



Bazy Danych

Wprowadzenie do zagadnień baz danych

Krzysztof Regulski

WIMiP, KISiM,

regulski@metal.agh.edu.pl

Konsultacje: wtorek, godz. 15:00 – 16:30

B5, pok. 409

Definicja bazy danych:

- **Baza danych** jest zbiorem struktur danych służących do organizowania i przechowywania danych. W takim modelu danych musi istnieć zbiór reguł określających wykorzystanie takich struktur danych w aplikacji.
- **Baza danych** to zintegrowana grupa ogólnie dostępnych zbiorów danych. Podstawową formą organizacji danych w bazie jest relacja.
- **Baza danych** – zbiór danych reprezentujących dany obszar analizy. W bazie danych mamy do czynienia z faktami, które mają miejsce. Dane powinny być trwałe.

System Zarządzania Bazą Danych (SZBD) - Database Management System (DBMS)

- Baza danych: Zbiór powiązanych ze sobą danych
- Baza danych jest zarządzana przez tzw. **system zarządzania bazą danych**, w skrócie SZBD.
- BD + SZBD = **system bazy danych**
- Z systemem bazy danych współpracują programy użytkowników, zwane **aplikacjami**. Zadaniem tych programów jest przetwarzanie danych, tj. wstawianie nowych danych, modyfikowanie danych już istniejących, usuwanie danych nieaktualnych, wyszukiwanie danych.
- Wszystkie omówione wyżej komponenty (tj. baza danych, SZBD i aplikacje) wchodzą w skład tzw. systemu informatycznego.

System Zarządzania Bazą Danych (SZBD) - Database Management System (DBMS)

- SZBD to pakiety programowe służące do zarządzania danymi. Podstawowymi elementami tych systemów są:
 - » język zapytań (SQL)
 - » generator raportów
 - » język manipulacji danymi
 - » język definicji danych
- SZBD umożliwiają oddzielenie aplikacji od bazy.
- SZBD dostarcza wygodnego i łatwego do używania środowiska

Geneza baz danych:

- Dawno, dawno temu... typowe aplikacje bazodanowe budowane były bezpośrednio w oparciu o system plików
- Wady bezpośredniego stosowania systemu plików w zagadnieniach bazodanowych:
 - » **Redundancja** i niespójność danych
 - Wiele formatów danych, duplikacja danych w różnych plikach
 - » Trudności z **dostępem** do danych
 - Potrzeba pisania nowych aplikacji dla nowych zadań
 - » Rozczłonkowanie danych — **wiele formatów** i plików
 - » Problem **integralności**
 - Warunki integralności (np. $\text{account balance} > 0$) stają się częścią kodu aplikacji
 - Trudności z nałożeniem nowych warunków integralności lub modyfikacją już istniejących

Geneza baz danych:

- Wady bezpośredniego stosowania systemu plików w zagadnieniach bazodanowych (c.d.):
 - » **Granulacja modyfikacji danych**
 - Błędy mogą pozostawić bazę w stanie nie zakończonej modyfikacji
 - Np. poprawny transfer pomiędzy rachunkami to dokonanie zmian salda obu rachunków
 - » **Współbieżny dostęp wielu użytkowników**
 - Dostęp współbieżny potrzebny dla efektywności
 - Niekontrolowany dostęp współbieżny może prowadzić do niespójności
 - Np. dwóch użytkowników odczytuje i modyfikuje saldo jednocześnie
 - » **Problemy związane z bezpieczeństwem**
- SZBD oferują rozwiązanie wszystkich tych problemów

Geneza baz danych:

- W latach 60 i 70 dominował model hierarchiczny (IMS) i sieciowy.
- W roku 1970 naukowiec z firmy IBM - E.F.Codd opublikował pracę ***A relational model for large shared data banks***, w której po raz pierwszy przedstawił założenia modelu relacyjnego.
- 1979 powstaje komercyjna wersja relacyjnej bazy danych – Oracle
- W 1983 IBM przedstawił DB2 pierwszą relacyjną bazę danych dla dużych komputerów.
- W tym samym czasie pojawił się inny konkurent na rynku relacyjnych baz - firma Relational Database Systems, która wkrótce zmieniła nazwę na Informix.
- W tym czasie istniały już relacyjne bazy danych dla komputerów osobistych dBase II (Ashton Tate).

- Architektura klient – serwer i wielowarstwowa
- Dane **multimedialne**
 - » problemy z porównywaniem wartości
 - » podejście obiektowo-relacyjne
 - » zarządzanie obiektami o dużych rozmiarach
- **Integracja** danych
 - » konieczność zarządzania rozproszonym (logicznie i fizycznie) środowiskiem
 - » hurtownie danych (opóźnienie w aktualizacji)

Typy baz danych:

- **operacyjne bazy danych** – znajdują zastosowanie w codziennym funkcjonowaniu organizacji, instytucji i firm. Baza taka przechowuje dane dynamiczne, czyli takie, które ulegają ciągłym zmianom i odzwierciedlają aktualny stan obiektu.

OLTP – *OnLine Transaction Processing*
(bieżące przetwarzanie transakcji)

- **analityczne bazy danych** – wykorzystywane są do przechowywania danych historycznych i informacji związanych z pewnymi wydarzeniami. Przechowywane tutaj dane są statyczne, bardzo rzadko ulegają zmianom i zawsze odzwierciedlają stan obiektów z pewnego ustalonego momentu.

OLAP – *OnLine Analytical Processing*
(oprogramowanie do analiz wielowymiarowych)

Cele systemów bazodanowych:

- **Elastyczny (efektywny) dostęp do danych** (przetwarzanie) - łatwe selekcjonowanie i prezentacja danych
- **Integralność danych** – dane są prawidłowe, spójne i aktualne
- **Bezpieczeństwo danych** – ochrona przed nieupoważnionym dostępem i uszkodzeniem
- **Obniżenie redundancji**
- **Niezależność od aplikacji** – fizyczna i logiczna organizacja danych jest oddzielona od aplikacji
- **Współdzielenie danych** (współbieżny dostęp) – umożliwienie różnym użytkownikom korzystania z tych samych (nie nadmiarowych) danych
- **Standaryzacja opisu** – jednolite definicje danych dotyczące ich nazw i opisu
- **Metadane** - dane o danych, strukturach dostępu, użytkownikach i ich prawach

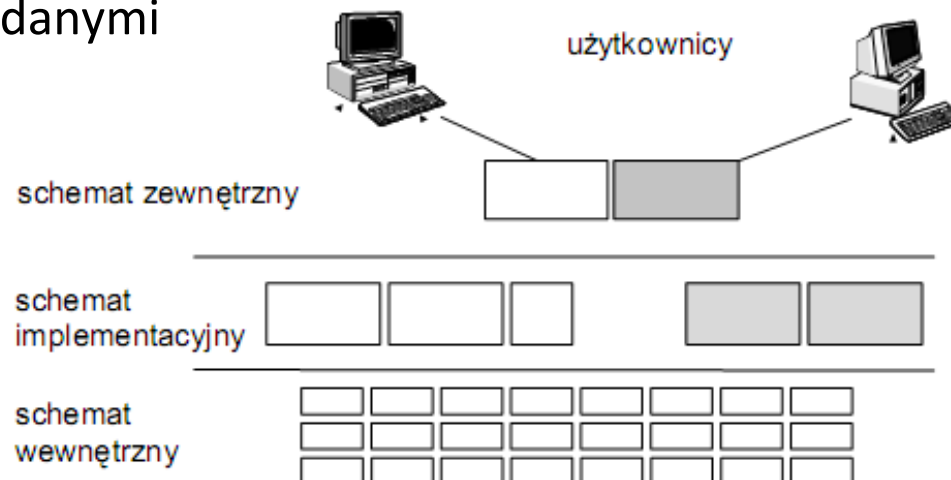
- Poprawność danych z punktu widzenia przyjętych kryteriów
 - » wierne odzwierciedlenie danych rzeczywistych
 - » wszystkie dane w bazie, na które nałożono pewne ograniczenia integralnościowe muszą te ograniczenia spełniać
- Odporność na anomalie będące wynikiem współbieżności dostępu do baz danych
- Odporność na błędy, awarie i inne anormalne sytuacje wynikające z zawodności środowiska sprzętowo-programowego
- Odporność na błędy użytkowników

- **Funkcje baz danych:**
 - » *aktualizujące* – zamieniają jeden stan w drugi (np. dodaj nowy produkt, zmień stawkę wynagrodzenia).
 - » *zapytań* – sprawdzanie stanu bazy danych (kwerendy pytające).
- **Funkcje SZBD:**
 - » *Zarządzanie plikami*
 - » *Wyszukiwanie informacji*
 - » *Zarządzanie bazą*
 - tworzenie i monitorowanie użytkowników
 - ograniczanie dostępu do plików
 - monitorowanie działania bazy danych

Architektura logiczna:

– Ujęcie ANSI/SPARC:

- » **Konceptualny** - poziom na którym zdefiniowany jest model świata w kategoriach pojęciowych użytkownika (interfejs użytkownika do bazy danych)
- » **Zewnętrzny** (implementacyjny) - języki wysokiego poziomu, umożliwiające definiowanie i dostęp do danych bez konieczności znajomości reprezentacji tych danych; odwzorowuje schemat wewnętrzny w struktury modelu danych wykorzystywanego w bazie danych
- » **Wewnętrzny** - służy do niezawodnego i trwałego składowania danych na nośnikach, na poziom ten składają się różnego rodzaju pamięci zewnętrzne, a także sposoby organizowania danych w tych pamięciach i techniki efektywnego zarządzania danymi



Architektura logiczna:

- Model poziomów abstrakcji:
 - » **Poziom widoków** – opis danych jest podobny do poziomu logicznego (pozbawiony takich szczegółów jak specyfikacje typów danych); zakres semantyczny każdego z widoków jest ściśle zorientowany (np. na konkretnego użytkownika) i stanowi zwykle jedynie małą część zakresu znaczeniowego całej bazy.
 - » **Poziom logiczny** opisuje dane i relacje pomiędzy nimi zgodnie z ich semantyką w świecie rzeczywistym.
 - » **Poziom fizyczny** opisuje, jak dane są składowane na odpowiednich nośnikach i jak realizowany jest elementarny dostęp do nich.

Niezależność danych:

Konsekwentnie skonstruowana trypoziomowa architektura bazy danych daje w efekcie niezależność schematu danych na wyższym poziomie od danych na niższym.

- **Logical Data Independence** – możliwość modyfikacji schematu logicznego bez konieczności przepisywania aplikacji.
- **Physical Data Independence** – możliwość modyfikacji schematu fizycznego bez zmiany schematu logicznego

W ogólności, interfejsy pomiędzy różnymi poziomami i komponentami powinny być dobrze zdefiniowane, tak aby zmiany w pewnych elementach nie wpływały istotnie na inne.

Schematy i instancje:

- **Schemat** – struktura bazy danych, część intensjonalna bazy danych
 - » **Schemat fizyczny**: projekt poziomu fizycznego bazy
 - Np. baza danych zawiera informację o zbiorze klientów i rachunków oraz relacjach pomiędzy nimi
 - » **Schemat logiczny**: projekt poziomu logicznego bazy

- **Instancja** – aktualna (w danym momencie czasu) zawartość bazy danych, część ekstensjonalna bazy danych

- Schemat: System wytwarzania
- Tabele:
 - » agregaty
 - » produkty
 - » plany produkcji
- Związki:
 - » plany produkcji pokazują obciążenie agregatów przez produkty
- Atrybuty:
 - » agregaty mają nazwę i czas dysponowany

- Zawartość: Buczek Technologie
- Agregaty:
 - » rozcinarka
 - » zgrzewarka
- Produkty:
 - » rura z/s okrągła 12/2
 - » rura b/s owalna 10/2
- Zaplanowano:
 - » rura z/s okrągła 12/2 na zgrzewarce 15.02. w ilości 500 kg

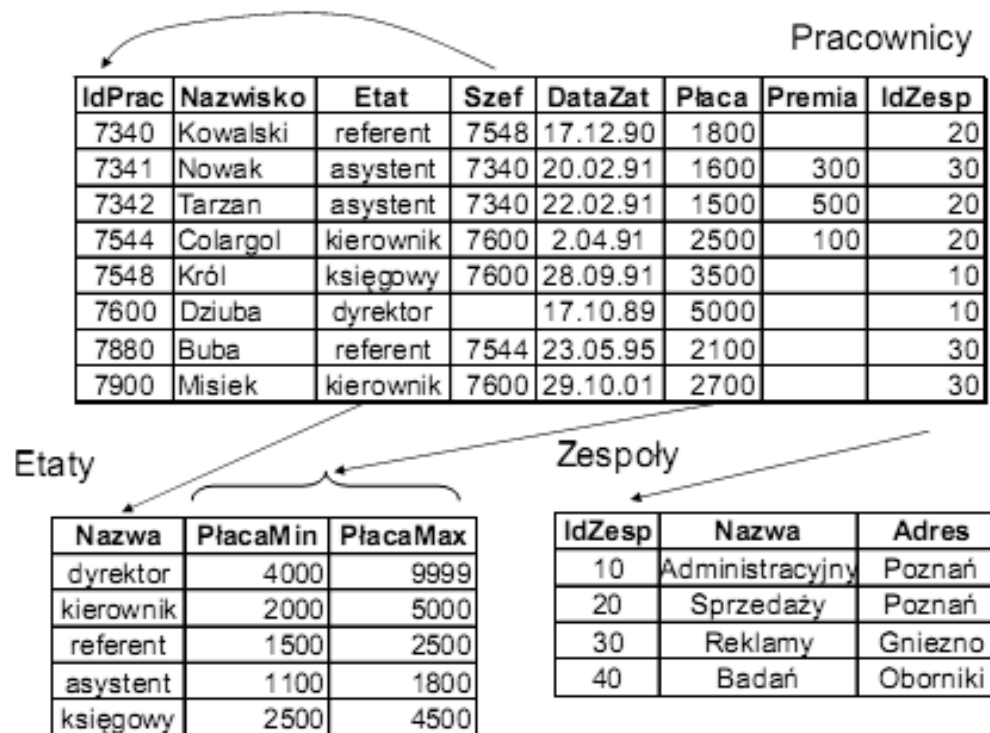
Formalizmy:

- **Formalizm reprezentacji** (*Patrick H. Winston*) to zbiór składniowych i semantycznych konwencji, które umożliwiają opisywanie rzeczy
- W terminologii baz danych idea formalizmu reprezentacji odpowiada pojęciu **modelu danych**

Modele danych:

- **Model danych** – zbiór narzędzi formalnych opisujący: syntaktykę danych, semantykę danych, relacje pomiędzy danymi oraz ograniczenia danych.
- Modele dla poziomów widoków i logicznego oparte na koncepcji **obiektu**
 - » Entity-Relationship model
 - » Model obiektowy
- Modele dla poziomów widoków i logicznego oparte na koncepcji **rekordu**
 - » Model relacyjny
 - » Modele: sieciowy i hierarchiczny
- Fizyczne modele danych

Przykładowa baza danych



- **Strukturami danych** modelu są w tym przypadku trzy relacje: Pracownicy, Zespoły, Etaty.
- Pierwsza z nich przechowuje dane o pracownikach, druga - o zespołach, w których ci pracownicy są zatrudnieni, a trzecia - zawiera katalog widełek płacowych.

- **Język SQL** (projektanci aplikacji, projektanci baz danych i administratorzy baz danych)
 - » jedyny sposób interakcji z bazą danych
 - » język deklaratywny
 - specyfikujemy tylko co chcemy otrzymać, nie w jaki sposób
 - » ustandaryzowany
 - producenci systemów komercyjnych i niekomercyjnych starają się implementować ten standard

```
SELECT nazwisko, płaca  
FROM pracownicy  
WHERE idZesp=30  
AND etat='kierownik'
```

- Aplikacje (użytkownicy końcowi)

- » formularze

- elektroniczne formularze z polami, listami, elementami wyboru
 - umożliwiają wstawianie, modyfikowanie, usuwanie, wyszukiwanie danych

- » raporty

- umożliwiają prezentowanie zawartości bazy danych:
 - teksty
 - wykresy
 - grafika

Więzy integralności

- Więzy integralności wskazują jakie asercje są poprawne a jakie nie
- Więzy statyczne to ograniczenia określone na stanie bazy danych, np.:
 - » nie można zaplanować operacji dla produktu na agregacie, którego nie ma w odpowiedniej karcie technologicznej
- Więzy przejść to reguły wiążące ze sobą stany bazy danych, np.:
 - » obciążenie agregatów nie może przekroczyć ich zdolności produkcyjnych

- **encja – relacja – klasa – tabela**: zbiór podobnych obiektów opisanych w jednolity sposób
- **krotka – obiekt (instancja klasy) – rekord**: zestaw wartości atrybutów opisujących jeden obiekt identyfikowany przez wyróżnione atrybuty lub nazwę
- **wież – asocjacja**: związek pomiędzy dwoma encjami (klasami) pokazujący jakie rekordy (obiekty) z jednej encji odpowiadają rekordom z drugiej i jaki jest charakter tej odpowiedniości

Przykład

Encja (klasa): WydziałyProdukcyjne

KodWydziału	NazwaWydziału	CharakterPracy
1	Odlewnia	ciągły
2	Obróbka	2 zmiany

Encja (klasa): Agregaty

KodWydziału	KodAgregatu	NazwaAgregatu
1	1	piec tyglowy
1	2	formierka
2	3	tokarka