

Egzamin z baz danych

01 lutego 2010

Zaprojektować bazę danych do przechowywania danych o drużynach sportowych, zawodnikach i kibicach. Dla każdej drużyny przechować jej nazwę, zawodników i kapitana drużyny (jednego z zawodników) oraz barwy klubowe. Dla każdego zawodnika zapisać imię, nazwisko, datę urodzenia, telefon i datę następnego badania lekarskiego. Dla każdego kibica zapisać imię, nazwisko, datę urodzenia, telefon i datę ostatniego wyroku. Kibic może kibicować tylko jednej drużynie. Podobnie, zawodnik może grać tylko w jednej drużynie. Baza danych ma być w co najmniej trzeciej postaci normalnej. Milc widziany diagram wymagań informacyjnych. Z odpowiedzi musi też wynikać, jakie są tabele, kolumny i więzy integralności.

2. Relacja R jest w pierwszej postaci normalnej. R ma $n > 4$ atrybutów A_1, A_2, \dots, A_n i dokładnie dwa klucze $\{A_1, A_2\}$ i $\{A_1, A_3\}$. Podać liczbę jej nadkluczy. Ustalić, czy musi być w 2NF, 3NF, BCNF. Jeśli musi, uzasadnić dlaczego. Jeśli nie musi, skonstruować kontrprzykład.

3. Zakładamy, że relacja R ma r krotek, a relacja S ma s krotek. X jest kolumną relacji R , a C pewnym warunkiem logicznym. Ile minimalnie i maksymalnie krotek będzie w relacjach $R \bowtie S$, $\pi_X(R)$, $S - \sigma_C(S)$, $(S \bowtie R) \bowtie S$.

4. Dane są relacje: Okret(nazwa, wodowanie), Bitwa(nazwa, data), Rezultat(okret, bitwa, wynik). Podkreślono klucze główne. Kolumny okret i bitwa w relacji Rezultat są kluczami obcymi do oczywistych tabel. Kolumna wynik zawiera jedną z trzech wartości: cały, trafiony, zatoniony. (a) Napisać zapytanie SQL wyszukujące nazwy wszystkich okrętów, które były trafione w pewnej bitwie, a potem wzięły udział w jakiejś innej bitwie, a następnie zostały zatopione w pewnej kolejnej bitwie (b) Napisać zapytanie SQL wyszukujące nazwy bitew, w których brało udział co najmniej 10 okrętów i co najmniej 33% wszystkich zostało zatopionych i co najmniej 33% wszystkich zostało trafionych, ale nie zatopionych.

5. Relacje są takie jak w zadaniu 4. Relacja Okret ma 1000 krotek i zajmuje 100 stron dyskowych. Relacja Bitwa ma 100 krotek i zajmuje 10 stron dyskowych. Relacja Rezultat ma 10000 krotek i zajmuje 1000 stron dyskowych. Dane jest zapytanie (okręty zatopione niecały rok po wodowaniu):

```
SELECT o.Nazwa
FROM Okret o, Bitwa b, Rezultat r
WHERE o.nazwa = r.okret AND b.nazwa = r.bitwa
AND r.wynik = 'zatoniony'
AND o.wodowanie + 365 > b.data;
```

Zaprojektować fizyczne struktury danych najlepiej wspomagające to zapytanie, opracować plan wykonania tego zapytania i oszacować jego pesymistyczny koszt jako sumę operacji wejścia-wyjścia, przy założeniu, że mamy do dyspozycji 100 ramek pamięci operacyjnej.

Egzamin poprawkowy z baz danych

01 marca 2010

Relacja *Cząsteczka* opisuje zawartość zamkniętego pojemnika z pewną substancją gazową. Ma siedem pól: *ID* — identyfikator cząsteczki, *x*, *y*, *z* — współrzędne kartezjańskie położenia cząsteczki i *dx*, *dy*, *dz* — współrzędne kartezjańskie wektora prędkości cząsteczki. Podać zależności funkcyjne w tej relacji, wskazać klucze. Określić, w której postaci normalnej jest ta relacja. Uzasadnić krótko odpowiedź. Zignorować odkrycia Heisenberga.

2. Dana jest relacja *Połączenie*(*stacjaA*, *stacjaB*, *długość*). Dwie pierwsze kolumny łącznie stanowią klucz tej relacji. Krotka takiej relacji oznacza, że istnieje jednokierunkowe bezpośrednie połączenie *stacjaA* ze *stacjąB* o długości *długość*. Zapisać w algebrze relacji zapytanie o takie pary stacji, między którymi można przejechać z dokładnie jedną przesiadką i nie ma między nimi połączenia bezpośredniego.
3. Relacja taka jak w 2. Zapisać w SQL zapytanie o pary stacji, między którymi można przejechać z dokładnie dwiema przesiadkami i nie ma między nimi połączenia bezpośredniego ani połączenia jedno-przesiadkowego. Dla każdej pary wypisać minimalną długość trasy z dwiema przesiadkami, która łączy te dwa miasta.
4. Relacja taka jak w 2. Załóżmy, że połączenia między stacjami nie zawierają cykli. W dowolnym znanym języku zapisać zapytanie/program o minimalny czas podróży między każdą parą różnych połączonych stacji przy założeniu, że czas przesiadki to zawsze 10 minut, a pociągi pokonują 60 jednostek długości na godzinę (jak to zwykle w Izraelu).
5. Relacja taka jak w 2. Dane jest zapytanie o długość bezpośredniego połączenia Warszawy z Krakowem:

```
SELECT długość FROM Połączenie  
WHERE stacjaA = 'Warszawa' AND stacjaB = 'Kraków'
```

W relacji *Połączenie* jest 100000 wierszy zapisanych w 1000 blokach dyskowych. Jest dostępnych 100 ramek pamięci operacyjnej. Wskazać jak najlepszy indeks do realizacji tego zapytania i obliczyć koszt realizacji tego zapytania, jeśli taki indeks jest już założony.

Egzamin z Baz Danych

1 lutego 2007, 9:00 – 12:00

+ **Zadanie 1:** Zapisz w algebrze relacji następujące zapytanie:

```
SELECT a, b
FROM c, d
WHERE c.e=d.f
UNION
SELECT g, h
FROM i
```

Pisz wyraźnie, skreślenia i pomazania będą interpretowane na Twoją niekorzyść.

+ **Zadanie 2:** Podaj przykład realizacji trzech transakcji, która unika kaskadowych wycofań, ale nie jest ścisła. Odpowiedź uzasadnij.

+ **Zadanie 3:** Masz daną tabelę Oceny(Przedmiot, Student, Stopień) (stopień to liczba, pozostałe pola to napisy, pola podkreślone tworzą klucz). Stopniem S nazwiemy taki stopień, który jest najrzadziej wystawiany spośród stopni mających nieparzystą liczbę wystawień (możesz założyć, że istnieje tylko jeden taki stopień). Napisz w SQL zapytanie, wybierające przedmiot (ew. przedmioty), z którego zostało wystawione najwięcej ocen o stopniu S .

? **Zadanie 4:** Jakie są modelowe fazy przetwarzania zapytania? Jaką przestrzeń przeszukuje algorytm przeszukiwania podczas optymalizacji? Jakie typy indeksów są najczęściej używane w bazach danych i czym się one od siebie różnią?

+ **Zadanie 5:** Mamy następującą relację i jej zależności funkcyjne:

$$R = \{ABCDEFGHI; A \rightarrow ABCDEFG; BC \rightarrow A; BD \rightarrow E; F \rightarrow A\}$$

- Co jest kluczem? Udowodnij. Ile jest kluczy?
- Dokonaj dwóch dekompozycji, do trzeciej postaci normalnej i BCNF.

Rozwiązanie Następujące zapytanie ma niepusty wynik wtedy i tylko wtedy, gdy istnieje kolejne połączenie Gdyni z Warszawą z co najwyżej dwoma przesiadkami.

$$\sigma_{(a_1='Warszawa' \vee b_1='Warszawa') \wedge (a_3='Gdynia' \vee b_3='Gdynia')}$$

$$(\rho_{O_1(a_1, b_1, d_1)}(\text{odcinki}) \bowtie_{a_1=a_2 \vee a_1=b_2 \vee b_1=a_2 \vee b_1=b_2}$$

$$\rho_{O_2(a_2, b_2, d_2)}(\text{odcinki}) \bowtie_{a_2=a_3 \vee a_2=b_3 \vee b_2=a_3 \vee b_2=b_3}$$

$$\rho_{O_3(a_3, b_3, d_3)}(\text{odcinki}))$$

Zadanie 4 (ZJ)

Dany jest schemat tabeli reprezentującej odcinki dwukierunkowych linii kolejowych z zadanie 3. Napisz w SQL zapytanie znajdujące minimalną długość podróży z Warszawy do Gdyni przy co najwyżej dwóch przesiadkach?

Rozwiązanie

```
WITH odc AS (SELECT stacjaA as stA, stacjaB as stB, dlugosc AS dl FROM odcinki
            UNION ALL
            SELECT stacjaB, stacjaA, dlugosc FROM odcinki)
SELECT MIN(razem.dl)
FROM (
  SELECT stA, stB, dl FROM odc
  UNION ALL
  SELECT o1.stA AS stA, o2.stB AS stB, o1.dl + o2.dl AS dl
  FROM odc o1, odc o2
  WHERE o1.stB=o2.stA
  UNION ALL
  SELECT o1.stA AS stA, o3.stB AS stB, o1.dl+o2.dl + o3.dl AS dl
  FROM odc o1, odc o2, odc o3
  WHERE o1.stB = o2.stA AND o2.stB = o3.stA
) razem
WHERE razem.stA = 'Warszawa'
AND razem.stB = 'Gdynia';
```

Rozwiązanie zabawne z kilku prac, w niektórych nawet prawidłowo. Pomysł podobny do zaprezentowanego rozwiązania zadania 3:

```
WITH odc AS (SELECT stacjaA as stA, stacjaB as stB, dlugosc AS dl FROM odcinki
            UNION ALL
            SELECT stacjaB, stacjaA, dlugosc FROM odcinki
            UNION ALL
            SELECT stacjaA, stacjaA, 0 FROM odcinki
            UNION ALL
            SELECT stacjaB, stacjaB, 0 FROM odcinki)
SELECT MIN(razem.dl)
FROM (SELECT o1.stA AS stA, o3.stB AS stB, o1.dl + o2.dl + o3.dl AS dl
      FROM odc o1, odc o2, odc o3
      WHERE o1.stB = o2.stA AND o2.stB = o3.stA) razem
WHERE razem.stA = 'Warszawa'
AND razem.stB = 'Gdynia';
```


Zadanie 5 (KS)

Wykonaj zadania 3 i 4 przy braku założenia o ograniczeniu na liczbę przesiadek. Czy da się użyć tu SQL lub algebry relacji? Jeśli nie, użyj jednego z języków poznanych na wykładzie lub w najgorszym wypadku PL/SQL.

Rozwiązanie Oto najkrótsze i najbardziej estetyczne rozwiązanie, jakie pojawiło się na egzaminie. W pierwszym kroku korzystamy z Datalogu. W drugim kroku wykonujemy już zwykłe zapytanie SQL. *Trasy* to relacja intensjonalna (perspektywa).

Trasy(*stacjaA*, *stacjaB*, *dlugosc*)

Trasy(*a*, *b*, *d*) \leftarrow *Odcinki*(*a*, *b*, *d*)

Trasy(*a*, *b*, *d*) \leftarrow *Odcinki*(*b*, *a*, *d*)

Trasy(*a*, *b*, *d*) \leftarrow *Trasy*(*a*, *e*, *d*₁) \wedge *Trasy*(*e*, *b*, *d*₂) \wedge *d* = *d*₁ + *d*₂

```
SELECT MIN(dlugosc)
FROM Trasy
WHERE stacjaA = 'Warszawa'
AND stacjaB = 'Gdynia';
```

Nieistnienie trasy jest równoważne zwróceniu NULL przez to zapytanie.

Jaar

- Nowy Dokument tekstowy.txt
1. Dana sa dwie tabele: Zamowienia(nr, klient) [kluczem glownym jest nr]
i Pozycje(zam, nr, produkt, ilosc) [kluczem glownym jest zbior {zam, nr},
zam to klucz obcy do tabeli Zamowienia.]

Napisac w SQL zapytanie znajdujace klientow, ktorzy zlozyli co najmniej jedno zamowienie, takie ze dla kazdej pozycji tego zamowienia, istnieje zamowienie z pozycja opiewajaca na wieksza ilosc tego samego produktu.

2. Schemat tabel taki jak w zadaniu 1. zapytanie:

```
SELECT z.klient, p.produkt
FROM Zamowienie z, Pozycja p
WHERE z.nr = p.zam
AND ilosc > ALL (SELECT ilosc FROM Pozycja p1
                 WHERE p1.zam = p.zam
                 AND p1.towar <> p.towar);
```

Pamieci operacyjnej mamy 100 ramek, Zamowienia miesci sie na 1000 stronach i ma 50000 wierszy a Pozycja na 10000 stron i ma 1000000 wierszy. Podaj optymalny plan wykonania tego zapytania i oszacuj jego koszt. Mozesz zalozyc istnienie dowolnych indeksow [nalezy oczywiscie wskazac z ktorych sie korzysta].

3. Dana jest tabela o n kolumnach {A1, A2, ..., An}. Obowiazuja w niej wszystkie zaleznosci funkcyjne postaci $A_i \rightarrow A_j$, gdzie $i < j$. Sprowadz te tabele do trzeciej postaci normalnej. Odpowiedz nalezy uzasadnic.

4. Na tabeli Magazyn(nazwa, ilosc) wykonano dwie transakcje T1 i T2:

```
T1: SELECT ilosc INTO v_ilosc FROM Magazyn where nazwa = 'gredziel';
     UPDATE Magazyn SET ilosc = v_ilosc + 3 where nazwa = 'gredziel';
```

```
T2: SELECT ilosc INTO v_ilosc FROM Magazyn where nazwa = 'gredziel';
     UPDATE Magazyn SET ilosc = v_ilosc + 1 where nazwa = 'gredziel';
```

Podaj wszystkie nieszeregowalne plany wykonania tych dwoch transakcji. Odpowiedz uzasadnij.

5. Dany jest nastepujacy schemat tabeli reprezentujacej odcinki dwukierunkowych linii kolejowych [w tabeli odcinki nie powtarzaja sie]:

```
CREATE TABLE odcinki (
  stacjaA    VARCHAR(15) NOT NULL,
  stacjaB    VARCHAR(15) NOT NULL,
  dlugosc    INTEGER
);
```

Napisz w jakims formalizmie zapytanie:
* czy z warszawy mozna dojechac do Gdyni w co najwyzej 8 przesiadkach?

Zadanie 1.
 Tabela w postaci normalnej Boyce'a Codd'a ma 3 kolumny i jeden klucz złożony z 2 kolumnami.
 Podaj minimalną i maksymalną moc domknięcia zbioru zależności funkcyjnych z danej tabeli.
2 BOC⁷ min = max = 1

Zadanie 2.
 Dane jest zapytanie:

```
SELECT * FROM Emp
WHERE mgr IN (SELECT empno FROM Emp WHERE Sal > 1000);
```

 Zapisz je w postaci wyrażenia algebry relacji (może być w postaci drzewa) i podaj plan jego wykonania. (możliwie najlepszy ale nie jest to tu wymagane)

Zadanie 3.
 SZBD ma bufory pamięci operacyjnej na 200 bloków dyskowych.
 Tabela Emp (empno NUMBER(10), ename VARCHAR(20), sal NUMBER(10), mgr NUMBER(10)) zajmuje 50000 bloków dyskowych. W jednym bloku mieszczą się informacje o 80 pracownikach. Kolumna empno jest kluczem. Wykonano następujące zapytanie:

```
SELECT m.empno, m.ename, sum(e.sal)
FROM Emp e, Emp m
WHERE e.mgr = m.empno
GROUP BY m.empno, m.ename
HAVING count(*) > 5;
```

Sort. po mgr
10-50.000 bloków 22
 Podaj jak najefektywniejszy plan wykonania tego zapytania i oszacuj koszt tego planu (podaj liczbę operacji odczytu i zapisu bloku dyskowego) przy założeniu, że nie korzystamy z żadnych indeksów i że na początku wszystkie bufory są puste a dane są tylko na dysku.
100.000 pracowników
1000.000 pracowników
1000.000 pracowników

Zadanie 4.
 Tabela Emp jest taka sama jak w zadaniu 3. Kolumna empno identyfikuje pracownika (klucz główny), ename to jego nazwisko, sal wynagrodzenie, a mgr to identyfikator szefa (klucz obcy odwołujący się do tej samej tabeli). Napisz zapytanie SQL, które wypisze nazwisko każdego pracownika, którego wszyscy podwładni zarabiają mniej niż on, ale jeden z tych podwładnych (oznaczymy jego wynagrodzenie S) do podwładnego o zarobkach większych niż S.

Zadanie 5.
 Tabela jest taka sama jak w zadaniu 4. Jej schemat niestety pozwala na istnienie podwładności cyklicznej, np:

EMPNO	ENAME	SAL	MGR
7782	CLARK	2450	7839
7839	KING	5000	7934
7934	MILLER	1300	7782

Napisz wyzwalacze (w języku, którego używałeś na laboratorium), które przypilnują, by do takiej sytuacji nigdy nie doszło.

Zadanie 6.
 Autobusy
 Firma komunikacji autobusowej chce gromadzić dane o:
 1. Posiadanych taborze autobusowym, m.in. typ, numer rejestracyjny, liczba miejsc.
 2. Rozkładzie jazdy, m.in. linia, skąd, dokąd, czas odjazdu, czas przyjazdu, dni kursowania.
 3. Kierowcach, m.in. nazwisko, imię, uprawnienia, adres, stan.
 4. Rezerwacji biletów na poszczególne kursy.

1. "Każda tabela mająca dokładnie dwie kolumny jest w trzeciej postaci normalnej"

Czy to zdanie jest prawdziwe? Jeśli tak, uzasadnij krótko odpowiedź. Jeśli nie, podaj kontrprzykład.

2. Podaj schemat relacyjnej bazy danych w co najmniej 3 postaci normalnej dla następującego zastosowania:

Ministerstwo Finansów pragnie uzbroić urzędy skarbowe w system, w którym rejestrowane będą nieruchomości i ich burżuazyjni właściciele. Celem systemu jest wykrywanie nieujawnionych dochodów. Każda nieruchomość leży na terenie pewnej gminy; każda firma jest zarejestrowana w pewnej gminie; każda osoba fizyczna jest zameldowana w pewnej gminie. Każda nieruchomość może mieć wielu współwłaścicieli. Każdy ze współwłaścicieli jest posiadaczem pewnego ułamka danej nieruchomości. Właścicielem może być osoba fizyczna, firma, gmina lub skarbowo państwa. Firmy mogą ukrywać swoje nieruchomości poprzez transferowanie ich do swoich spółek-córek. Dlatego w systemie potrzebna będzie jeszcze informacja o udziałowcach każdej firmy. Każdy z udziałowców jest posiadaczem pewnego ułamka danej firmy. Udziałowcem może być osoba fizyczna, firma, gmina lub skarbowo państwa.

3. Mamy tabele:

```
Abonament( Nazwa, Cena)
Osoby( PESEL, Imie, Nazwisko, Miasto)
Telefony( Numer, PESEL, Abonament)
```

Pola podkreślone są kluczami. Abonament w tabeli telefony test nazwa abonamentu, a PESEL w tej samej tabeli jest numerem PESEL jakiejś osoby. Żadne pole nie przyjmuje wartości NULL.

Napisz zapytanie SQL odpowiadające na pytanie: ile jest miast, w których najpopularniejszy (najczęściej płacony) rodzaj abonamentu jest inny niż najpopularniejszy abonament w całej Polsce (w przypadku gdy najpopularniejszych abonamentów płaconych w całej Polsce jest kilka, najpopularniejszy abonament w zliczanych miastach nie może być żadnym z nich). Wolno używać widoków.

4. Schemat danych taki jak w zadaniu 3. Dane jest zapytanie:

```
SELECT PESEL
FROM Telefony, Osoby
WHERE Telefony.PESEL = Osoby.PESEL
AND Miasto BETWEEN 'A' AND 'CZZZZ'
AND Abonament = 'Standardowy';
```

Wskazać struktury pomocnicze, które najlepiej wspierają wykonanie tego zapytania. Podać najlepszy plan wykonania zapytania działający przy założeniu istnienia tych struktur. Oszacować jego koszt w terminach operacji wejścia-wyjścia.

5. Schemat danych taki jak w zadaniu 3. Napisać w algebrze relacji następujące zapytanie:

```
SELECT o.PESEL klient
FROM Osoby o
WHERE NOT EXISTS (SELECT 6+7 FROM Telefony WHERE o.PESEL = PESEL);
```