



# Zestaw 1



## Egzamin z Baz Danych 2022.

### Punktacja

Do zdobycia jest 50 punktów, progi to: 26 — dst, 30 — dst+, 34 — db, 38 — db+, 42 — bdb.

Punkty za zadanie są przyznawane wyłącznie, jeśli student zdobędzie **przynajmniej 25% liczby punktów przypisanej danemu zadaniu** (tzn. jeśli rozwiązanie zadania zostanie ocenione na mniej niż 25% punktów możliwych do zdobycia za nie, wynik ten jest zaokrąglany do zera).

### Opis bazy danych (do wszystkich zadań)

Rozważmy następującą bazę danych *Magazyn*, która zawiera dane o firmach, osobach je reprezentujących, produktach, zamówieniach oraz o poszczególnych pozycjach zamówień:

- osoba(id int, imie text, nazwisko text, firma\_id int)
- firma(id int, nazwa text, dane text)
- produkt(id int, nazwa text, cena\_szt decimal, stan int, dostawca text, dane\_dostawcy text)
- pozycja(id int, zamowienie\_id int, produkt\_id int, liczba\_sztuk int)
- zamowienie(id int, osoba\_id int, czas timestamp, adres text)

Dla wszystkich relacji atrybuty id są ich kluczami głównymi (*primary key*). Atrybuty produkt\_id, zamowienie\_id, osoba\_id, firma\_id są kluczami obcymi wskazującymi na atrybuty id relacji produkt, zamowienie, osoba, firma (odpowiednio).

Wszystkie atrybuty za wyjątkiem firma\_id i dane\_dostawcy są oznaczone jako not null. Atrybuty imie i nazwisko to dane osoby reprezentującej firmę. Może się zdarzyć, że więcej niż jedna osoba reprezentuje tę samą firmę. Atrybut czas to czas złożenia zamówienia.

W bazie zachodzą następujące więzy:

1. Produkty o tej samej nazwie zawsze kosztują tyle samo za sztukę niezależnie od ich dostawcy.
2. Dla każdej pary nazwa, dostawca w relacji produkt jest co najwyżej jeden taki produkt w bazie.
3. Atrybut dane\_dostawcy zawiera dane kontaktowe dostawcy dostawca (takie same dla każdego produktu konkretnego dostawcy).
4. Stan magazynowy produktu (stan) jest liczbą nieujemną.

Imię i nazwisko, indeks: ..... 1

**Zadanie 1.** (8 punktów) Napisz zapytanie w algebrze relacji, które zwróci nazwy produktów, które nie zostały ani razu zamówione po 1 stycznia 2022 roku (do porównywania dat można używać standardowych operatorów  $<$ ,  $\leq$ ,  $=$ ,  $\geq$ ,  $>$ ). Uwzględnij, że może zdarzyć się, że różne produkty o takiej samej nazwie dostarczają różni dostawcy i napisz zapytania w dwóch wersjach:

- Nazwa produktu, który nie został zamówiony po 1 stycznia 2022 powinna zostać wypisana, nawet jeśli po tej dacie zamówiono produkt o tej samej nazwie u innego dostawcy.

- Nazwa produktu nie powinna zostać wypisana, jeśli po 1 stycznia 2022 zamówiono produkt o tej samej nazwie, ale u innego dostawcy.

Imię i nazwisko, indeks: ..... 1

**Zadanie 2.** (8 punktów) Napisz zapytanie SQL, które wyliczy sumaryczną wartość zamówień dla każdej firmy. Wynikiem powinny być pary składające się z nazwy firmy oraz obliczonej kwoty. Uwzględnij również firmy, które nie składały zamówień.

Imię i nazwisko, indeks: ..... 1

**Zadanie 3.** (8 punktów) Powiemy, że pewne dwa produkty są *powiązane bezpośrednio* jeśli występują razem jako pozycje pewnego zamówienia. Produkty  $a$  i  $b$  są *powiązane*, jeśli istnieje produkt  $c$  taki, że  $a$  i  $c$  są powiązane bezpośrednio, a  $c$  i  $b$  są powiązane. Napisz zapytanie Datalogu definiujące relację **Powiązane** zawierającą pary powiązanych produktów. Następnie wykorzystaj tę relację w zapytaniu algebry relacji, które zwraca `id` osób, które zamawiają jedynie powiązane produkty (tj. nigdy nie zamówiły dwóch produktów, które nie są ze sobą powiązane).

*Jeśli wolisz, zamiast Datalogu i algebry relacji możesz użyć SQL, ale wtedy całe rozwiązanie zadania musi być jednym zapytaniem SQL.*

**Zadanie 4.** (4 punktów) Napisz zapytanie SQL zwracające listę nazw tych firm, dla których baza nie zawiera informacji o osobach je reprezentujących.

Imię i nazwisko, indeks: ..... 1

**Zadanie 5.** (8 punktów) W jaki sposób najlepiej wymusić spełnianie czterech wymienionych w opisie bazy więzów? Zaproponuj odpowiednie zmiany (np. opisz jak zreorganizować schemat albo jakie nowe więzy zdefiniować za pomocą SQL) wraz z uzasadnieniem ich poprawności.



Imię i nazwisko, indeks: ..... 1

**Zadanie 6.** (8 punktów) Używamy systemu z kontrolą współbieżności MVCC, tj. taką, jak w systemie PostgreSQL. Rozważmy następujące zapytanie, w którym :d oraz :n są parametrami zadanymi przez aplikację i oznaczają nazwę i dostawcę produktu (takie pary są unikalne w relacji produkt).

```
BEGIN TRANSACTION ISOLATION LEVEL REPEATABLE READ;
UPDATE produkt SET stan = stan - 1
    WHERE stan > 0 AND nazwa = :n AND dostawca = :d;

SELECT sum(stan) INTO liczba FROM produkt WHERE nazwa = :n;

IF (liczba=0) THEN ROLLBACK ELSE COMMIT;
```

Zapytanie powinno zmniejszać stan magazynowy podanego produktu o 1 tylko wówczas, gdy nie powoduje to wyzerowania sumy stanów produktów o podanej nazwie :n. Czy faktycznie to rozwiązanie daje taką gwarancję? Odpowiedź uzasadnij lub podaj kontrprzykład. (5 pkt.)

Czy Twoja odpowiedź zmieni się, jeśli w zapytaniu ustawiany będzie poziom izolacji READ COMMITTED? Uzasadnij. (3 pkt.)

Imię i nazwisko, indeks: ..... 1

**Zadanie 7.** (6 punktów) Aby przyspieszyć wykonywanie zapytań, w których szukane są produkty według nazwy lub dostawcy, założono indeks następującym poleceniem:

```
CREATE INDEX idx ON produkt (nazwa, dostawca);
```

Dla poniższych zapytań krótko opisz, czy należy się spodziewać ich przyspieszenia w wyniku wykorzystania tego indeksu. Podaj zwięzłe uzasadnienie. Tak jak poprzednio :d oraz :n są parametrami zadanymi przez aplikację i oznaczają nazwę i dostawcę produktu.

a) `SELECT id from produkt WHERE nazwa = :n`

b) `SELECT id from produkt WHERE dostawca = :d`

c) `SELECT id from produkt WHERE nazwa = :n AND dostawca = :d`