliwił integracji funkcji na wszystkich poziomach zarządzania,
 - nie spowodował poważnych zmian w pozycji firmy na rynku,
 - nie stworzył strategicznych szans dla firmy,
 - nie doprowadził do zmian w praktyce zarządzania oraz strukturach organizacyjnych firmy.

Roadmapa dla systemów informatycznych

- Pojedyncze systemy dziedzinowe: sprzedaż, kadry, płace, itp.
- Częstkowość rozwiązań
- Wielu dostawców komponentów systemu
- Niespójność funkcjonalna i słaba integracja
- Zależność od jednej platformy sprzętowej
- Słabe wspomaganie procesów zarządzania
- Brak rachunkowości zarządczej i kontrolingu
- Rosnące koszty utrzymania
- Ograniczone możliwości rozwojowe

STAN OBECNY



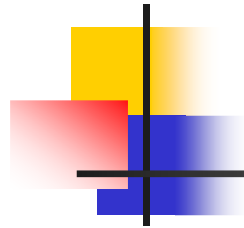
- Kompleksowość funkcjonalna
- Integracja danych i procedur
- Jeden dostawca - integrator rozwiązania
- Niezależność sprzętowo-programowa
- Wykorzystanie możliwości intra-, ekstra- i internetu oraz multimediów
- Pełne wspomaganie procesów zarządzania w ramach orientacji procesowej
- Realizacja rachunkowości zarządczej i kontrolingu zgodnie z systemem MRPII/ERP
- Uwzględnienie systemu jakości wg przyjętych standardów

STAN POŻĄDANY



Co jest eksploatowane w organizacjach

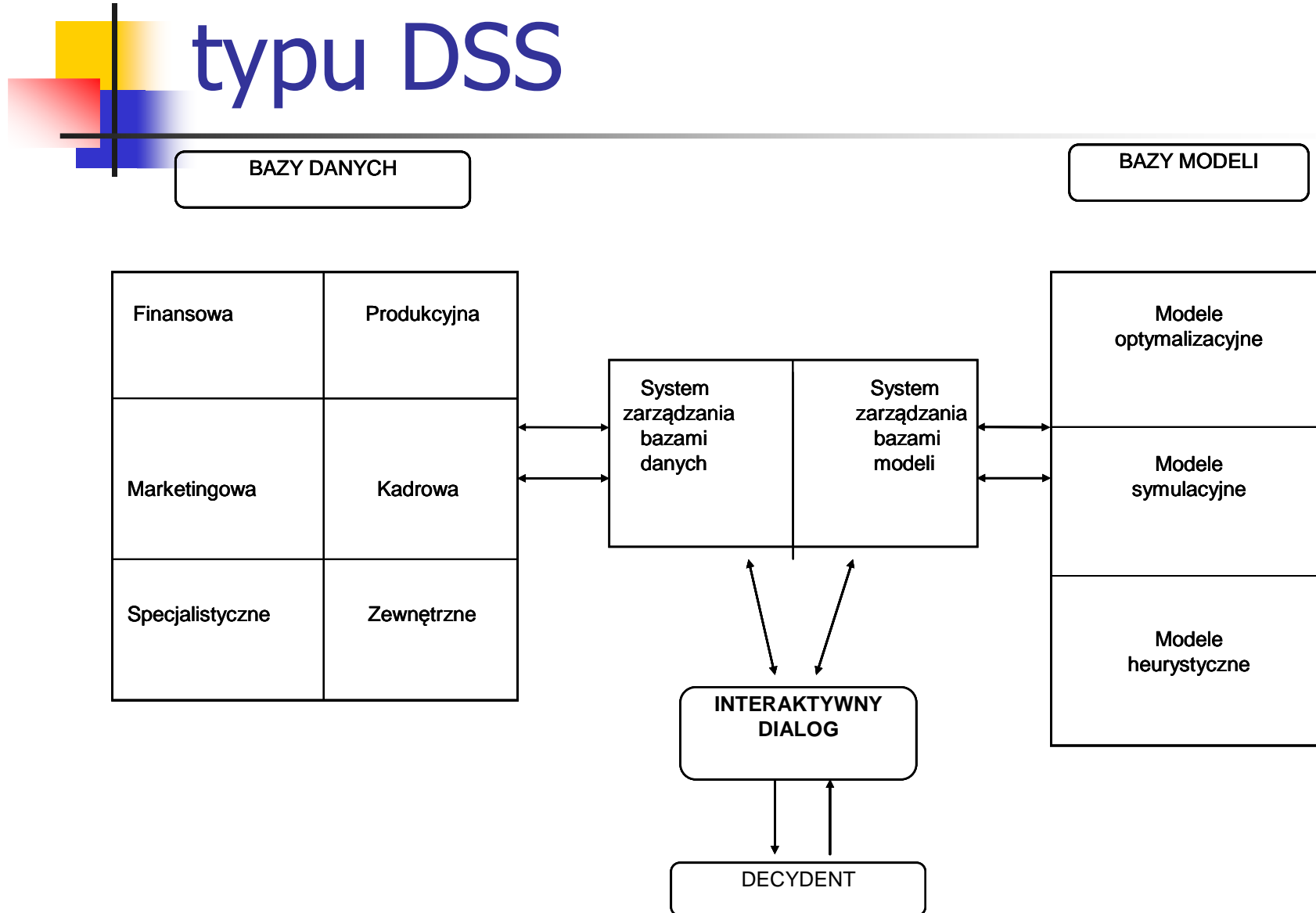
- Edytory tekstu
- Arkusze kalkulacyjne
- Uniwersalne bazy danych
- Aplikacje do prezentacji danych
- Aplikacje wspomagające organizację pracy
- Oprogramowanie do Internetu
- Systemy wspomagania decyzji
- Systemy informowania kierownictwa



Systemy wspomagania decyzji

- DSS to system informatyczny, który dostarcza informacje w danej dziedzinie przy wykorzystaniu analitycznych modeli decyzyjnych z dostępem do baz danych w celu wspomagania decydentów w skutecznym działaniu w kompleksowym i źle ustrukturalizowanym środowisku.

Ogólny schemat systemów typu DSS





Główne cechy systemu wspomagania decyzji

- Umożliwia szybkie analizy w tym także prowadzenie analizy wrażliwości.
- Zwiększa znacznie liczbę możliwych ocen i szacunków.
- Wymusza podniesienie wiedzy o zarządzaniu na szczeblach kierowniczych.
- Zintegrowany z istniejącymi w organizacji bazami danych.
- Zwiększa profesjonalność podejścia do procesów planowania i podejmowania decyzji w organizacji.



Systemy wspomagania decyzji grupowych

- GDS - Group Decision Support,
- GDSS - Group Decision Support Systems lub coraz częściej GSS - Group Support Systems.
- CSCW – Computer Supported Co-operative Work.
- EMS - Electronic Meeting Support.

Systemy informowania kierownictwa (1)



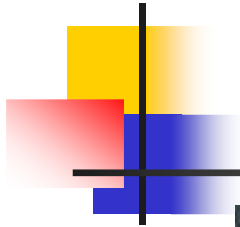
- EIS to wykorzystanie najnowszych rozwiązań z dziedziny informatyki dla stworzenia jak najbardziej komfortowych warunków dla podejmowania decyzji menedżerom najwyższego szczebla.
- EIS pomaga kierownictwu prowadzić tzw. monitoring skupiając jednak uwagę bardziej na ogólnym sprawnym działaniu firmy czy organizacji niż na optymalizacji konkretnych decyzji. Służą temu rozbudowane systemy zapytań oraz indywidualizacja przedstawianych raportów i narzędzi do komunikacji z systemem.



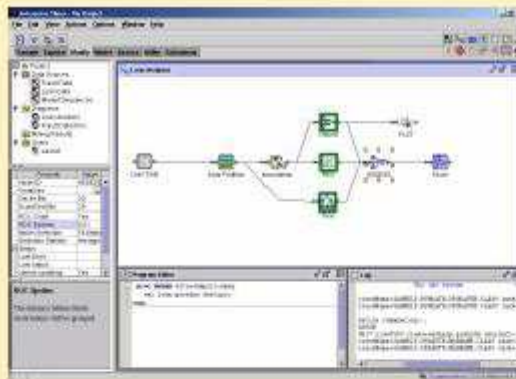
Systemy informowania kierownictwa (2)

- Narzędzie realizacji celów strategicznych:
 - zebranie informacji dotychczas rozproszonych (przez to niedostępnych) w jednym miejscu)
 - monitorowanie realizacji celów strategicznych
 - automatyzacja procesów informacyjnych

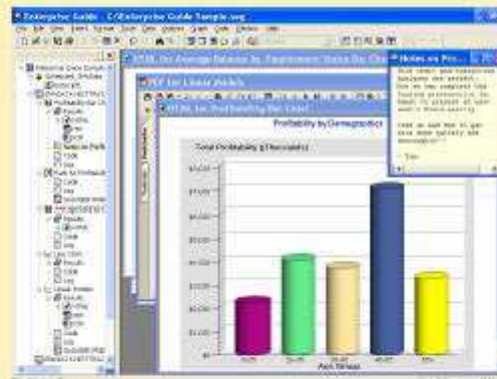
Filozofia systemu informowania kierownictwa



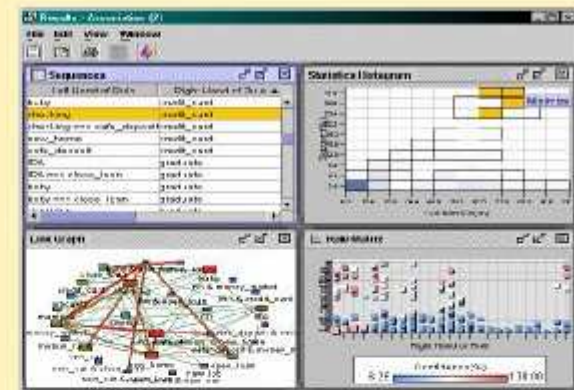
Rozbudowane „fundamenty”



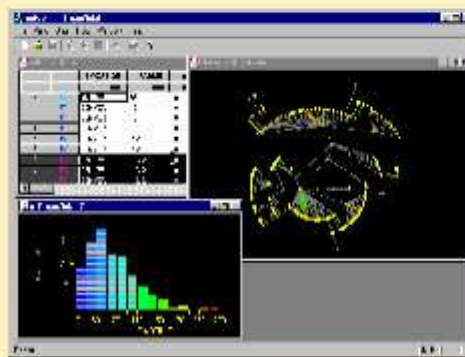
Modelowanie predykcyjne



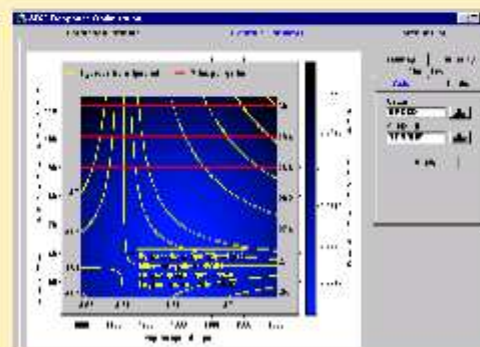
Business Intelligence



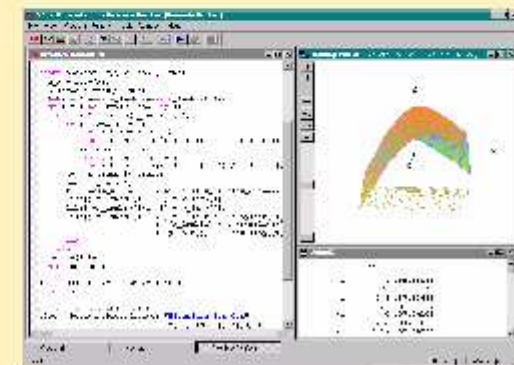
Data i Text Mining



Optymalizacja

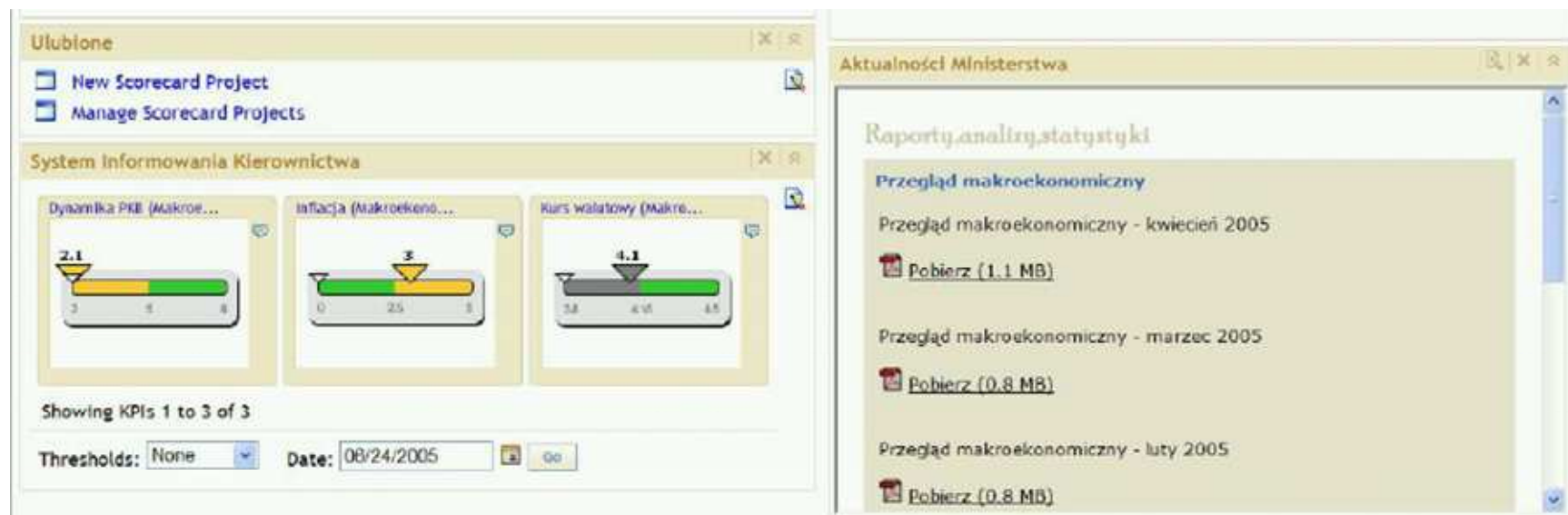


Zarządzanie jakością



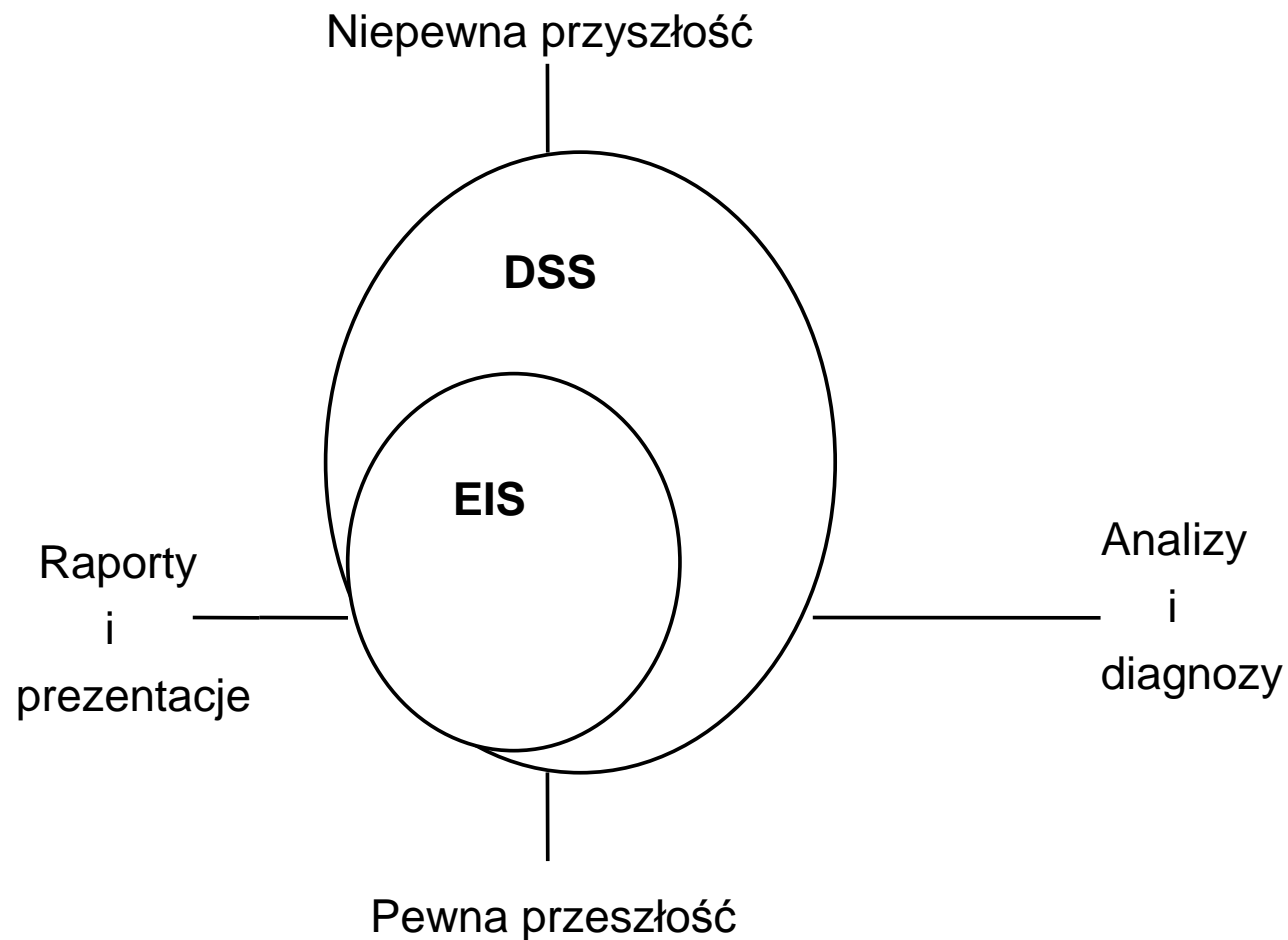
Modele matematyczne

Przykład kokpitu EIS





Relacje pomiędzy DSS a EIS





Systemy eksperckie

- Systemy eksperckie starają się naśladować decyzje eksperta-człowieka w konkretnej wybranej dziedzinie i potrafią to robić wielokrotnie w przyjazny dla użytkownika sposób.
- Na obecnym etapie rozwoju technologicznego system ekspertowy jest programem komputerowym, który wykonuje złożone zadania o dużych wymaganiach intelektualnych i robi to tak dobrze jak człowiek, będący ekspertem w tej dziedzinie.



Jak to działa (1)?

- **Fakty:**

- F1: Jan ma rybki
- F2: Jan ma grzałkę

- **Reguły:**

- R1: IF x ma rybki
THEN x ma akwarium
- R2: IF x ma grzałkę AND x ma akwarium
THEN x ma rybki żyjące

- **Wnioski:**

- W1: Jan ma akwarium.
- W2: Rybki Jana żyją.



Jak to działa (2)?

Proszę odpowiadać na pytanie tylko tak (T) lub nie (N)

Czy jest paliwo w zbiorniku ? T

Czy kranik dopływu paliwa jest zamknięty ? N

Odczekaj 5 min. Czy następna próba się powiodła ? N

Czy paliwo wypływa z przewodu doprowadzającego je do pompy ? T

MOIM ZDANIEM: Uszkodzenie pompy paliwa

EKSPERTYZA ZAKOŃCZONA

Czy chcesz następnej diagnozy ? T

Czy jest paliwo w zbiorniku ? T

Czy kranik dopływu paliwa jest zamknięty ? N

Odczekaj 5 min. Czy następna próba się powiodła ? N

Czy paliwo dopływa do gaźnika ? T

Czy ustawienie przesłony rozruchowej jest prawidłowe ? T

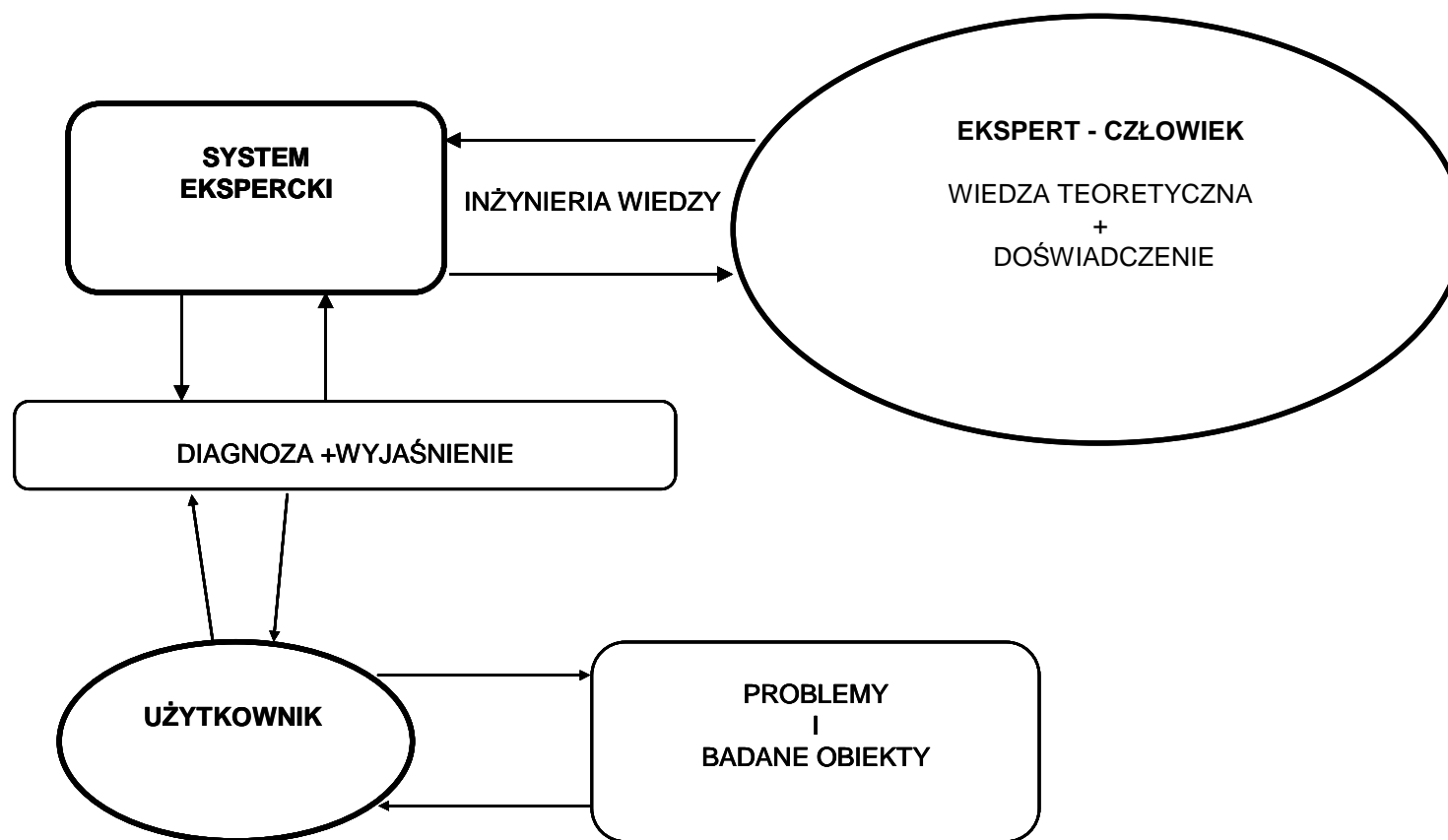
Czy główna dysza paliwa jest drożna ? T

MOIM ZDANIEM: Zanieczyszczona dysza biegu jałowego lub awaria w układzie zapłonowym

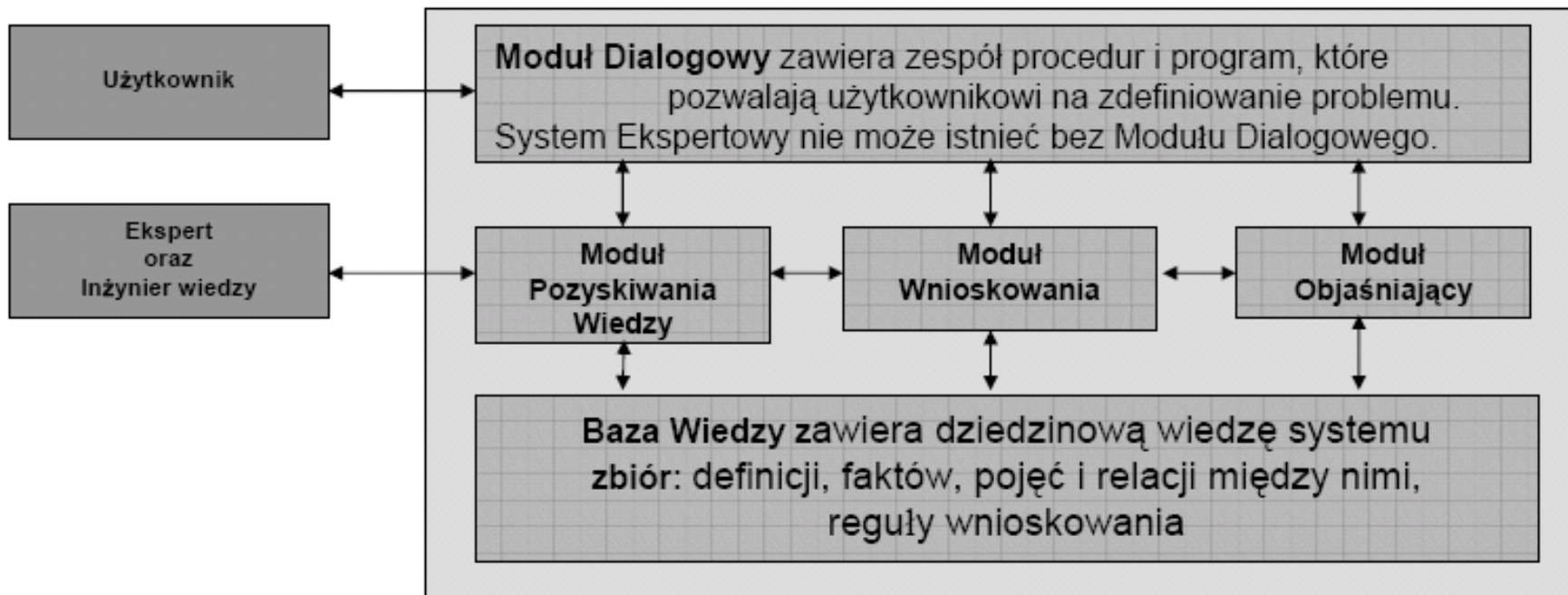
EKSPERTYZA ZAKOŃCZONA

Czy chcesz następnej diagnozy ? N

Schemat systemu eksperckiego



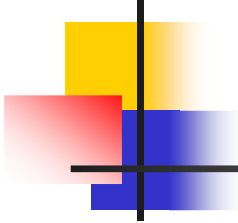
Konfiguracja systemu eksperckiego





System eksperckie vs. klasyczne

- Wykorzystanie w procesie przetwarzania mechanizmu wnioskującego
- Uzasadnienia podawanych odpowiedzi za pomoc modułu objaśniającego
- Dotyczą wąskiej dziedziny wiedzy
- Możliwość wnioskowania z niepełnej wiedzy
- Możliwość wyjaśniania łańcucha wnioskowania w sposób zrozumiały dla użytkownika
- Rozdział mechanizmu wnioskowania od bazy wiedzy
- Zastosowanie reguł wnioskowania postaci „if ... then ...”.



Przykłady systemów eksperckich

- MYCIN – system medyczny,
- EMYCIN – Empty Mycin: pusty szkielet dla innych dziedzin,
- GUIDON – składnik MYCIN służący do objaśniania,
- HEARSAY II – do rozpoznawania mowy,
- MACSYMA – system matematyczny,
- DENDRAL – system ekspertowy w dziedzinie chemii.
- PROSPECTOR – odkrywanie złóż rud żelaza – geologia.



Bazy danych

- Baza danych - zbiór zorganizowanych w logiczny sposób elementów danych, tworzący określoną strukturę.
- Komputerowa baza danych – uporządkowany zbiór danych przechowywany w pamięci komputera
- Element danych - podstawowy składnik informacji.



Ewolucja baz danych

- Bazy:
 - Przetwarzanie plików płaskich - lata 50
 - Sieciowe oraz hierarchiczne bazy danych - lata 60 i 70
 - Relacyjne bazy danych - lata 70 i 80
 - Obiektowo-relacyjne bazy danych - lata 90
 - Obiektowe bazy danych – lata 00?



Pliki płaskie

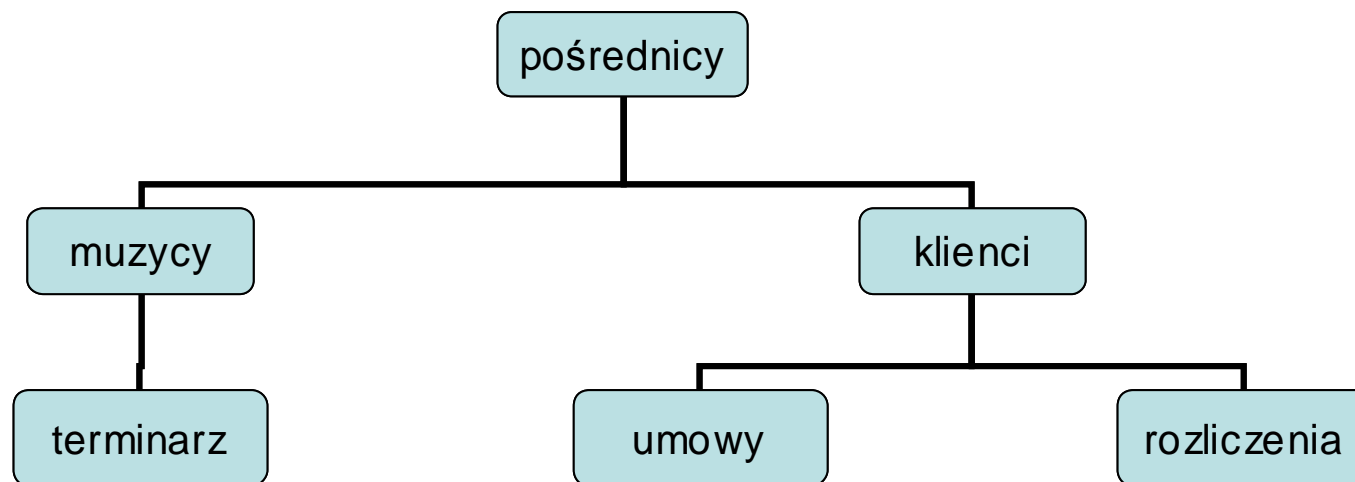
- Płaskie pliki (flat files) to sekwencyjne pliki, przechowujące dane w sposób liniowy.
- Wady:
 - Problemy związane z dzielonym przez kilka aplikacji dostępem do danych, co w efekcie prowadziło do duplikowania tych samych informacji i bardzo utrudniało ich aktualizację.
 - W przypadku zmiany struktury plików, wymagana była aktualizacja wszystkich programów działających w oparciu o nie.
 - Pliki płaskie nie posiadały mechanizmów zabezpieczających, takich jak kopie zapasowe, czy metody odzyskujące dane.
 - Brak jednolitego formatu przechowywania danych, co znacznie ograniczało możliwości przetwarzania przez różne systemy.

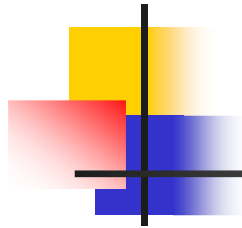


Hierarchiczny model danych

- Model ten wykorzystuje wskaźniki jako logiczne połączenia pomiędzy jednostkami danych.
- Zalety:
 - Łatwość zastosowania;
 - Bardzo duża szybkość działania.
- Wady:
 - Ścisłe reguły dotyczące relacji.
 - Wstawianie i kasowanie danych może okazać się bardzo skomplikowane.
 - Dostęp do niższych warstw jest możliwy tylko poprzez warstwy nadrzędne.
 - Trudności w modelowaniu relacji typu wiele - do - wielu.

Przykład modelu hierarchicznego



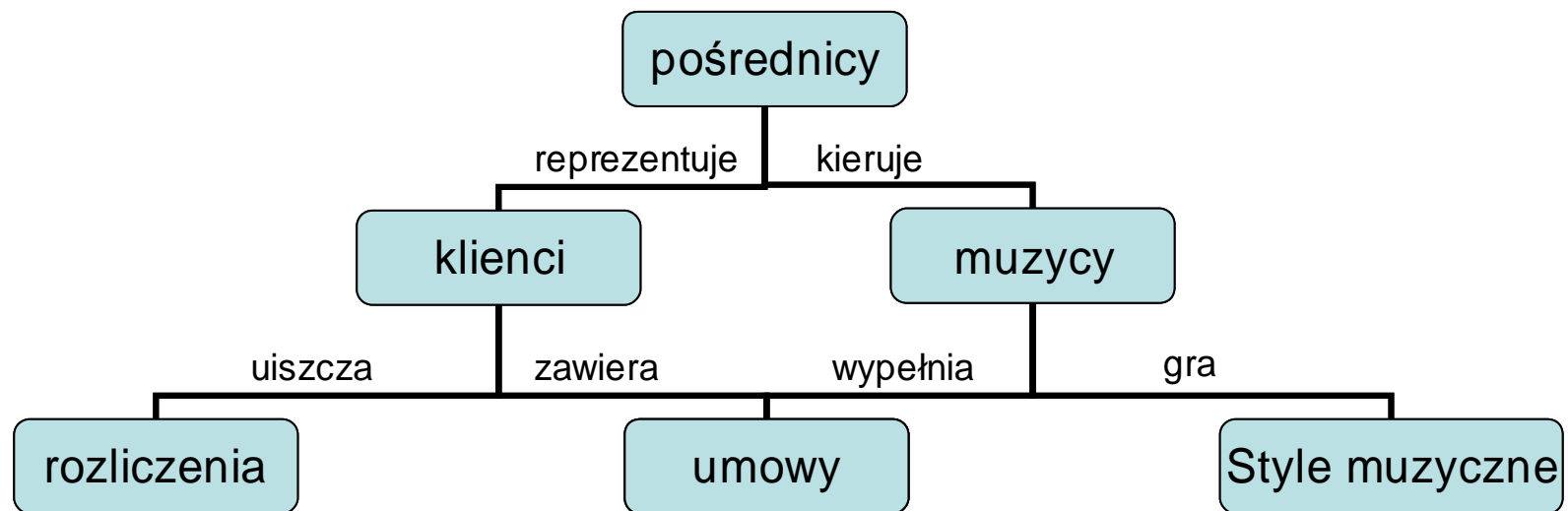


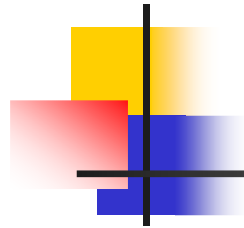
Sieciowy model danych

- W sieciowym modelu danych CODASYL dane posiadają wewnętrzne identyfikatory (lub adresy), zaś związki semantyczne pomiędzy danymi reprezentowane są poprzez powiązania referencyjne lub wskaźnikowe. Struktura danych tworzy, więc graf, czyli sieć.
- Wady:
 - Model ten jest bardzo złożony i skomplikowany w użyciu.
 - Sieciowe bazy danych, podobnie jak hierarchiczne, są bardzo trudne w nawigacji.
 - Implementacja strukturalnych zmian jest niezwykle trudna w przypadku baz danych tego typu.



Przykład modelu sieciowego





Relacyjny model danych

- W roku 1970 dr Edgar Ted Codd z firmy IBM zaprezentował relacyjny model danych. W modelu tym dane miały być przechowywane w prostych plikach liniowych, które to pliki nazywane są „relacjami” bądź „tabelami”.
- Codd wprowadził język SQL w celu ułatwienia dostępu do danych i ich modyfikacji. Tabele są dwuwymiarowymi tablicami składającymi się z wierszy i kolumn.

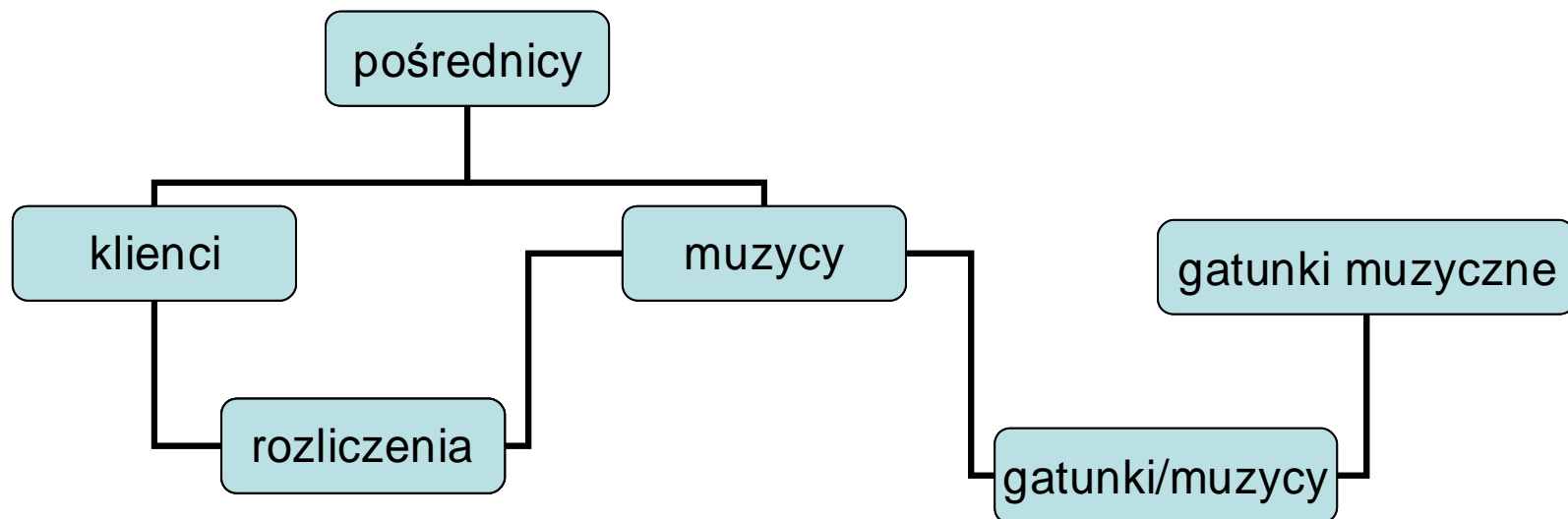


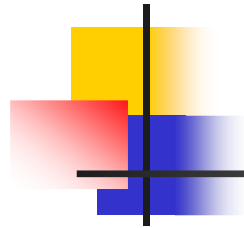
Relacyjny model danych

- Relacyjne bazy danych wprowadziły następujące ulepszenia w stosunku do hierarchicznych i sieciowych baz danych:
 - Prostota. Złożone diagramy, wykorzystywane w przypadku hierarchicznych i sieciowych baz danych, nie występują w przypadku relacyjnych baz danych.
 - Niezależność danych pozwala na modyfikowanie struktury danych bez wpływu na istniejące programy. Tabele nie są ze sobą połączone na sztywno. Można dodawać kolumny, tabele mogą być dołączane do bazy danych, a nowe relacje mogą być tworzone bez konieczności wprowadzania istotnych zmian do tabel.
 - Deklaratywny język dostępu do danych. Model relacyjny wprowadził język SQL w celu ułatwienia dostępu do danych i ich modyfikacji. Nawigacja w bazie danych jest ukryta przed użytkownikiem końcowym, w odróżnieniu od języka DML w systemie CODASYL, gdzie użytkownik musi znać wszelkie szczegóły określające ścieżkę dostępu do danych



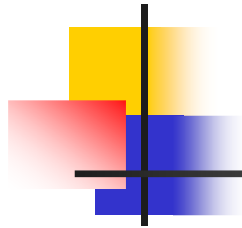
Relacyjny model danych





Postulaty Codda - wybór

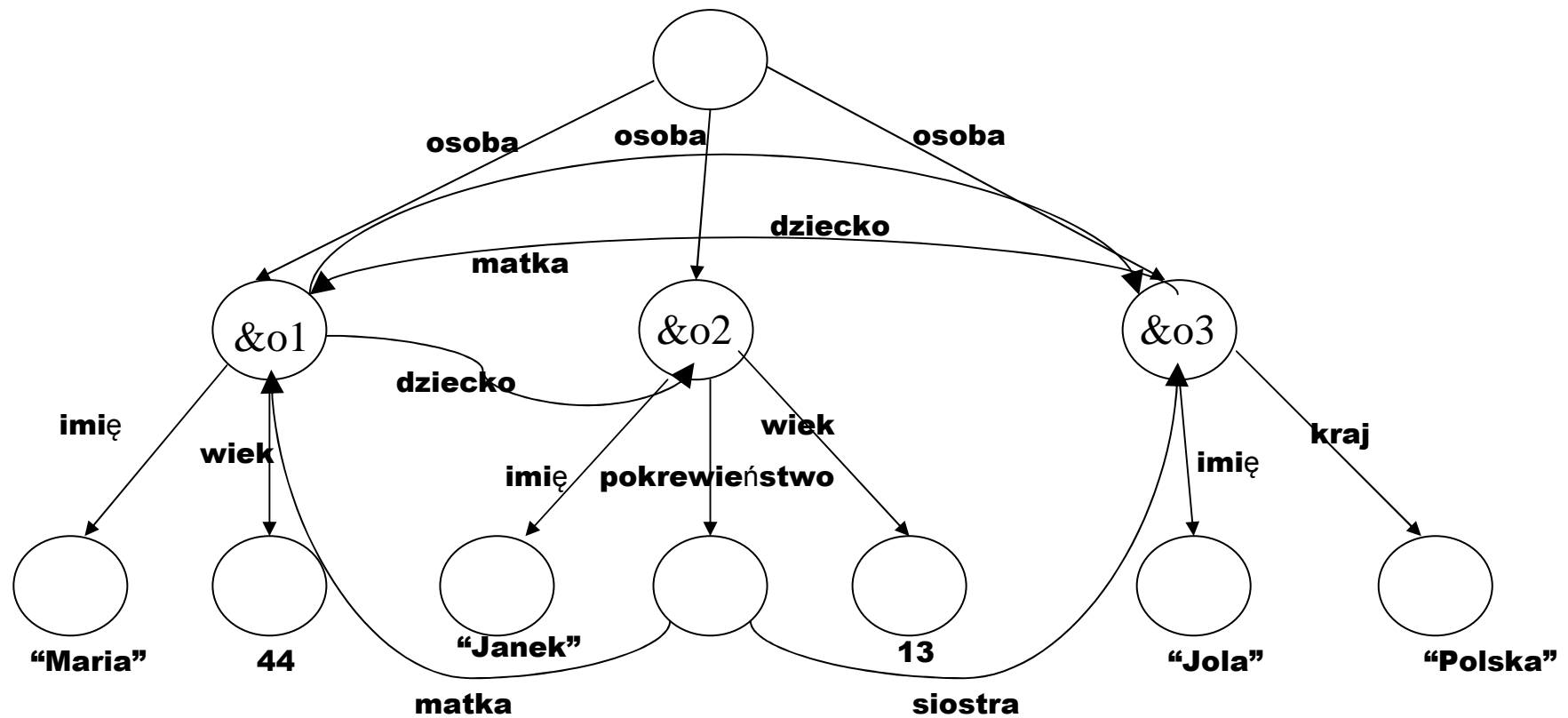
- System wdraża język zapytań
- Programy, za pomocą których manipuluje się bazą danych, są niezależne od tego, jak baza danych jest fizycznie zorganizowana
- Relacyjna baza danych powinna działać tak samo, niezależnie od tego, czy pracuje na pojedynczej maszynie, czy jest rozproszona przez sieć



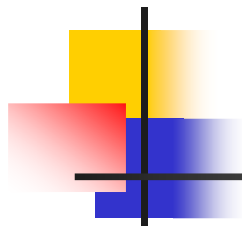
Obiektowy model danych

- Obiekty, w sensie zarządzania danymi, zostały wprowadzone po raz pierwszy na początku lat siedemdziesiątych przez Palo Alto Research Center z firmy Xerox.
- Podejście obiektowe podkreśla bardziej naturalną reprezentację danych. Ich zadaniem jest przetwarzanie dźwięku, obrazu, tekstu, grafiki itp. Potrzeby te wymagają dużo bardziej elastycznego formatu przechowywania danych niż hierarchiczne, sieciowe, czy relacyjne bazy danych mogą zapewnić.

Obiektowy model danych



Obiektowo - relacyjny model danych



- Obiektowo-relacyjne bazy danych są najnowszym osiągnięciem w rozwoju hybrydowych architektur baz danych. Systemy te pojawiły się między innymi z powodu ogromnych inwestycji poczynionych przez różne organizacje w systemy relacyjnych baz danych
- Nazywane są Object Relational lub Extended Relational. Są wynikiem ostrożnej ewolucji systemów relacyjnych w kierunku obiektowych.



System zarządzania bazą danych

- System zarządzania bazą danych - jest to zbiór narzędzi umożliwiający dostęp i zarządzanie jedną lub więcej bazami danych.
- Umożliwia zarządzanie elementami danych tak, aby można było wyszukiwać i wykorzystywać dane zawarte w bazie przy minimalnym nakładzie pracy.



Przykłady systemów zarządzania bazą danych

- Oracle 9i/10g (www.oracle.com)
- IBM DB 2 (www.ibm.com)
- Microsoft Access (www.microsoft.com)
- Microsoft SQL Server .Net
(www.microsoft.com)
- Tamino (www.softwareag.com/)
- Postgres (www.postgresql.org)
- MySQL (www.mysql.com)



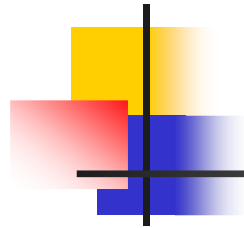
Problemy organizacji z przetwarzaniem danych

- Ilość danych
 - Różnorodność formatów zapisu danych
 - Wiele rozproszonych źródeł danych
- Charakter wydobywanej informacji
- Szybkość transakcji
- Dostępność danych



Problemy organizacji z przetwarzaniem danych

- Systemy do zarządzania przedsiębiorstwami intensywnie korzystają z transakcyjnych baz danych. Jedni użytkownicy wprowadzają do nich informacje np. o wysyłce towaru lub sprzedaży. Drudzy pobierają dane, generując z nich faktury, rozliczenia, raporty.
- Zasoby informacji zgromadzone w bazie danych systemu ERP są niezbędne do bieżącej działalności firmy jednak po pewnym czasie stają się "historią", stanowiącą dla systemu informatycznego niepotrzebny balast.



Czym jest hurtownia danych

- Hurtownia danych jest to:
 - scentralizowana nietransakcyjna baza danych przeznaczona do przechowywania informacji w długim horyzoncie czasowym globalnie w skali instytucji, w wielowymiarowych układach analitycznych i ukierunkowana na wyszukiwanie informacji bezpośrednio przez końcowych użytkowników.
- **Angielskie pojęcia: data warehouse, data mart (uwaga: to nie jest to samo!!!)**



Czym jest tematyczna hurtownia danych

- Data Mart - ściśle związana z mniejszą grupą użytkowników, tematycznie dostosowana, zoptymalizowana dla ich potrzeb hurtownia danych.
- "Data mart" można opisać jako podzbiór hurtowni danych przeznaczony do zaspokajania potrzeb pojedynczej jednostki biznesowej (np. oddziału firmy), czy też mniejszych jednostek gospodarczych, w zakresie dostępu do informacji za pomocą specjalizowanej aplikacji.



Porównanie hurtowni danych z data marts

Cechy:	Data marts	Hurtownie danych
Obszar wdrożenia	oddział/przedsiębiorstw o	Korporacja
Zakres tematyczny	Jeden	Wiele
Źródła danych	Kilka	Wiele
Rozmiary	<100 GB	100GB – 1TB
Platforma	NT, Unix	Unix



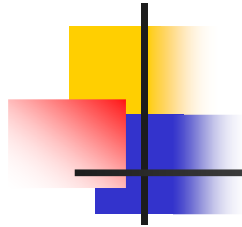
Trochę szczegółów

- Hurtownia danych jest, zgodnie z definicją, skonsolidowanym repozytorium danych historycznych.
 - Repozytorium zawiera dane historyczne, nie podlegające już zmianom (bo nie zmieniamy danych o sprzedaży sprzed np. roku).
 - Dane są skonsolidowane, co rzutuje na architekturę bazy danych – jest ona znacznie mniej złożona niż w przypadku baz OLTP (On Line Transactional Processing – baz danych dla przetwarzania transakcyjnego).
 - Dane są zorganizowane w strukturę wielowymiarową, w której fakty (liczby przechowywane w bazie) są zależne od wielu parametrów (nazywanych wymiarami).



Dlaczego hurtownia danych?

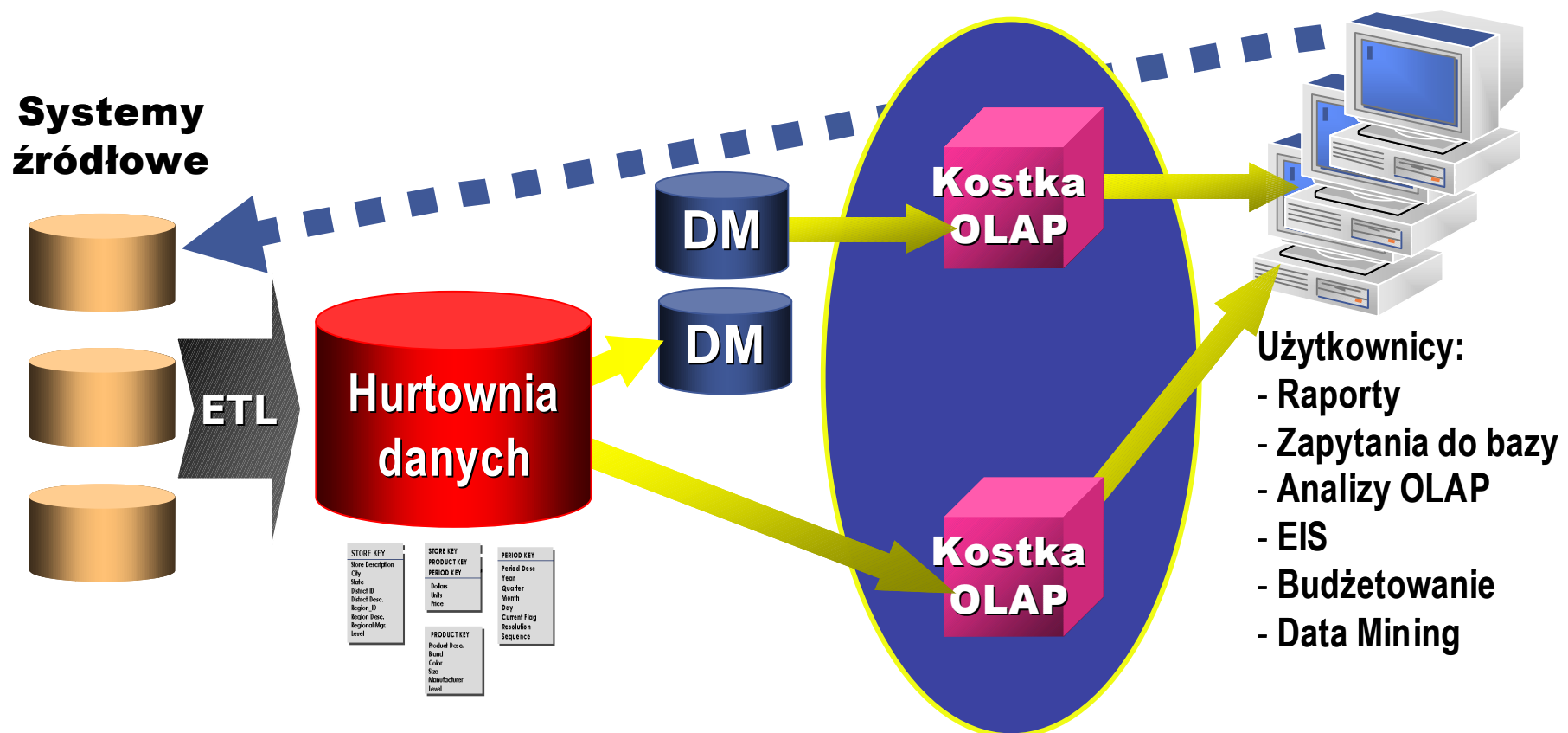
- Zdenormalizowana struktura bazy hurtowni danych powstała w wyniku procesu poszukiwania modelu logicznego bazy, który umożliwiałby wydajne, złożone i wszechstronne raportowanie.
- Próby usprawnienia raportowania i prowadzenia analiz w oparciu o tradycyjne bazy OLTP poniosły fiasko.



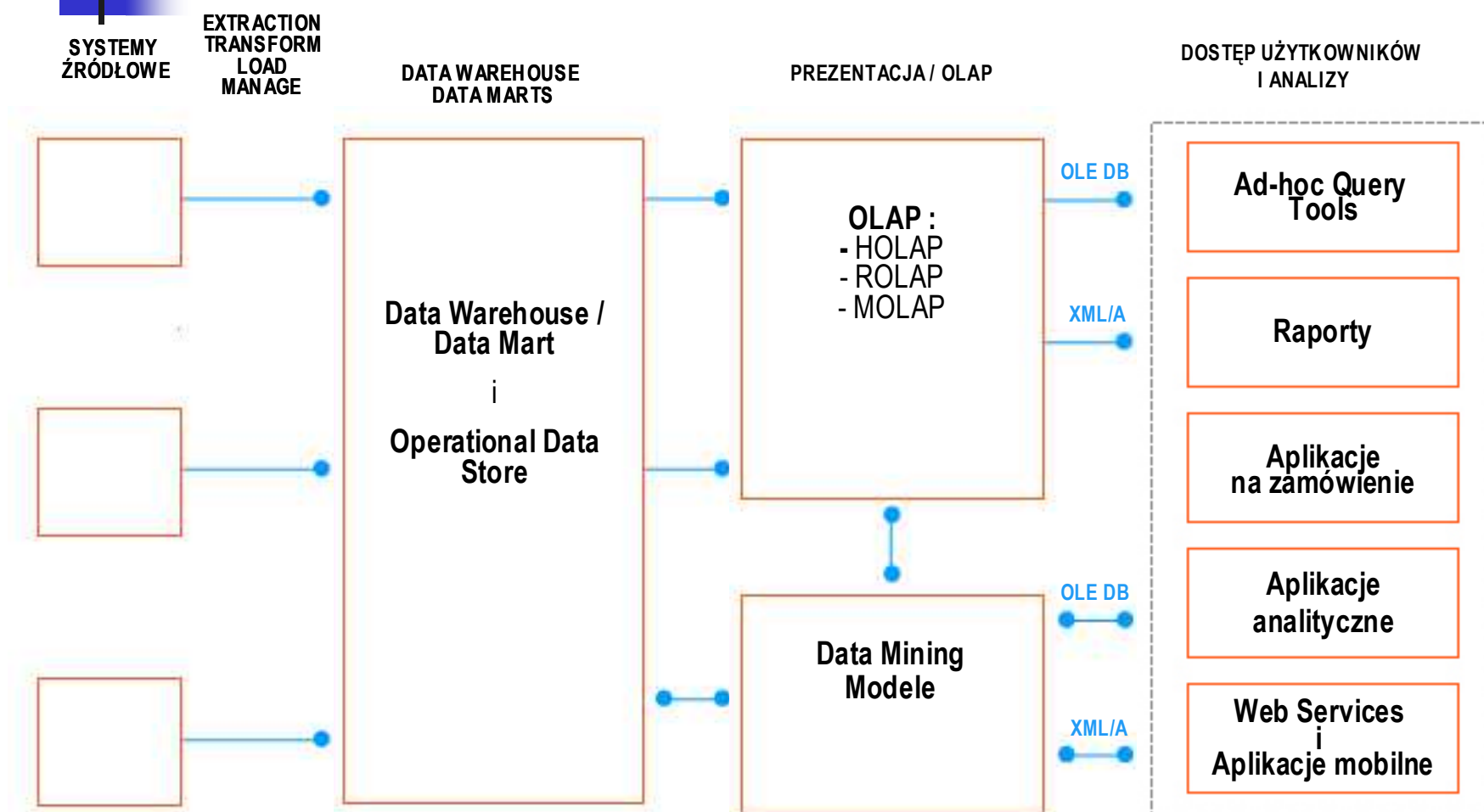
Cechy hurtowni danych

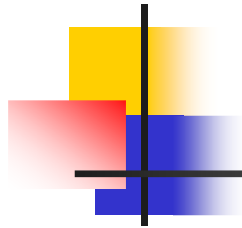
- Tematyczność
 - Grupowanie danych
- Integralność
 - Normalizacja pól, wspólne typy danych
- Oznaczenie czasowe
 - Analiza w długich okresach
- Niezmiennność

Jak to działa?



Jak to działa?





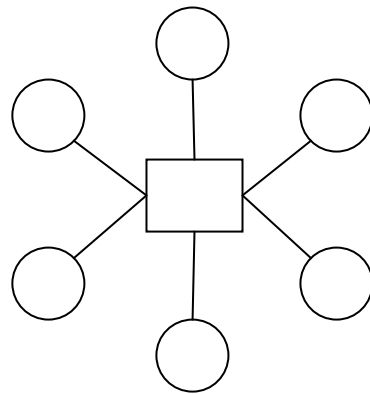
Zasilanie hurtowni danych

- Ważnym elementem tych systemów są procedury ekstrakcji, czyszczenia, transformacji i ładowania danych do bazy (ang. Extract, Transformation, Load – ETL).
 - Procedury ekstrakcji danych z systemów OLTP uruchamiane są w czasie minimalnego obciążenia tych systemów.
 - Dane wyekstrahowane są następnie weryfikowane względem reguł i danych słownikowych przechowywanych w repozytorium metadanych, przekształcane do pożądanej w hurtowni postaci i następnie ładowane do bazy.
 - Dzięki procedurom ETL dane w hurtowni charakteryzują się wysoką jakością, przewyższającą znacznie jakość danych systemów OLTP.
 - Przykład: deduplikacja – procedura usuwająca powtórzenia danych: w systemach OLTP panowie Jan Kowalski i Jan Piotr Kowalski mogą być różnymi klientami, mimo iż mieszkają pod tym samym adresem, w hurtowni (dzięki procesowi deduplikacji) zostaną zidentyfikowani jako ta sama osoba umożliwiając tym samym rzeczywistą analizę zachowań klienta.

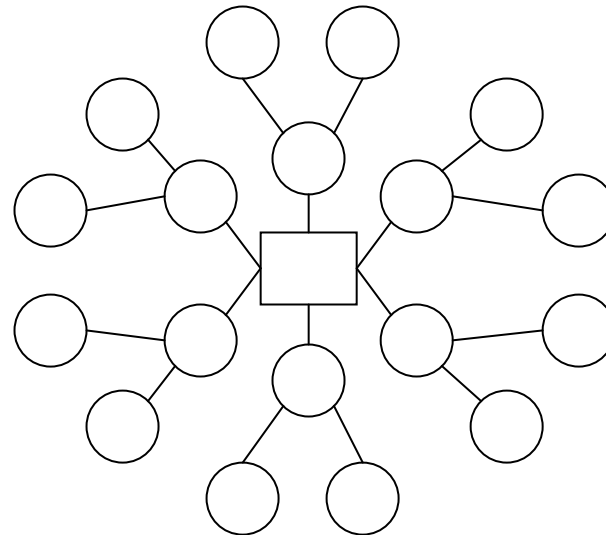
Struktura danych w hurtowni danych

- Dwie podstawowe struktury

Model płatka śniegu



Model gwiazdy



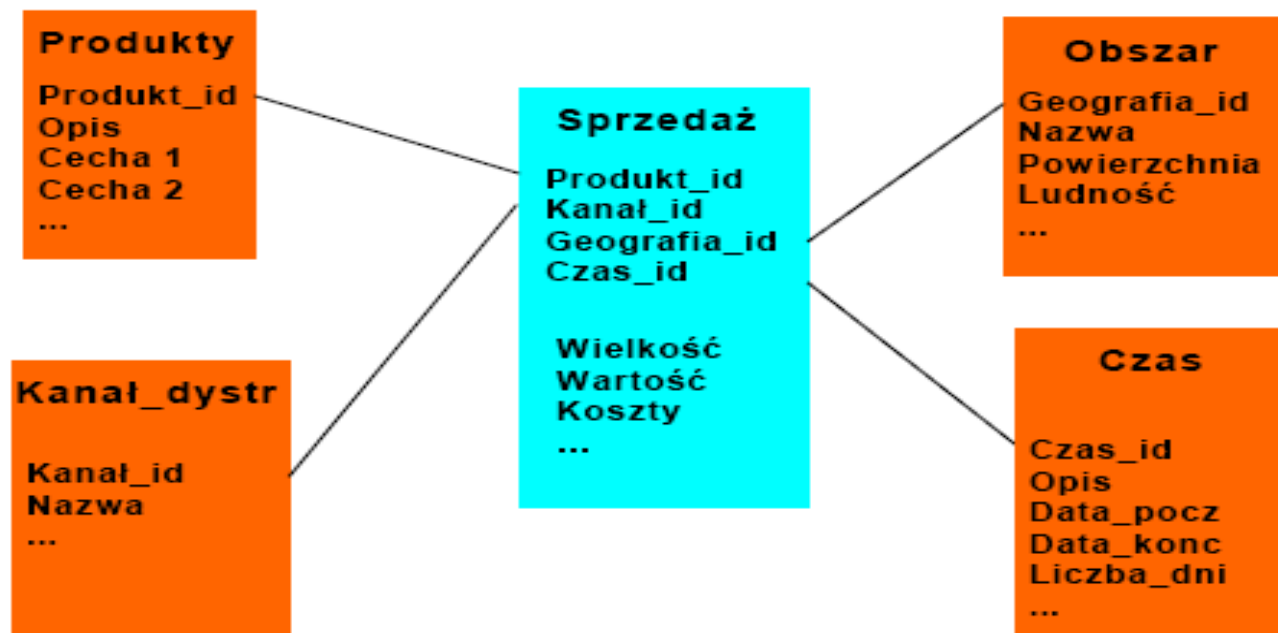
- tablica faktów

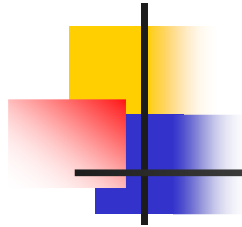


- hierarchia wymiaru



Model gwiazdy



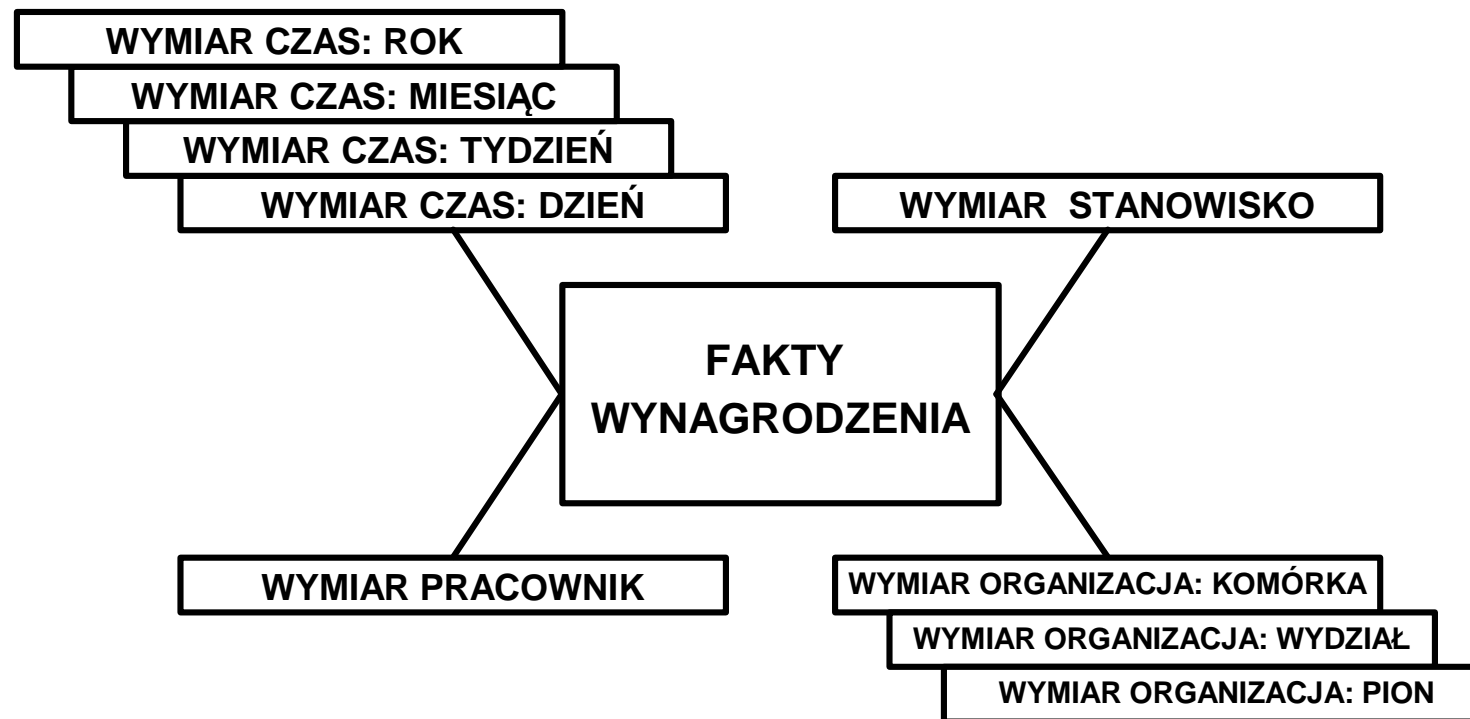


Co oznacza model gwiazdy

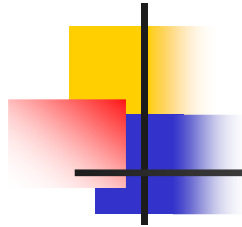
- Dane dotyczące sprzedaży (wielkość, wartość, ...) gromadzone są w centralnej tabeli faktów
- Parametry (wymiary) od których sprzedaż zależy znajdują się w mniejszych tabelach takich jak np. produkt, geografia czy czas.
- Elementy wymiarów (dla geografii np. województwa) mogą posiadać cechy charakterystyczne (np. rozmiar, liczba ludności), które będą wykorzystywane w analizach.
- Schemat gwiazdy odpowiada strukturze najbardziej zdenormalizowanej



Model płatka śniegu



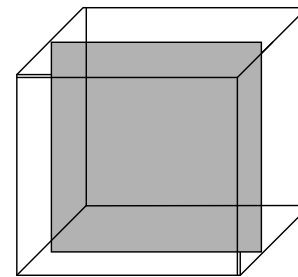
Zastosowania hurtowni danych



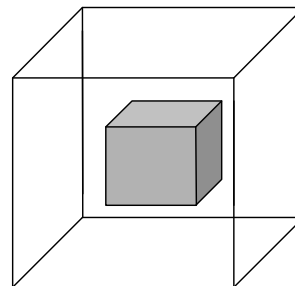
- Obsługa danych księgowych – spełnia funkcję archiwalną przechowując dane, które dają obraz sytuacji na konkretny dzień.
- Obsługa sprawozdawczości - wykorzystuje informacje z baz klientów.
- Obsługa finansowa – przeznaczona do zarządzania aktywami i pasywami, umożliwia obliczenie ogólnego przepływu finansowego.
- Obsługa marketingowo-klientowska – umożliwia podział klientów na segmenty, określenie koszyków produktowych, sprawdzenie solidności klientów i skutków działań marketingowych.

Narzędzia analityczne

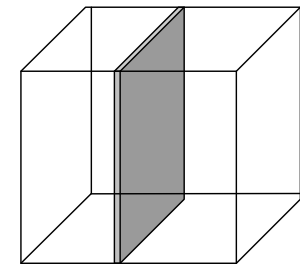
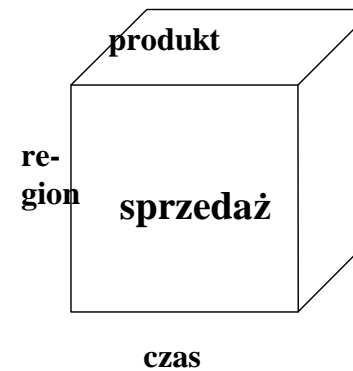
- Aplikacje OLAP
- Data mining



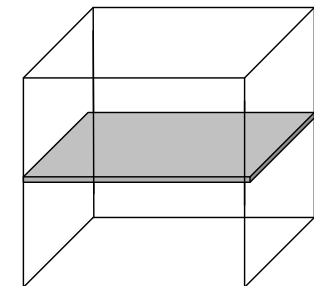
Analiza produktu



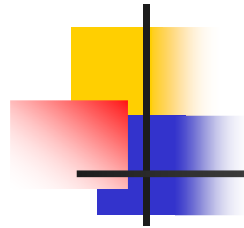
Analiza Ad-hoc



Analiza w czasie



Analiza regionu



Narzędzia analityczne c.d.

- Aby uniknąć konieczności budowania dla kolejnych raportów coraz to nowych programów stworzono narzędzia analityczne umożliwiające tworzenie w bardzo prosty sposób, praktycznie tylko za pomocą myszy, nawet bardzo złożonych raportów.
- Stworzono klasę narzędzi pozwalających prowadzić nawet skomplikowane analizy w trybie on-line (On Line Analytical Processing - OLAP).
 - Systemy klasy OLAP przełamały wszystkie ograniczenia systemów raportujących z baz OLTP umożliwiając prowadzenie różnego rodzaju analiz biznesowych także na bardzo dużych bazach danych.

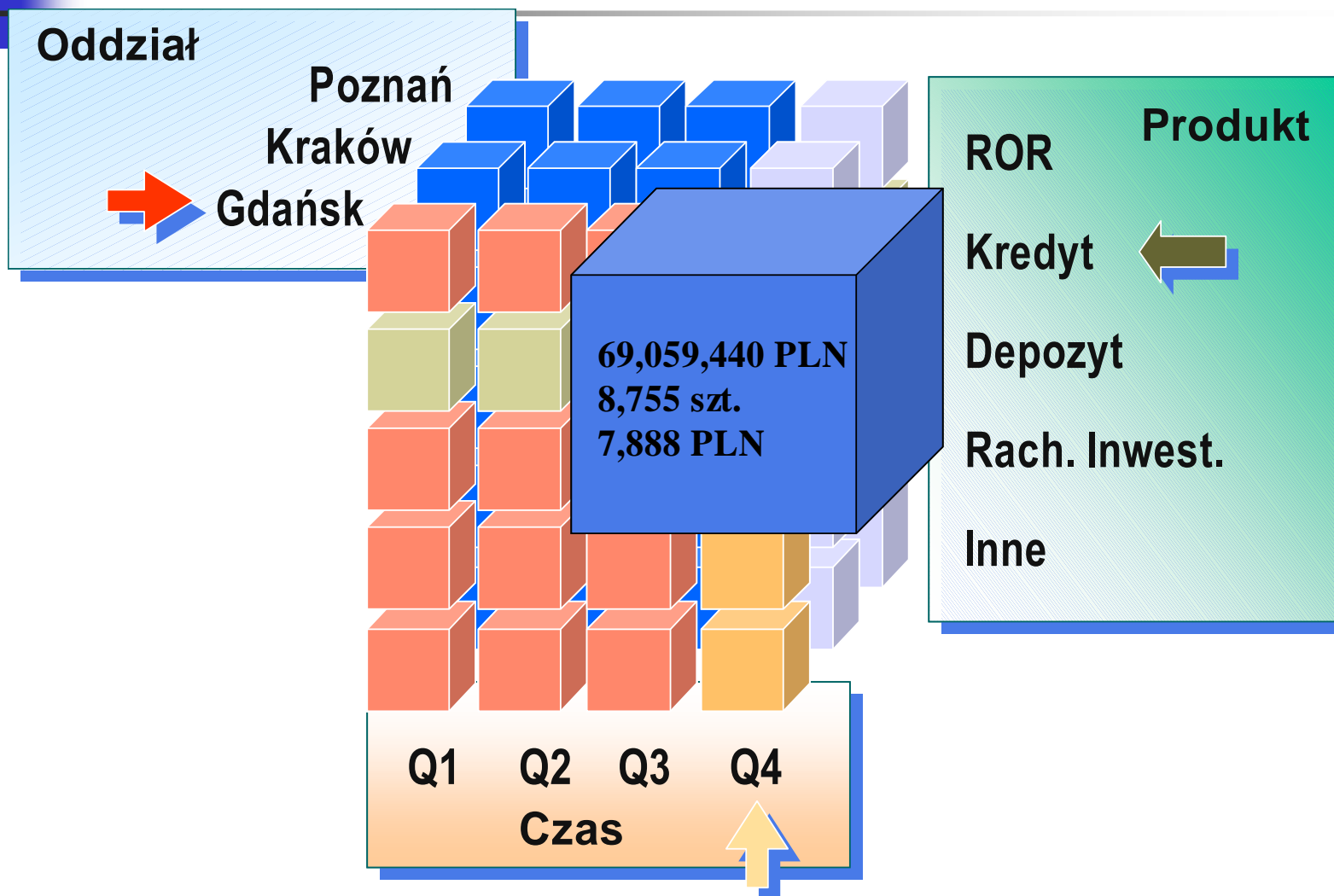


Porównanie narzędzi wspomagających analizy

Zapytania i raporty	OLAP	Data Mining
Wybór pojedynczych i zsumowanych danych	Podsumowania, trendy i prognozy	Odkrywanie ukrytych zależności
Informacja	Analiza	Wgląd i przewidywanie
Kto zakupił produkt w ostatnich dwóch latach?	Jaki jest średni przychód od klienta, który kupił produkt w ciągu ostatnich dwóch lat z podziałem na regiony?	Kto i dlaczego kupi produkt w najbliższym półroczu?

Tradycyjne bazy danych, narzędzia OLAP oraz Data Mining

Przykład kostki analitycznej





Data mining

- Narzędzia drążenia i eksploracji danych (ang. Data Mining) służą do automatycznego znajdowania związków między danymi.
- Narzędzia klasy DataMining wykorzystują wiele wyrafinowanych technik takich jak na przykład sieci neuronowe, drzewa decyzyjne, sieci Bayesa, algorytmy genetyczne, clustering czy regresja.
- Narzędzia tej klasy są wykorzystywane przez analityków między innymi do segmentacji bazy klientów, prognozowania, pozycjonowania produktu na rynku, a także do wykrywania oszustw w czasie rzeczywistym.



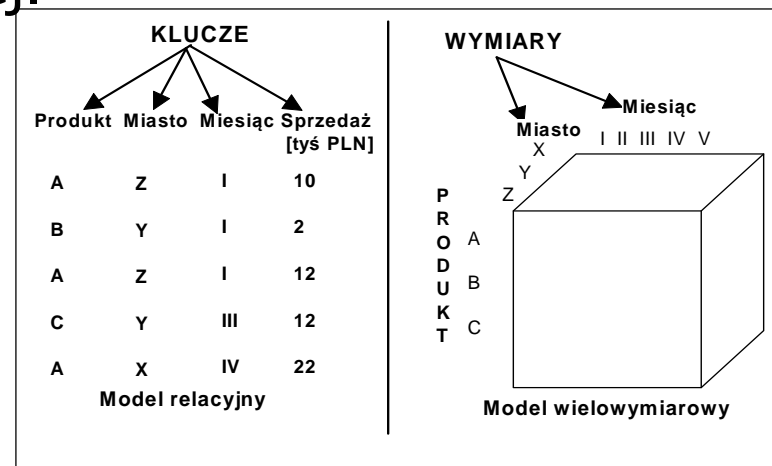
Data mining

- Data Mining, choć oferuje automatyczne generowanie wyników, wymaga dobrego ich zrozumienia (w celu uniknięcia pułapek) i dlatego prowadzony jest zwykle przez zaawansowanych analityków często posiadających doktorat z matematyki czy nauk przyrodniczych.
- Dla ułatwienia i usystematyzowania analiz drążenia danych opracowano w 1996 roku metodykę CRISP DM (Cross-Industry Standard Process for Data Mining), w której określono sposób prowadzenia analizy od zrozumienia zadań biznesowych i dostępnych danych poprzez przygotowanie danych i modelowanie aż po oszacowanie poprawności modelu i jego wdrożenie do eksploatacji. Metodyka ta jest dziś wspierana przez praktycznie wszystkich wytwórców oprogramowania klasy DataMining.

Relacyjne a wielowymiarowe bazy danych

w hurtowniach danych

- Od samego początku hurtownie danych realizowane były w oparciu o bazy danych dwojakiego rodzaju: relacyjne i wielowymiarowe.
- Omawiana struktura gwiazdy jest przykładem modelu danych bazy relacyjnej. W takiej strukturze fakty są funkcją wielu zmiennych (wymiarów), i jest to jedna z bardzo wielu możliwych struktur danych bazy relacyjnej.



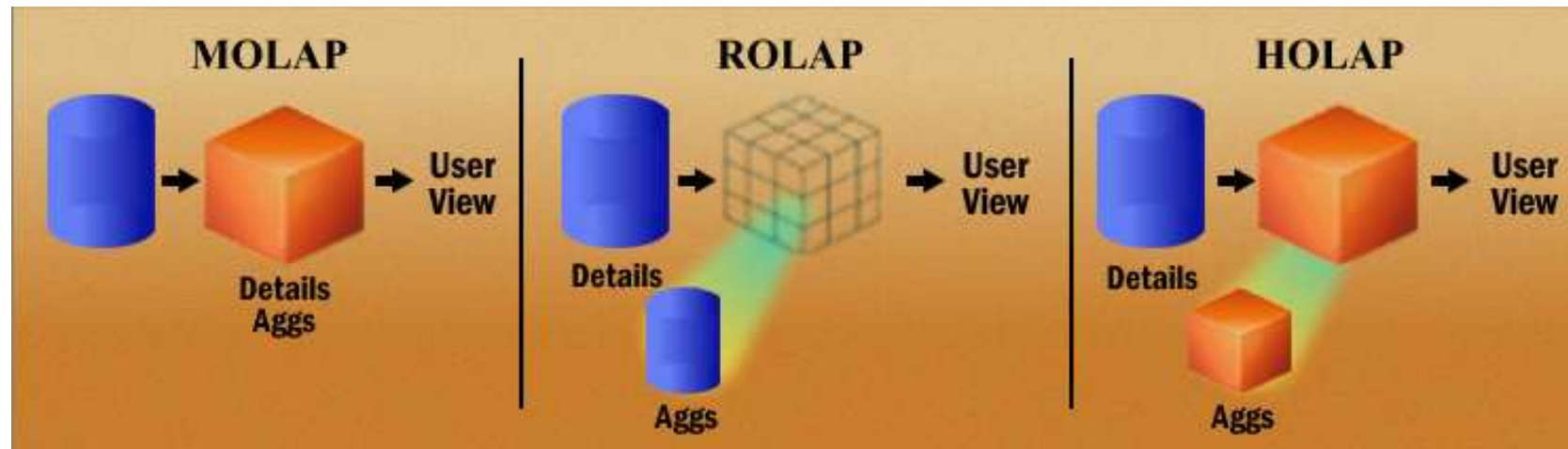


Model relacyjny vs. wielowymiarowy

- W bazie wielowymiarowej dane są przechowywane z definicji w postaci faktów – funkcji kilku zmiennych, i jest to podstawowa forma gromadzenia danych w tej bazie.
- Relacyjne bazy danych mogą gromadzić olbrzymie ilości danych osiągając rozmiar wielu terabajtów.
- Bazy wielowymiarowe natomiast mają ograniczenie na wielkość i przeważnie nie przekraczają rozmiarem kilkudziesięciu gigabajtów. Oferują natomiast strukturę danych przygotowaną do prowadzenia wszechstronnych i złożonych raportów na podstawie różnorodnych przecięć przez dane wielowymiarowe.
- Przedstawicielami baz wielowymiarowych są np. Oracle Express czy Essbase (Hyperion).

MOLAP, ROLAP, HOLAP

Do baz relacyjnych
stworzono narzędzia typu
ROLAP
(Relational OLAP)



Do baz
wielowymiarowych
stworzono narzędzia typu
MOLAP
(Multidimensional OLAP)



Jak działa baza hybrydowa

- Dane gromadzone są w bazie relacyjnej a następnie odwzorowywane (przez aplikację mapującą) na strukturę wielowymiarowej bazy danych (zaznaczonej symbolicznie kostką).
- Baza wielowymiarowa nie zawiera faktów a tylko elementy wymiarów, jej rozmiar jest zatem bardzo mały, nawet jeśli baza relacyjna jest wielkości terabajtów.
- Do analiz stosowane jest narzędzie analityczne klasy MOLAP. W czasie analizy zapytanie z narzędzia MOLAP jest przesyłane do wielowymiarowej bazy danych, gdzie generowane jest dynamiczne zapytanie SQL, które jest przesyłane przez warstwę odwzorowującą do bazy relacyjnej.
- Wynik zapytania powraca do narzędzia analitycznego tą samą drogą (oczywiście w odwrotnej do zapytania kolejności).



Badanie „Co daje naprawdę hurtownia danych”

- Temat: **Analiza przydatności hurtowni danych w procesie podejmowania decyzji**
- Grupa docelowa:
 - 37 firm w oparciu o następujące kryteria:
 - Rok od wdrożenia HD
 - Rozproszone struktury organizacyjne
 - Osoby z kadry kierowniczej



Co daje naprawdę hurtownia danych?

- Referenci uznali za opłacalne przedsięwzięcia wdrożenia hurtowni danych tylko wtedy, gdy integrowała ona dane z wielu firm, zatem organizacji o charakterze organizacji wirtualnej.
- Po wdrożeniu hurtowni danych pracownicy starają się wykorzystać jej możliwości do podejmowania wszystkich decyzji niezależnie czy są one strategiczne dla firmy czy nie (jest to dość zaskakujący wynik, gdyż z hurtownie danych wdrażane są z reguły w celu zidentyfikowania trendów występujących w otoczeniu)
- Firmy, które nie radziły sobie z nadmiarem danych uznały wdrożenie hurtowni danych za opłacalne – hurtownie danych znacznie ułatwiają odnalezienie koniecznych informacji z TB danych
- Wyznaczanie trendów wynikających ze zbiorów danych powinno być przydatne w podejmowaniu decyzji strategicznych, ankietowani, jednak nie potwierdzają tej hipotezy.
- Najistotniejszym wnioskiem z ankiet jest fakt, iż osoby pracujące z hurtownią danych uważają, że przedsięwzięcie jest opłacalne i pomaga w radzeniu sobie z nadmiarem ilości danych.



Gdzie hurtownie danych mają zastosowanie?

- W telekomunikacji hurtownie danych wykorzystujące dane bilingowe umożliwiają między innymi segmentację klientów na grupy w różny sposób korzystających z usług operatora.
 - Pozwala to na ustanowienie taryf dedykowanych specjalnie dla tych grup.
 - Ponadto, poprzez obliczanie tzw. wskaźnika 'churn' można określić, którzy klienci noszą się z zamiarem zrezygnowania z usług firmy telekomunikacyjnej.
- Dzięki analizie danych dziesiątek i setek tysięcy transakcji dla tysięcy produktów oferowanych w supermarketach (jest to tzw. analiza koszyka) można dobrze określić preferencje klientów i korelacje pomiędzy produktami. Umożliwia to handlowcom wyjście naprzeciw oczekiwaniom klientów i zwiększenie sprzedaży.



Producenci hurtowni danych

- SAS
- Oracle
- Sybase
- IBM
- Microsoft
- SAP



Dziękuję za uwagę
