



# Bazy Danych

**Andrzej M. Borzyszkowski**

**Instytut Informatyki  
Uniwersytetu Gdańskiego**

materiały dostępne elektronicznie  
<http://inf.ug.edu.pl/~amb>

© Andrzej M. Borzyszkowski

Bazy Danych

## Program wykładu

- Wstęp: historia, systemy zarządzania bazą danych
- Modelowanie danych: encje, związki, atrybuty
- Model relacyjny i algebra relacyjna
- Projektowanie baz danych
  - postaci normalne
- Język SQL – realizacja algebry relacyjnej
  - definiowanie danych
  - operowanie na danych: dostęp, aktualizacja
- Integracja ze środowiskiem programistycznym
- Współbieżność, blokady, transakcje
- Fizyczna organizacja plików, indeksy, optymalizacja

© Andrzej M. Borzyszkowski

Bazy Danych

2

## Bibliografia, oprogramowanie

- R. Elmasari, S. B. Navathe, *Wprowadzenie do systemów baz danych*, Helion, 2005
- R. Stones, N. Matthew, *Bazy danych i PostgreSQL*, Helion, 2002 (seria Wrox)
- Jeffrey D. Ullman, Jennifer Widom, *Podstawowy wykład z systemów baz danych*, WNT, 2001 (seria Klasyka Informatyki)
- Praca w laboratorium:
  - system PostgreSQL (open source, dostępny na systemy Linux i Windows)
  - wersja online: <https://sqliteonline.com/> pozwala emulować wiele SZBD, n.p. Postgres

© Andrzej M. Borzyszkowski

Bazy Danych

3

## Dane

- Przechowywanie danych:
  - bank: wszystkie transakcje
  - sprzedaż: klienci, towary
  - produkcja: części, dostawcy, proces produkcyjny
  - administracja państwowa: dane osobowe, miejsce zamieszkania, samochody
  - urząd skarbowy: dochody, podatki
  - szkoła wyższa: studenci, pracownicy, proces dydaktyczny
- Komputery służą (były zaprojektowane) do obliczeń
  - ale używamy je również do przechowywania i przetwarzania danych

© Andrzej M. Borzyszkowski

Bazy Danych

4

# Technologia

- Technologia komputerowa
  - plik: sekwencyjny zapis danych, dobry np. dla muzyki/filmu
  - albo trochę struktury: wiersze z polami, znaczniki

```
hplip:x:107:7:HPLIP system user,,,:/var/run/hplip:/bin/false
gdm:x:108:113:Gnome Display Manager:/var/lib/gdm:/bin/false
amb:x:1000:1000:Andrzej Borzyszkowski,,,:/home/amb:/bin/bash
postgres:x:111:115:PostgreSQL
administrator,,,:/var/lib/postgresql:/bin/bash
```
  - rekordy i pola
  - indeksy: dodatkowy plik zawierający adresy rekordów wyszukiwanych wg klucza
  - tzw. hasze (skrót): sam klucz wyznacza adres
    - gdy zachodzi kolizja adresów, to wyszukiwanie sekwencyjne

© Andrzej M. Borzyszkowski

Bazy Danych

5

# Technologia – problemy

- Problemy:
  - nieregularność danych, różna liczba pól w rekordach, różna wielkość rekordów
  - wielkość danych (tzn. liczba rekordów), wydajność
  - warunki spójności (=integralności)
    - nie są wyrażone bezpośrednio
    - nie są też gwarantowane
  - problem dostępu współbieżnego

© Andrzej M. Borzyszkowski

Bazy Danych

6

# Baza danych

- Zbiór powiązanych ze sobą informacji
  - opisujących *fakty i zdarzenia*
  - zachodzące w pewnym wycinku rzeczywistości
  - przechowywanych w sposób *trwały*
  - zorganizowanych w *strukturę* pozwalającą na ich szybkie wyszukiwanie i analizę
- Baza danych jest projektowana, konstruowana i wypełniana danymi
  - w określonym *celu*, ma określona zastosowania,
  - ma określonych *użytkowników*.
- Przykład: systemy ewidencyjne (rejestrują i analizują dane)
  - finansowo-księgowo, śledzenia aktywności użytkowników, centrale telefoniczne, systemy alarmowe

© Andrzej M. Borzyszkowski

Bazy Danych

7

# System Zarządzania Bazą Danych

- Zestaw programów umożliwiających *definiowanie, konstruowanie* baz danych, *manipulowanie i udostępnianie* zawartych w nich danych oraz *ochronę i konserwację*
  - SZBD (DBMS - *database management system*)
- *System bazy danych* = baza danych + system zarządzania bazą danych
- Operacje wykonywane na danych
  - tworzenie nowych plików
  - dodawanie danych do plików
  - wyszukiwanie i odczytywanie danych z plików
  - modyfikowanie i usuwanie danych w plikach
  - usuwanie całych plików (pustych lub nie)

© Andrzej M. Borzyszkowski

Bazy Danych

8

# Znane i lubiane implementacje SZBD

- Oracle
  - MySQL
  - Microsoft SQL Server
  - PostgreSQL
  - MongoDB
  - IBM DB2
  - dBASE
  - Sybase
  - Adabas
  - Informix
  - SQLite (niezupełnie SZBD)
  - MS Access (czy to w ogóle SZBD?)
- <http://is.gd/buzdrE> wikipedia:  
Lista\_systemów\_zarządzania\_relacyjnymi\_bazami\_danych
- [http://db-engines.com/en/ranking\\_trend](http://db-engines.com/en/ranking_trend) lista rankingowa

9

© Andrzej M. Borzyszkowski

Bazy Danych

# Otoczenie programistyczne

- Aplikacje zewnętrzne
- Programy do budowy aplikacji
- Programy narzędziowe (np. kopie zapasowe)
- Arkusze kalkulacyjne,
  - pakiety statystyczne,
  - inne programy do analiz wykorzystujących zgromadzone dane,
  - programy do grafiki,
  - edytory raportów, etc

10

© Andrzej M. Borzyszkowski

Bazy Danych

# Architektura klient-serwer

- Aplikacja bazodanowa
  - na serwerze zawierającym SZBD
  - na komputerze użytkownika (klient)
- Funkcje serwera
  - zarządzanie bazą danych (w tym dbałość o spójność danych)
  - zarządzanie kontami użytkowników
  - wykonywanie poleceń przekazanych przez klienta
- Funkcje klienta
  - kontakt z użytkownikiem (interfejs, np. graficzny)
  - wykonywanie lokalne obliczeń
  - komunikacja z serwerem
  - prezentacja danych otrzymanych od serwera

11

© Andrzej M. Borzyszkowski

Bazy Danych

# Języki zapytań

- Języki zapytań (*query*)
  - dawniej przewidywane zapytania sterowały projektem bazy
  - inne zapytania były bardzo nieefektywne
  - bazy relacyjne są neutralne
- Pierwsze języki zapytań
  - QBE (*query by example*) – zapytanie przez przykład
  - SQL – standardowy język zapytań
  - dziś rozwinął się do powszechnego standardu
- SQL
  - manipulowanie danymi – wstawianie, usuwanie, wyszukiwanie
  - definiowanie danych – tworzenie tabel
  - sterowanie danymi – np. prawa dostępu w bazie danych

12

© Andrzej M. Borzyszkowski

Bazy Danych

## Przykład

- Baza danych „uniwersytet”:
  - student (imię i nazwisko, rok i kierunek studiów, indeks)
  - przedmiot (nazwa i skrót, kierunek, liczba godzin)
  - obsada (przedmiot, prowadzący, semestr)
  - oceny (student, przedmiot, ocena)
  - wymagania (przedmiot, co wymaga)
- Dane podzielone są na wiele plików,
  - plik składa się z rekordów (wiersze/krotki, *tuple*)
  - kolumny/pola/attributy, stała liczba i format
  - komórki są atomowe
  - mają one ustalone z góry typy (napis/liczba/data ...)

© Andrzej M. Borzyszkowski

Bazy Danych

13

## Baza danych vs. dane rozproszone

- Integracja danych
  - baza może być traktowana jako połączenie informacji zawartych w odrębnych plikach danych
  - przykład: dziekanat posługuje się tabelą ocen by analizować zaliczenia studentów, inny dział oblicza wynagrodzenie pracowników korzystając z obsady danych
- Gdyby każdy z działów miał swoje dane:
  - nadmiarowość
    - niepotrzebnie zajęte miejsce
    - niebezpieczeństwo niespójności danych
  - odmienny format danych w każdym dziale
    - być może utrudniłoby to wymianę

© Andrzej M. Borzyszkowski

Bazy Danych

14

## Cechy systemów baz danych

- Opis struktury
  - SZBD przechowuje *katalog* czyli informację o strukturze wszystkich plików bazy danych
  - również informacje o użytkownikach i ich uprawnieniach
  - SZBD jest przystosowany do obsługi dowolnej bazy, struktura nie jest częścią aplikacji

© Andrzej M. Borzyszkowski

Bazy Danych

15

## Cechy systemów baz danych, c.d.

- Abstrakcja danych
  - aplikacja jest niezależna od struktury bazy danych
  - istnieje możliwość dodania pól, połączenia tabel, zmiany organizacji wewnętrznej, etc.
  - istnieje możliwość zmiany sposobu dostępu do danych
  - w podejściu obiektowym aplikacja może być niezależna od operacji: wywołuje metody, implementacja operacji w metodzie

© Andrzej M. Borzyszkowski

Bazy Danych

16

## Cechy systemów baz danych, 3

- Spójność (integralność) danych
  - system pozwala zdefiniować własności wymagane od danych
  - system sprawdza te własności
    - zmniejsza ryzyko zapisania błędnych danych w bazie (błędy użytkowników)
    - zmniejsza/niweluje ryzyko błędów sprzętowych/awarii
    - zmniejsza/niweluje ryzyko błędów wskutek dostępu współbieżnego

## Cechy systemów baz danych, 4

- Poufność danych, obsługa perspektyw
  - różni użytkownicy mają dostęp do różnych danych
  - nie każdy użytkownik bazy powinien mieć dostęp do wszystkich danych,
    - dane zbiorcze/indywidualne, czytanie/zapis
  - może nawet nie wiedzieć o istnieniu niektórych danych
- Autoryzacja dostępu: system kont z hasłami, różne systemy identyfikacyjne

## Cechy systemów baz danych, 5

- Współdzielenie danych
  - fragmenty danych mogą być używane przez wielu użytkowników jednocześnie (współbieżność)
  - problem czytelników i pisarzy
  - *transakcja* - niepodzielna operacja dokonana przez jednego z użytkowników, izolowana od innych operacji
- Niezawodność
  - możliwość odtworzenia bazy sprzed awarii sprzętowej czy programowej
- Wydajność
  - struktura odpowiednia do wyszukiwania danych (indeksy)

## Zalety rozwiązań bazodanowych

- Ograniczanie nadmiarowości (redundancji)
  - większy wysiłek przy wprowadzaniu danych
  - większa zajętość miejsca (dziś mniej ważny argument)
  - niespójność danych (błędy w niektórych kopiach, rozbieżność wprowadzanych danych)
- Ale: nadmiarowość może być pożyteczna
  - dane bliżej użytkownika końcowego
  - dane wynikowe przechowywane w celu dalszego użycia
  - wniosek: nadmiarowość trzeba kontrolować

## Zalety rozwiązań bazodanowych, c.d.

- Różnorodność interfejsów
  - baza danych jest jedna, ale różni użytkownicy mogą różnie ją widzieć
  - narzędzia graficzne, formatki do wprowadzania danych, graficzne przedstawienie danych
  - dostęp przez witrynę internetową, przez aplikację mobilną, bezpośrednio poprzez SQL
- Definiowanie reguł
  - automatyczne wnioskowanie na podstawie danych
  - automatyczne podejmowanie odpowiednich działań
  - również dbałość o spójność (wartości czy zależności pomiędzy danymi, *business rules*)

© Andrzej M. Borzyszkowski

Bazy Danych

21

## System BD – użytkownicy

- Użytkownik końcowy
  - komunikuje się z bazą
  - realizuje swoje cele za pomocą udostępnionego mu interfejsu (system formularzy, procesora zapytań SQL)
- Rodzaje użytkowników końcowych:
  - dorywczy: inne potrzeby za każdym podejściem
  - naiwny użytkownik: standardowe i powtarzalne czynności, używa formularzy
  - doświadczony użytkownik: wykonuje niestandardowe operacje, używa SQL

© Andrzej M. Borzyszkowski

Bazy Danych

22

## System BD – użytkownicy

- Programista aplikacji
  - określa wymagania użytkowników końcowych (naiwnych)
  - tworzy programy umożliwiające użytkownikom końcowym dostęp do bazy
- Projektant bazy
  - identyfikuje dane do przechowania, projektuje struktury, przewiduje perspektywy dla różnych użytkowników
- Administrator
  - specjalista z dziedziny IT (*Information Technology*)
  - zakłada bazę danych, implementuje kontrolę dostępu do bazy, monitoruje wykorzystanie, odpowiada za wydajność systemu i za bezpieczeństwo danych

© Andrzej M. Borzyszkowski

Bazy Danych

23