

Rozdział 1 Wprowadzenie do baz danych

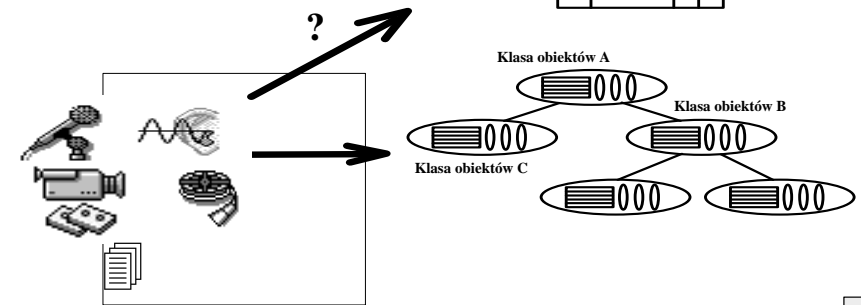
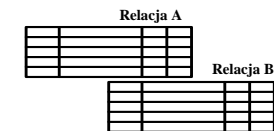


Model danych

Obiekty świata
rzeczywistego



Obiekty modelu danych



Funkcje systemu zarządzania bazą danych

Wymagania

- spójność bazy danych po awarii
- trwałość danych
- wielodostęp
- poufność danych
- wydajność
- rozproszenie danych

Funkcje DBMS

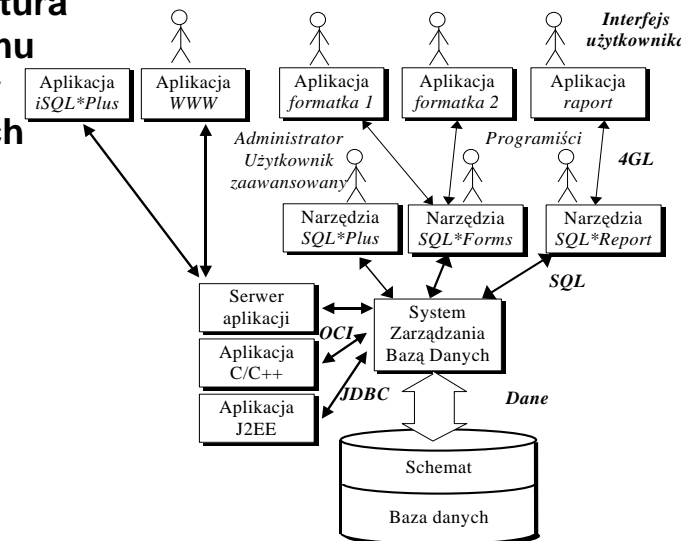
- odtwarzanie bazy danych
- archiwizacja bazy danych
- zarządzanie współbieżnością transakcji
- identyfikacja użytkowników, autoryzacja dostępu, szyfrowanie danych, kontrola dostępu
- fizyczne struktury danych, optymalizacja zapytań
- dwufazowe zatwierdzanie transakcji, replikacja danych



Architektura systemu bazy danych

Użytkownicy końcowi (naiwni)

Interfejs
użytkownika



Definicja transakcji

Transakcja

Jest sekwencją logicznie powiązanych operacji na bazie danych, która przeprowadza bazę danych z jednego stanu spójnego w inny stan spójny. Systemy bazy danych umożliwiają łączenie operacji w transakcje i gwarantują poprawne zarządzanie transakcjami.

- Transakcja przelewu kwoty N z konta A na konto B:

```
BEGIN.  
// Odejmij kwotę N z konta A;  
UPDATE konta.  
SET stan = stan - N.  
WHERE id_konta = A;  
// Dodaj do konta B kwotę N;  
UPDATE konta.  
SET stan = stan + N.  
WHERE id_konta = B;  
COMMIT;  
END.
```



Spójność bazy danych

Definicja

- Baza danych jest spójna jeżeli jej stan jest zgodny ze stanem reprezentowanego przez nią fragmentu świata rzeczywistego.
- Baza danych jest spójna jeżeli są w niej spełnione wszystkie ograniczenia referencyjne i integralnościowe.

Zagrożenia spójności bazy danych

- awarie sprzętu komputerowego i oprogramowania
- utrata danych w wyniku uszkodzenia pamięci masowej
- użytkownicy
- współbieżny dostęp do danych



Spójność bazy danych. Własności transakcji

A(tomicity) C(onsistency) I(solation) D(urability)

Atomowość

- zbiór operacji wchodzących w skład transakcji jest niepodzielny; albo zostaną wykonane wszystkie operacje transakcji, albo żadna.

Spójność

- poprawne wykonanie transakcji przeprowadza bazę danych z jednego stanu spójnego do innego stanu spójnego.

Izolacja

- transakcje są od siebie logicznie odseparowane. Mogą wzajemnie oddziaływać na siebie w taki sposób jak gdyby były wykonywane sekwencyjnie.

Trwałość

- wyniki zatwierdzonych transakcji nie mogą zostać utracone, niezależnie od awarii systemu.



Relacyjny model danych

Historia

Podstawy teoretyczne:

- E.Codd: „*A Relational Model for Large Shared Data Banks*”, rok 1970

Projekty badawcze (lata siedemdziesiąte):

- IBM System R, SEQUEL (1978)
- Berkley Ingres

Produkty komercyjne (lata osiemdziesiąte):

- Oracle (1979), Ingres (1979), DB2 (1982), Sybase, Informix, SQL Server
- Paradox, dBase
- Postgres, MySQL

Standardy

- ANSI X3.135, X3.168
- ISO „Database Language SQL”: SQL89, SQL92, SQL99



Struktury danych (2)

Definicja I.

- Relacja (lub wystąpienie relacji) r o schemacie $R(A_1, \dots, A_k)$, oznaczona $r(R)$, jest zbiorem krotek:
$$r = \{t_1, t_2, \dots, t_n\}.$$
- Wystąpienie relacji $r(R)$ nazywane jest również stanem relacji.
- Każda krotka jest uporządkowaną listą k wartości:

$$t = \langle v_1, v_2, \dots, v_k \rangle,$$

gdzie każda wartość v_i , $1 \leq i \leq k$, jest elementem dziedziny $\text{dom}(A_i)$ lub wyróżnioną wartością null (wartość pusta).

Definicja II.

- Relacja $r(R)$ jest podzbiorem iloczynu kartezjańskiego dziedzin definiujących schemat relacji r :
$$r(R) = (\text{dom}(A_1) \times \text{dom}(A_2) \times \dots \times \text{dom}(A_k)).$$



Jawne ograniczenia integralnościowe

Ograniczenie na unikalność krotek relacji.

- Każdy podzbiór SK atrybutów relacji R , taki że dla dwóch dowolnych krotek ze zbioru $r(R)$, prawdą jest, że: $t_1[SK] \neq t_2[SK]$, jest nazywany nadkluczem (ang. super key) relacji R . Każda relacja ma co najmniej jeden nadklucz - schemat relacji.

$SK(\text{STUDENT}) = \{\text{NrIndeksu}, \text{Nazwisko}, \text{Wiek}\}.$

Kluczem relacji R nazywamy taki nadklucz tej relacji, że żaden z jego podzbiorów nie jest nadkluczem.

$K(\text{STUDENT}) = \{\text{NrIndeksu}\}.$

- Jeżeli relacja zawiera więcej niż jeden klucz, to są one nazywane kluczami potencjalnymi (ang. candidate key). Jeden z nich może pełnić rolę klucza podstawowego (ang. primary key) relacji. Pozostałe są wtedy kluczami drugorzędnymi.

Ograniczenie integralności klucza podstawowego

- Wartości klucza podstawowego identyfikują krotki relacji i dlatego nie mogą przyjmować wartości pustych.

Ograniczenie referencyjne

- Dane są relacje R_1 i R_2 . Podzbiór FK atrybutów relacji R_1 nazywamy kluczem obcym R_1 (ang. foreign key), jeżeli:
 - Atrybuty w FK mają taką samą dziedzinę jak atrybuty klucza podstawowego PK relacji R_2 ;
 - Dla każdej krotki t_1 relacji R_1 istnieje krotka t_2 relacji R_2 , taka że: $t_1[\text{FK}] = t_2[\text{PK}]$, lub $t_1[\text{FK}] = \text{null}$.



Podjęcie relacyjne

Własności bazy danych

- Widziana jako zbiór relacji
- Nazwy tabel nie mogą się powtarzać
- Operatory: selekcja, projekcja, iloczyn kartezjański, połączenie, suma, przecięcie, różnica
- Dane nie są powiązane za pomocą żadnych wskaźników, powiązania między danymi tylko przez porównywanie wartości pól
- Nieproceduralny język komunikowania się z bazą danych
- Użytkownik nie określa sposobu dostępu do danych i nie zna ich fizycznej organizacji

Własności relacji

- W tabeli nie ma żadnych powtarzających się krotek
- W tabeli nie ma atrybutów o powtarzających się nazwach
- Kolejność krotek w tabeli jest całkowicie nieokreślona
- Wartościami pól są dane atomowe
- Krotki są identyfikowane tylko i wyłącznie na podstawie wartości pól, lokalizacja krotki nie ma żadnego wpływu na tożsamość krotki

