

BAZY DANYCH

LITERATURA

- ★ **Beynon-Davies P. - Systemy baz danych – WNT Warszawa 2003**
- ★ **Ullman J.D., Widom J. – Podstawowy wykład z systemów baz danych – WNT Warszawa 2001**
- ★ **Date C.J. - Wprowadzenie do systemów baz danych - WNT Warszawa, 2000**
- ★ **L. Banachowski, A. Chęczyńska, K. Matejewski, E. Mrówka-Matejewska, K. Stencel - Bazy danych. Wykłady i ćwiczenia – Wydawnictwo PJWSTK Warszawa 2003**
- ★ **Hernandez M.J. – Bazy danych dla zwykłych śmiertelników – Mikom 2004**
- ★ **Kowalski P – Podstawowe zagadnienia baz danych i procesów przetwarzania – Mikom 20005**



Wykład 1

Wprowadzenie do Baz Danych



Opis podstawowych cech systemu bazy danych, definicje podstawowych pojęć, znaczenie baz danych we współczesnych systemach informacyjnych

Jacek Bartman

jbartman@univ.rzeszow.pl

1. Bazy danych jako maszyny abstrakcyjne

1.1. System bazy danych

★ Większość nowoczesnych przedsiębiorstw musi rejestrować dane dotyczące swojej codziennej działalności.

■ Np.: Uniwersytet musi m.in. rejestrować:

- jakich ma studentów i wykładowców,
- jakie wykłady i moduły są przez nich prowadzone,
- którzy wykładowcy uczą których modułów,
- którzy studenci są zapisani na jakie moduły,
- którzy studenci zaliczają które moduły.

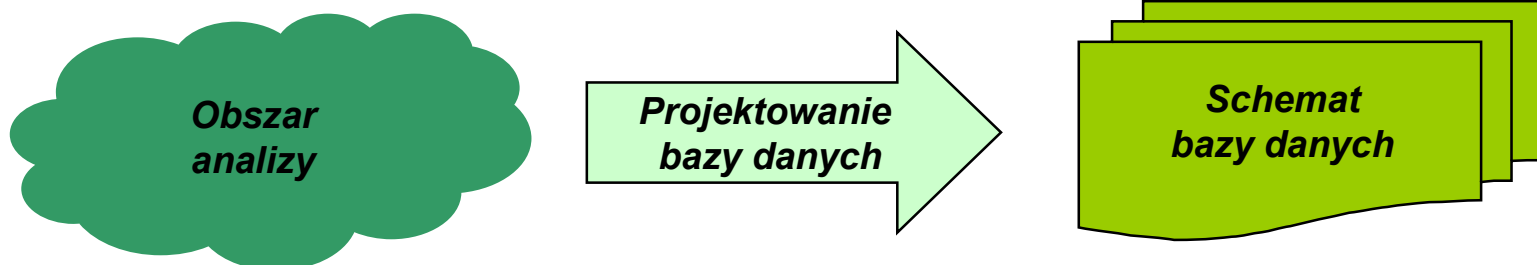
■ Dane mogą być wprowadzane przez różne osoby

■ Dane można wykorzystać do różnych celów

- np.: lista aktualnych studentów, wykorzystanie sal itp.

1.2 Obszar analizy

- ★ baza jest modelem pewnego aspektu rzeczywistości, rzeczywistość ta to obszar analizy
- ★ w przypadku uniwersytetu obszar analizy może obejmować m.in. moduły oferowane studentom i studentów zaliczających te moduły (rzeczy istotne dla uniwersytetu)
- ★ rzeczy istotne nazywa się **klasami** lub **encjami**,
 - np.: studenci, moduły,
- ★ klasy mogą być powiązane,
 - np.: którzy studenci jakie moduły zaliczają,
- ★ klasy mają **właściwości** lub **atrybuty**,
 - np.: studenci mają nazwiska, imiona, adresy itp..
- ★ klasy, atrybuty i związki muszą mieć jakąś reprezentację w bazie danych (baza musi być zaprojektowana)

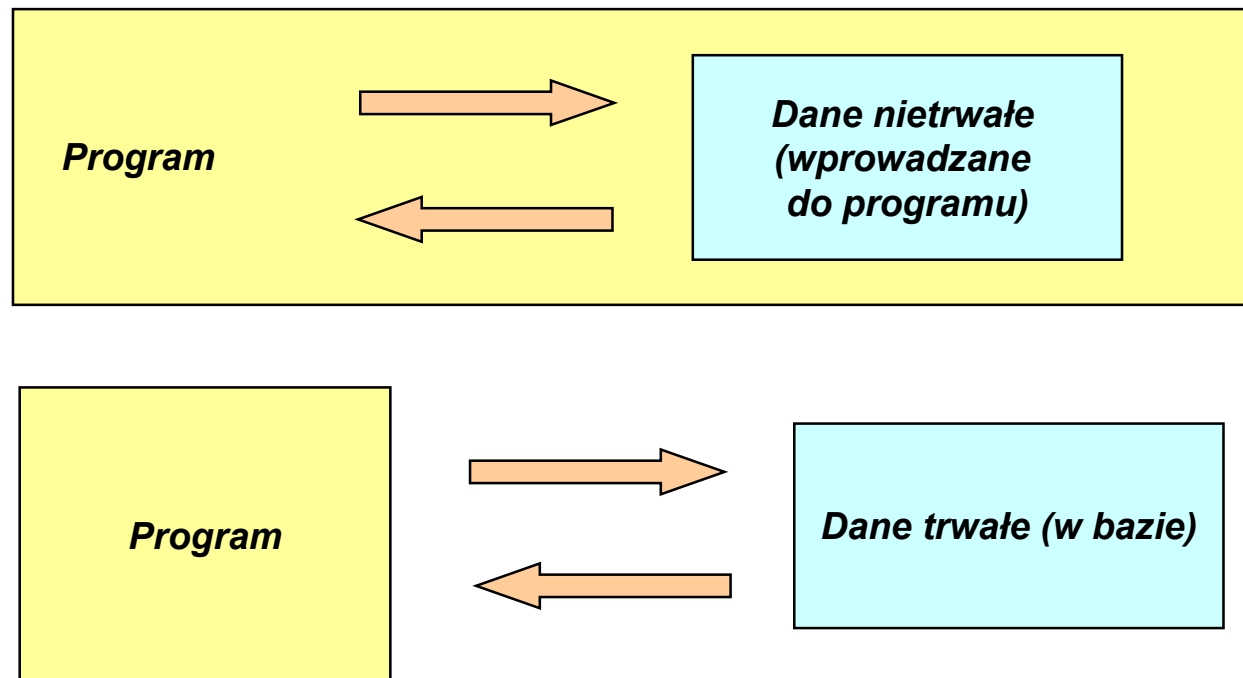


1.2.1. Bazy faktów

- ✱ **Dane** to **fakty**, dana jest symbolem lub zbiorem symboli
- ✱ Fakty same w sobie nie mają znaczenia, aby były użyteczne muszą być zinterpretowane.
- ✱ Zinterpretowane dane to **informacje** (mają przypisaną semantykę).
 - np. : napis 43 sam nic nie znaczy, musi być podany kontekst
- ✱ Bazę danych możemy uważać za zbiór faktów lub pozytywnych asercji na temat obszaru analizy za zbiór faktów.
- ✱ W określonej chwili baza danych znajduje się w jakimś stanie (zawiera fakty prawdziwe w danej chwili).

1.2.2 Trwałość

- ★ Dane w bazie danych traktowane są jako trwałe (mogą być przechowywane przez pewien czas).
- ★ Dane wprowadzane przy terminalu przeznaczone do przetworzenia nie są trwałe.



1.2.3 Część intensjonalna i ekstensjonalna

★ Baza danych składa się z dwóch części: intensjonalnej i ekstensjonalnej

■ **Część intensjonalna** – zbiór definicji które opisują strukturę danych bazy, inaczej jest to schemat bazy danych (uzyskany w wyniku projektowania)

■ **Część ekstensjonalna** – łączny zbiór danych w bazie (zawartość bazy)

Schemat: <i>Uniwersytet</i>	Ekstensja (zawartość)
Moduły: Systemy relacyjnych baz danych Projektowanie relacyjnych baz danych	
Studenci: Jan Kowalski urodzony 1 lipca 1980 Piotr Nowak urodzony 10 maja 1980 Maria Maj urodzona 18 stycznia 1981	
Zaliczają: Jan Kowalski zalicza Systemy relacyjnych baz danych	

Schemat: <i>Uniwersytet</i>	Intensja (schemat)
Klasy: Moduły – kursy prowadzone przez instytucję w semestrze Studenci – osoby zaliczające moduły w tej instytucji	
Związki: Studenci zaliczają moduły	
Atrybuty: Moduły mają nazwy Studenci mają nazwiska	

1.3. Integralność

- ★ Baza ma właściwości integralności gdy dokładnie odzwierciedla swój obszar analizy
 - Np.: w uniwersyteckiej bazie danych integralność oznacza zapewnienie, że baza daje poprawne odpowiedzi na pytania typu „*ilu studentów jest zapisanych na moduł Systemy relacyjnych baz danych*”
- ★ Integralność jest ważna szczególnie w bazach, których dane zmieniają się.
- ★ W zbiorze przyszłych stanów bazy są poprawne i niepoprawne integralność gwarantuje, iż baza nie przyjmie stanu niepoprawnego.

Schemat: Uniwersytet	
Moduły: Systemy relacyjnych baz danych Projektowanie relacyjnych baz danych	Ok. Stan poprawny
Studenci: Jan Kowalski urodzony 1 lipca 1980 Piotr Nowak urodzony 10 maja 1980 Maria Maj urodzona 18 stycznia 1981	
Zaliczają: Jan Kowalski zalicza Systemy relacyjnych baz danych Jan Kowalski zalicza Projektowanie relacyjnych baz danych	

Schemat: Uniwersytet	
Moduły: Systemy relacyjnych baz danych Projektowanie relacyjnych baz danych	Stan niepoprawny
Studenci: Jan Kowalski urodzony 1 lipca 1980 Piotr Nowak urodzony 10 maja 1980 Maria Maj urodzona 18 stycznia 1981	
Zaliczają: Jan Kowalski zalicza Systemy relacyjnych baz danych Piotr Nowak zalicza Podstawy programowania	

1.3.1. Redundancja (nadmiarowość)

- ★ **Redundancja to powtarzanie się danych w bazie** (należy jej unikać)
- ★ Fakty w bazie danych nie powinny się powtarzać

Schemat: Uniwersytet

Moduły:

Systemy relacyjnych baz danych
Projektowanie relacyjnych baz danych

Studenci:

Jan Kowalski urodzony 1 lipca 1980
Piotr Nowak urodzony 10 maja 1980
Maria Maj urodzona 18 stycznia 1981

Zaliczają:

Jan Kowalski zalicza Systemy relacyjnych baz danych
Piotr Nowak zalicza Projektowanie relacyjnych baz danych
Jan Kowalski zalicza Systemy relacyjnych baz danych

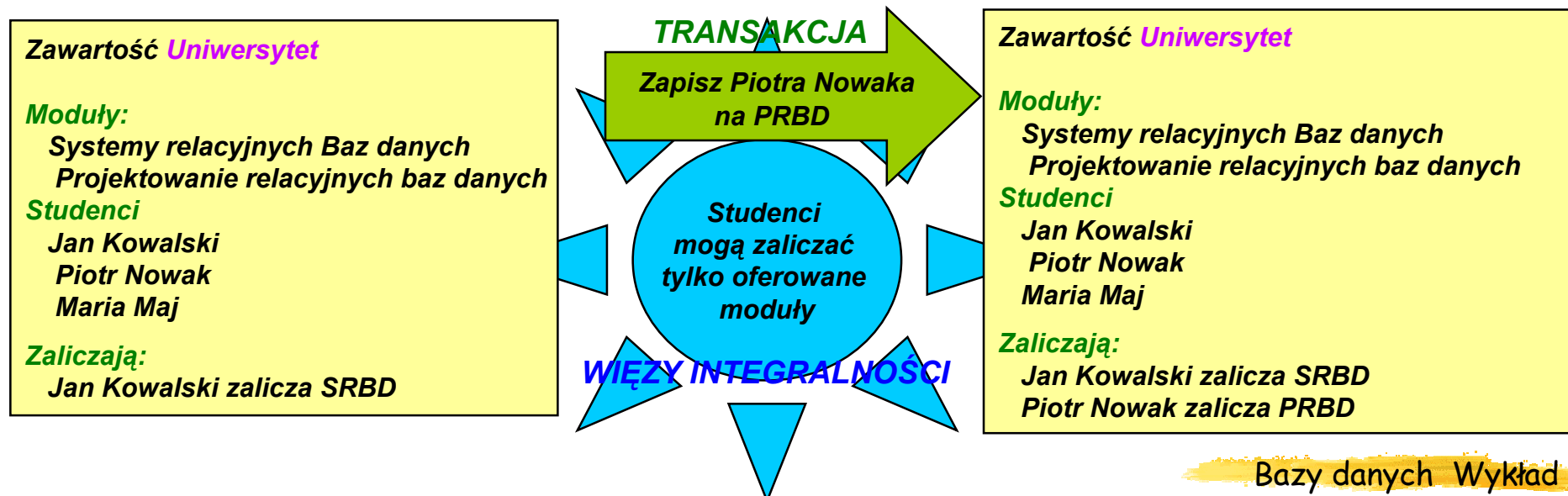
1.3.2 Transakcje

- ★ **Transakcje to zdarzenia powodujące zmianę stanu bazy danych**
- ★ Typem transakcji nazywamy transakcje o podobnej formie np.:

zapisz studenta na moduł

1.3.3. Więzy integralności

- ★ Integralność bazy danych jest realizowana przez więzy integralności
- ★ Więzy integralności to reguły określające w jaki sposób baza danych ma pozostać dokładnym odbiciem swego obszaru analizy
 - **Wizy statyczne (niezmienniki stanu)** – sprawdzają czy wykonana transakcja nie zmienia stanu bazy w stan niepoprawny.
 - np.: czy studenci zaliczają tylko oferowane moduły
 - **Więzy przejść** – reguły, które wiążą ze sobą stany bazy danych. Stanowią ograniczenie nałożone na przejście:
 - np.: liczba modułów zaliczanych przez studenta nie może w semestrze zmniejszyć się do zera



1.4. Funkcje bazy danych

- ★ Do wykonywania operacji na bazie są potrzebne dwa rodzaje funkcji:
 - funkcje aktualizujące,
 - funkcje zapytań.

1.4.1. Funkcje aktualizujące

- ★ **Funkcje aktualizujące dokonują zmian w danych**
- ★ Transakcje mogą inicjować funkcje aktualizujące, które zmieniają bazę z jednego stanu w inny
 - Np. na uniwersytecie:
 - Zainicjuj semestr
 - Zaoferuj moduł
 - Anuluj moduł
 - Wpisz studenta na kurs
 - Przenieś studenta między modułami

1.4.2. Funkcje aktualizujące i więzy integralności

- ★ Z funkcją aktualizującą związany jest ciąg warunków reprezentujących więzy integralności

■ np.:

Funkcja aktualizująca: *Przenieś studenta X z modułu 1 do modułu 2*

Warunki:

Student X zalicza moduł 1

Student X nie zalicza modułu 2

Moduł 2 jest oferowany

Akcje:

Zaprzecz, że student X zalicza moduł 1

Stwierdź, że student X zalicza moduł 2

- ★ Akcja danej funkcji aktualizującej może spowodować, że warunki innej funkcji aktualizującej staną się prawdziwe

■ np.:

Funkcja aktualizująca: *Anuluj moduł 1*

Warunki:

Moduł 1 istnieje

Liczba zapisanych studentów na moduł 1 jest mniejsza od 10

Akcje:

Zaprzecz, że moduł 1 istnieje

- Jeżeli funkcja „Anuluj moduł 1” jest aktualnie niewykonalna gdyż nie jest spełniony warunek 2. To wygenerowanie funkcji przeniesienia studenta z modułu 1 do innego, spowoduje zmniejszenie liczby studentów zapisanych na moduł 1 i może doprowadzić do sytuacji w której funkcja „Anuluj moduł 1” będzie wykonalna

1.4.3. Funkcje zapytań

★ **Funkcje zapytań wydobywają dane z bazy danych, nie modyfikując ich zawartości**

■ np.:

- Czy moduł X jest oferowany
- Czy student X zalicza moduł Y
- Którzy studenci zaliczają moduł X ?
- Które moduły są obecnie oferowane?

Zwraca odpowiedź tak/nie

Zwraca zbiór wartości

Ten typ pytań jest częściej stosowany

1.5. Formalizmy

- ★ Każdy system bazy danych musi używać jakiegoś formalizmu reprezentacji (konwencji zapisu).
- ★ **Formalizm reprezentacji zbioru składnikowych i semantycznych konwencji, które umożliwiają opisywanie rzeczywistości.**
 - Składnia reprezentacji określa zbiór reguł łączenia symboli i układów symboli w celu tworzenia wyrażeń
 - Semantyka reprezentacji określa w jaki sposób należy interpretować wyrażenia
- ★ **W bazach danych idea formalizmu reprezentacji odpowiada pojęciu modelu danych**
 - Modele relacyjne dobrze reprezentuje fakty, ale komplikują reprezentację funkcji aktualizujących
 - Modele obiektowe oferują dobrą reprezentację funkcji aktualizujących ale nie posiadają jednolitości dostępnych implementacji

1.6. Dostęp dla wielu użytkowników

- ★ Większość baz danych jest używana przez wielu użytkowników – powstaje problem **współbieżności**

- Np.: jeden użytkownik rejestruje studenta na jakiś moduł a inny w tym samym czasie usuwa ten moduł,
- Jeden użytkownik sprzedaje miejsce w samolocie na określony rejs, a inny robi to samo

■ **Działania takie prowadziły by do niespójności bazy danych**

- ★ Baza danych musi radzić sobie z konfliktami współbieżności

1.6.1. Perspektywy

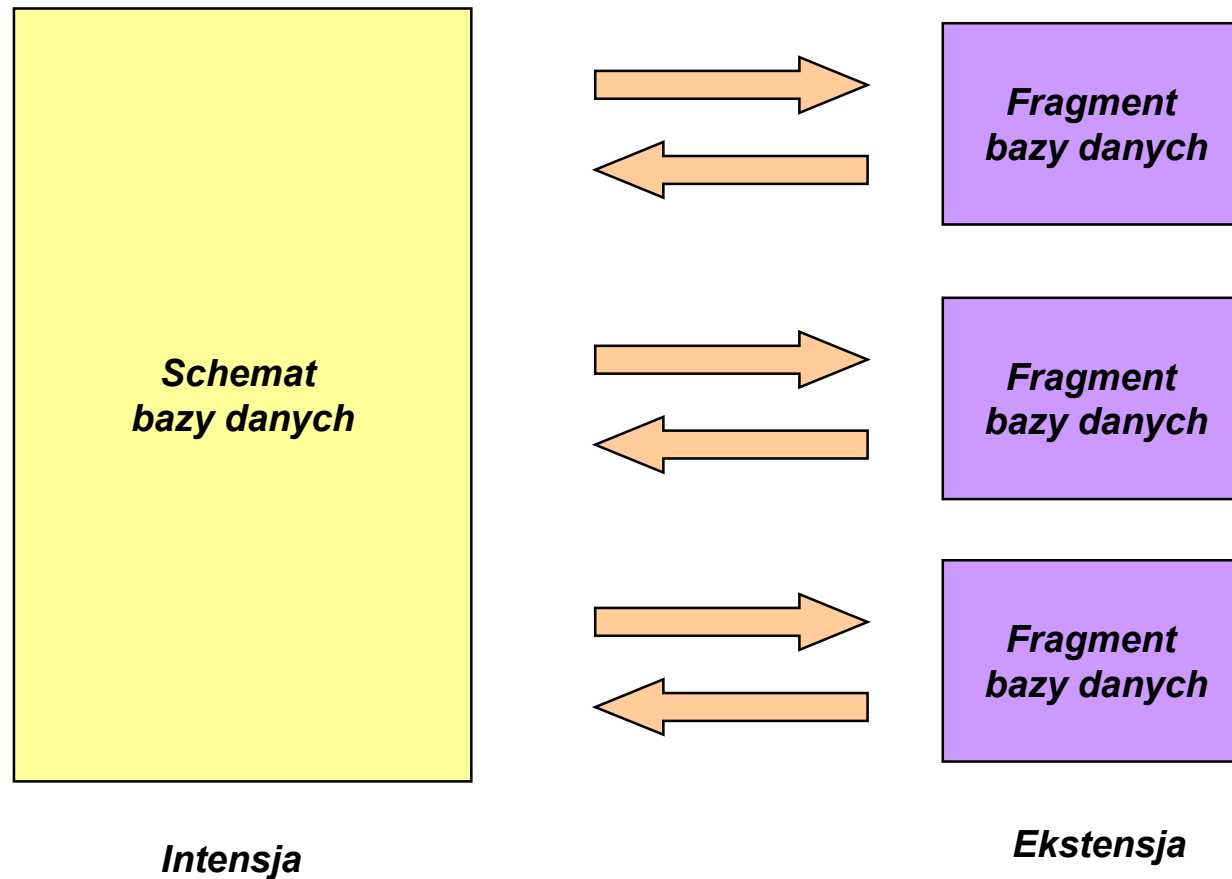
- ★ Jednym z powodów, dla których bazy danych są używane przez wiele osób, jest to, że baza może być wykorzystywana do różnych celów.

- Np.: w bazi uniwersyteckiej dziekanaty będą zainteresowana danymi o studentach, układający podziały godzin danymi o salach i wykładowcach.

- ★ **Podzbiór danych będących w czymś zainteresowaniu nosi nazwę perspektywy**

1.7. Rozproszenie

- ★ W systemie rozproszonych baz danych jeden rzeczywisty schemat jest przechowywany w kilku oddzielnych fragmentach i/lub kopiach.
 - Np.: w bazie studenckiej dane o studentach mogą być przechowywane na wydziałach a dane kadrowo-płacowe w dziale administracyjnym



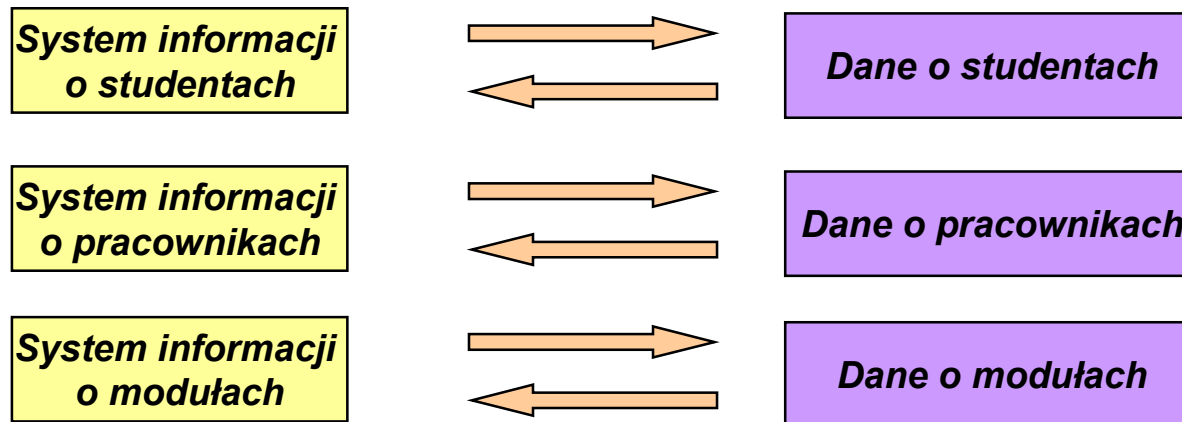
2. Podstawowe pojęcia

★ TEMATYKA

- Określenie potrzeb systemów baz danych
- Definicje pojęć:
 - Baza danych
 - System zarządzania bazą danych
 - Model danych
- Rozróżnienie wielu znaczeń terminu model danych
- Historia zarządzania danymi
- Wprowadzenie do procesu tworzenia bazy danych

2.1. Rozwój systemu informacyjnego

- ★ Informatyzacja zwyczajowo odbywała się po kawałku (każdy dział oddzielnie)
- ★ Informatyzacja „po kawałku” najczęściej prowadzi do powstania wielu oddzielnych systemów z własnymi plikami, danymi i programami



- ★ Izolowane systemy nie reprezentują sposobu w jaki działa przedsiębiorstwo.
- ★ Wymiana danych między systemami odbywa się poza nimi.
- ★ Informacje wygenerowane przez kilka systemów są dla personelu mniej wartościowe, gdyż nie dają pełnego obrazu działalności przedsiębiorstwa.
- ★ Dane mogą być powielane w wielu plikach używanych przez różne systemy.
- ★ W nowoczesnych systemach informatycznych dane są zintegrowane – dzięki czemu mogą być wykorzystywane na wiele sposobów

2.2. Bazy danych i systemy zarządzania bazą danych

- ★ W nowoczesnych systemach dane przechowuje się w sposób scentralizowany w jednym miejscu
- ★ Dąży się do integracji **Systemów Zarządzania Bazą Danych – SZBD** (**Database Management System – DBMS**)

2.2.1 Co to jest baza danych ?

- ★ **Baza danych to zbiór powiązanych i trwałych danych opisujących rzeczywistość**
- ★ Baza danych jest magazynem danych z nałożoną na niego strukturą wewnętrzną – zazwyczaj hierarchiczną
 - Np.: bazą danych może być zbiór informacji o studentach – kartoteka a w niej teczki z informacjami o studentach – informacje na oddzielnych kartkach: dane personalne, zaliczenia, praktyki itp..
- ★ Celem bazy danych jest najczęściej przechowywanie danych potrzebnych do codziennego funkcjonowania organizacji

2.2.1 Właściwości bazy danych

- ★ **Współdzielenie danych** – dane mogą być używane przez wiele osób nawet w tym samym czasie
- ★ **Integracja danych** – w bazie nie może być redundancji danych (danych powtarzających się oraz zbędnych)
- ★ **Integralność danych** – baza danych musi dokładnie odzwierciedlać swój obszar analizy (w przypadku danych powiązanych zmiany winny być propagowane). Integralność często zapewnia się ograniczając do niej dostęp.
- ★ **Bezpieczeństwo danych** – polega głównie na wydzielaniu zakresu danych do których poszczególni użytkownicy mają określone prawa dostępu
- ★ **Abstarkacja danych** – stanowią model rzeczywistości
- ★ **Niezależność danych** – dane winny być oddzielone od procesów na nich operujących (organizacja danych ma być niewidoczna dla użytkowników)

2.3. Co to jest System Zarządzania Bazą Danych ?

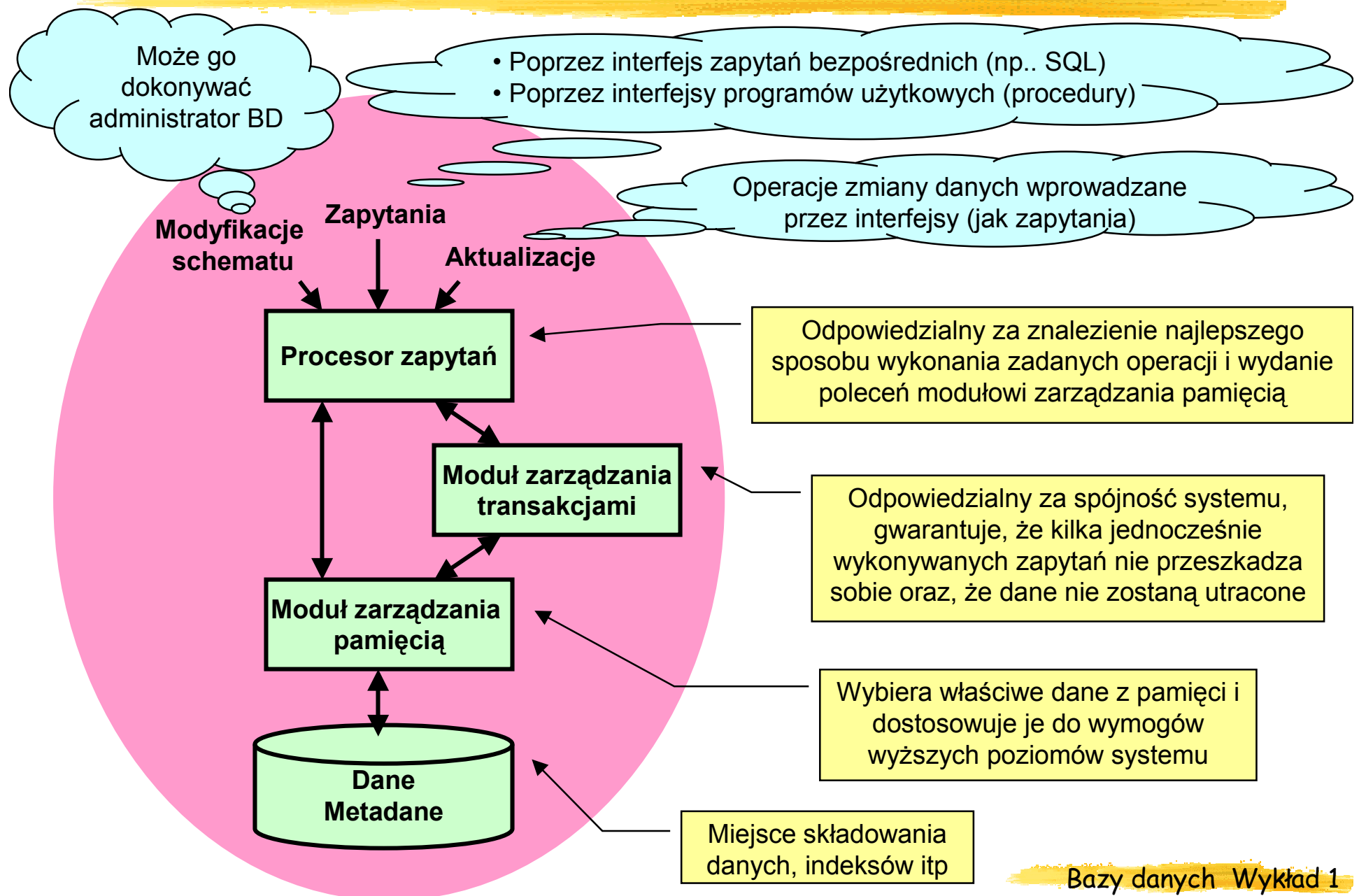
★ **System Zarządzania Bazą Danych SZBD (DBMS) jest to zbiór programów umożliwiających użytkownikom definiowanie, konstruowanie, manipulowanie bazą danych**

- definiowanie - specyfikacja typów danych, struktur, związków integralnościowych dla danych przechowywanych w bazie danych
- konstruowanie - proces zapisu danych na nośniku danych oraz ich kontrola przez SZBD
- manipulowanie - zapytania do bazy danych, modyfikacja, raportowanie

★ **Zadania Systemu Zarządzania Bazą Danych**

- **Pielęgnacja danych.** Umożliwienie użytkownikowi tworzenia nowych baz danych i określania ich struktury logicznej (przy pomocy specjalnego języka) oraz aktualizowania ich.
- **Wyszukiwanie danych.** Udostępnienie użytkownikowi możliwości tworzenia zapytań o dane.
- **Kontrola danych.** Sterowanie wielodostępem do danych

2.3.1. Architektura SZBD (DBMS)



2.4. Modele danych

- ★ Każda baza danych, a także każdy SZBD muszą stosować się do zasad określonego modelu danych.
- ★ Pojęcie model danych jest niejednoznaczne; najczęściej używane jest w znaczeniu:
 - architektura danych
 - projekt danych (zbiór wymagań w odniesieniu do danych)

2.4.1. Model danych jako projekt

- ★ W tym sensie model danych to zintegrowany i niezależny od implementacji zbiór wymagań w odniesieniu do danych
 - Np.: model danych dla kartoteki studentów, model danych dla płac itp..

2.4.2 Model danych jako architektura

- ★ W tym sensie model danych jest zbiorem ogólnych zasad posługiwania się danymi: model hierarchiczny, model sieciowy, model relacyjny, model obiektowy
- ★ Zbiór zasad które określają model danych można podzielić na trzy kategorie:
 - definicje danych (określają strukturę danych),
 - operowanie danymi (reguły operowania danymi),
 - integralność danych (określa które stany bazy są poprawne)

2.4.3. Typy architektonicznych modeli danych

- ★ Wyróżnia się trzy generacje architektonicznych modeli danych:
 - proste modele danych (pliki płaskie z rekordami)
 - klasyczne modele danych (hierarchiczne, sieciowe, relacyjne) - bazują na rekordach i trudno z samej bazy odczytać znaczenie informacji.
 - semantyczne modele danych (obektowe)

2.5. System baz danych

★ **System Baz Danych to skomputeryzowany system przechowywania i przetwarzania danych**

★ **System Baz Danych składa się z następujących elementów:**

- modelu danych
- oprogramowania
 - System Zarządzania Bazą Danych
 - Inne oprogramowanie
- baz danych

★ **Często do sytemu baz danych zalicza się również:**

- sprzęt
 - pamięci masowe
 - urządzenia systemowe
- użytkowników
 - programiści aplikacji – tworzą programy umożliwiające innym użytkownikom dostęp do bazy danych
 - użytkownicy końcowi – obsługujący bazę danych – wprowadzający dane, generujący raporty itp..
 - administratorzy BD – odpowiada za tworzenie i konserwację rzeczywistej bazy danych, odpowiada za jej bezpieczeństwo

2.6 Historia zarządzania danymi

- ★ **4000 p.n.e – 1900 n.e – ręczne zarządzanie zapisami 4000**
 - pierwszy znany zapis obejmujący majątek królewski i podatki w Sumatrze.
- ★ **1900 – 1955 – zarządzanie zapisami na kartach perforowanych**
 - Maszyna Hollerith'a (bazująca na wynalazku Jacquarda z 1801 roku).
- ★ **1955 – 1970 – programowane zarządzanie zapisami**
 - 1950 rok wynalezienie taśmy magnetycznej (na jednej mieściły się dane z ok. 10 000 kart perforowanych)
 - praca wsadowa na plikach i rekordach (nie daje możliwości wychwycenia błędów aż do zakończenia wsadu)
- ★ **1965 – 1970 – sieciowe zarządzanie danymi on-line**
 - praca on-line, przetwarzanie bieżących transakcji
 - pliki indeksowane
 - zastosowanie bębnow i dysków magnetycznych
 - trzypoziomowa architektura bazy danych

★ 1980-1995 – relacyjne zarządzanie danymi

- 1970 – model Codd'a
- powszechne SZBD na komputery PC
- najczęściej stosowany model
- architektura klient-server

★ 1995- Obiektowe bazy danych

- dane aktywne
- oparte o mechanizmy dziedziczenia
- Często przechowują dane różnych typów np.: głos, obraz itp..

2.7. Tworzenie baz danych

