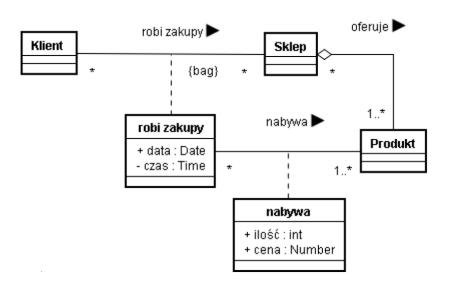
Sprawozdanie z laboratorium Hurtownie Danych

Mikołaj Kubś, 272662 8 marca 2025

1 Zadanie 1

Analiza konceptualnego modelu danych "Usługi", który jest niekompletny, ale klasy i relacje między nimi reprezentują rozpatrywany wycinek rzeczywistości.



Rysunek 1: Konceptualny model danych "Usługi"

Reguły i ograniczenia dziedzinowe:

- Reg/01 Klient może wielokrotnie robić zakupy w tym samym sklepie
- Reg/02 W sklepie może robić zakupy dowolny klient

- \bullet Reg/03 Każdy zakup realizowany jest przez klienta w sklepie w określonym dniu i godzinie
- Reg/04 Sklep musi oferować co najmniej jeden produkt
- Reg/05 ...

1.1 Weryfikacja i poprawa modelu danych

Ponieważ reguły są niekompletne i nie w pełni poprawne, zdecydowałem się wprowadzić szereg zmian.

Uznałem, że reguła "Reg/04 - Sklep musi oferować co najmniej jeden produkt" wprowadza niepotrzebną komplikację. Na przykład, według tej zasady, gdy sklep sprzedałby cały swój inwentarz, nie mógłby dalej istnieć w bazie danych.

Brakuje aktualnie informacji, jaka jest liczba dostępnego produktu w danym sklepie. Można wyciągnąć tą daną do nowej tabeli asocjacyjnej, do której można by dodatkowo dodać cenę produktu dla konkretnego sklepu, co zwiększa elastyczność na przyszłość i jest szeroko stosowaną praktyką w rzeczywistych sklepach. Tak więc sklep może oferować wiele produktów, każdy z własną cenę i ilością. Ponieważ tabela asocjacyjna "oferuje" ma cenę, można by usunąć cenę z tabeli asocjacyjnej "nabywa". Uznałem jednak, że ją zostawię, ponieważ cena oferty może się zmienić po zakupie produktu przez klienta. Sklepy mają często informację, że coś sprzedają, nawet jeśli nie ma tego chwilowo w magazynie, tak więc ustaliłem, że ilość oferowanego produktu to co najmniej 0, a nie koniecznie więcej niż zero, co wymuszałoby usunięcie oferty w przypadku braków w magazynie.

Można dodać parę atrybutów do encji. Do klienta dodam imię i nazwisko, a do produktu i do sklepu nazwę.

Warto również dodać standardowe ograniczenia wobec atrybutów encji. Zdecydowałem, że cena nie może być mniejsza od 0, ale może się mu równać - czasem są zaskakujące promocje w sklepach. Dodatkowo uznałem, że każdy produkt ma unikalną nazwę.

Oprócz tego dodałem reguły wynikające z diagramu klas, klaryfikujące, że:

- 1. Klient może robić zakupy w różnych sklepach
- 2. Ten sam produkt może być oferowany w wielu sklepach
- 3. Klient robiący zakupy musi zakupić przynajmniej 1 produkt

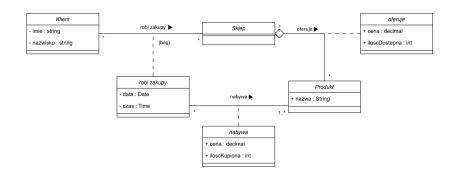
Dodatkowo zmieniłem typ atrybutów dotyczących kosztu na "decimal", a także zmieniłem atrybut "data" w "robi zakupy" na prywatny.

Na koniec dodałem regułę klaryfikującą, że wiele zakupów klienta nie może być zapisane do bazy danych w tym samym czasie.

1.2 Finalna postać reguł, ograniczeń i diagramu klas UML

Finalna postać reguł i ograniczeń:

- Reg/01 Klient może wielokrotnie robić zakupy w tym samym sklepie
- $\bullet~{\rm Reg}/02$ W sklepie może robić zakupy dowolny klient
- Reg/03 Każdy zakup realizowany jest przez klienta w sklepie w określonym dniu i godzinie
- \bullet Reg/04 Każdy sklep ustala własną cenę oraz ilość oferowanego produktu
- Reg/05 Klient nabywając produkt w danym sklepie kupuje go za cenę oferowaną w sklepie, która zostaje zapamiętana
- Reg/06 Sklep może zmienić cenę oferowanego produktu, co wpływa tylko na przyszłe zakupy
- Reg/07 Klient może robić zakupy w różnych sklepach
- $\bullet~{\rm Reg}/08$ Ten sam produkt może być oferowany w wielu sklepach
- Reg/09 Klient robiąc zakupy musi nabyć co najmniej 1 produkt
- Reg/10 Imię klienta nie może być puste
- Reg/11 Nazwisko klienta nie może być puste
- Reg/12 Ilosc oferowanego produktu nie może być mniejsza od zera
- Reg/13 Cena oferowanego produktu nie może być mniejsza od zera
- Reg/14 Ilość nabytego produktu musi być większa od zera
- Reg/15 Cena nabytego produktu nie może być mniejsza od zera
- Reg/16 Nazwa produktu nie może się powtarzać
- Reg/17 Klient nie może robić zakupów więcej niż raz o tym samym czasie i dacie.



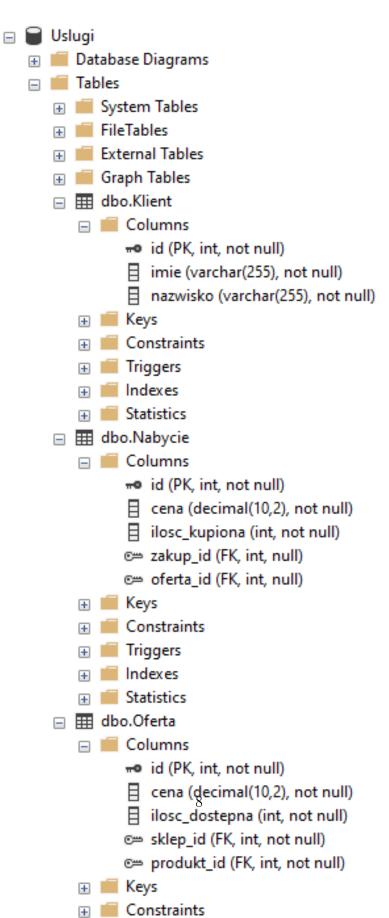
Rysunek 2: Finalny model danych "Usługi"

1.3 Skrypt DDL SQL

```
1
    CREATE DATABASE Uslugi;
2
    GO
3
    USE Uslugi;
5
6
    DROP TABLE IF EXISTS Nabycie;
7
    DROP TABLE IF EXISTS Zakup;
    DROP TABLE IF EXISTS Oferta;
9
    DROP TABLE IF EXISTS Produkt;
10
    DROP TABLE IF EXISTS Sklep;
    DROP TABLE IF EXISTS Klient;
12
13
    CREATE TABLE Klient (
14
        id INT PRIMARY KEY,
15
        imie VARCHAR(255) NOT NULL,
16
        nazwisko VARCHAR(255) NOT NULL
17
    );
18
19
    CREATE TABLE Sklep (
20
        id INT PRIMARY KEY,
21
        nazwa VARCHAR(255) NOT NULL
22
    );
23
^{24}
    CREATE TABLE Produkt (
25
        id INT PRIMARY KEY,
26
        nazwa VARCHAR(255) NOT NULL UNIQUE
```

```
);
28
29
    CREATE TABLE Oferta (
30
        id INT PRIMARY KEY,
31
        cena DECIMAL(10,2) NOT NULL CHECK (cena >= 0),
        ilosc_dostepna INT NOT NULL CHECK (ilosc_dostepna >= 0),
        sklep_id INT NOT NULL,
34
        produkt_id INT NOT NULL,
35
        FOREIGN KEY (sklep_id) REFERENCES Sklep(id) ON DELETE CASCADE,
36
        FOREIGN KEY (produkt_id) REFERENCES Produkt(id) ON DELETE CASCADE
37
    );
38
    CREATE TABLE Zakup (
40
        id INT PRIMARY KEY,
41
        data DATE NOT NULL,
42
        czas TIME NOT NULL,
43
        klient_id INT,
44
        sklep_id INT NOT NULL,
        FOREIGN KEY (klient_id) REFERENCES Klient(id) ON DELETE SET NULL,
46
        FOREIGN KEY (sklep_id) REFERENCES Sklep(id) ON DELETE CASCADE,
47
        CONSTRAINT unique_zakup UNIQUE (data, czas, klient_id, sklep_id)
48
    );
49
50
    CREATE TABLE Nabycie (
51
        id INT PRIMARY KEY,
52
        cena DECIMAL(10,2) NOT NULL CHECK (cena >= 0),
53
        ilosc_kupiona INT NOT NULL CHECK (ilosc_kupiona > 0),
54
        zakup_id INT,
55
        oferta_id INT,
56
        FOREIGN KEY (zakup_id) REFERENCES Zakup(id) ON DELETE SET NULL,
57
        FOREIGN KEY (oferta_id) REFERENCES Oferta(id)
    );
```

1.4 Inicjalizacja bazy danych w systemie MS SQL



1.5 Testy działania bazy danych

```
(4 rows affected)
(5 rows affected)
(6 rows affected)
(6 rows affected)
(6 rows affected)
(8 rows affected)
(9 rows affected)
(9 rows affected)
(10 rows af
```

Rysunek 4: Widok konsoli po wykonaniu poniższego kodu

```
INSERT INTO Klient (id, imie, nazwisko) VALUES
  (1, 'Jan', 'Kowalski'),
  (2, 'Anna', 'Nowak'),
  (3, 'Piotr', 'Wójcik');
  INSERT INTO Sklep (id, nazwa) VALUES
  (1, 'Biedronka'),
  (2, 'Biedronka'),
  (3, 'BIEDRONKA<sub>□</sub>pl.<sub>□</sub>GRUNWALDZKI');
10
11
  INSERT INTO Produkt (id, nazwa) VALUES
12
  (1, 'Masło'),
13
  (2, 'Bułki'),
   (3, 'Wedliny');
15
16
  INSERT INTO Oferta (id, cena, ilosc_dostepna, sklep_id, produkt_id) VALUES
17
  (1, 25.99, 10, 1, 1),
18
  (2, 45.50, 5, 1, 2),
19
  (3, 15.75, 15, 2, 1),
20
  (4, 35.30, 7, 2, 3),
^{21}
  (5, 99.99, 0, 3, 2);
22
  INSERT INTO Zakup (id, data, czas, klient_id, sklep_id) VALUES
```

```
(2, '2025-03-07', '11:00:00', 2, 2),
26
   (3, '2025-03-07', '12:00:00', 3, 3);
27
28
  INSERT INTO Nabycie (id, cena, ilosc_kupiona, zakup_id, oferta_id) VALUES
  (1, 25.99, 3, 1, 1),
30
   (2, 45.50, 2, 2, 2),
31
  (3, 15.75, 5, 3, 3),
32
   (4, 35.30, 1, 3, 4);
33
34
   -- Invalid operations
   INSERT INTO Oferta (id, cena, ilosc_dostepna, sklep_id, produkt_id) VALUES (6, -
37
38
   INSERT INTO Oferta (id, cena, ilosc_dostepna, sklep_id, produkt_id) VALUES (7, 2
39
40
  INSERT INTO Nabycie (id, cena, ilosc_kupiona, zakup_id, oferta_id) VALUES (5, 25
41
   INSERT INTO Nabycie (id, cena, ilosc_kupiona, zakup_id, oferta_id) VALUES (5, -2
43
44
  INSERT INTO Zakup (id, data, czas, klient_id, sklep_id) VALUES (4, '2025-03-07',
      Wszystkie operacje zadziałały, oprócz tych, które naruszały ograniczenia.
  Błędy w konsoli jasno opisują, które konkretnie ograniczenie zostało złamane.
```

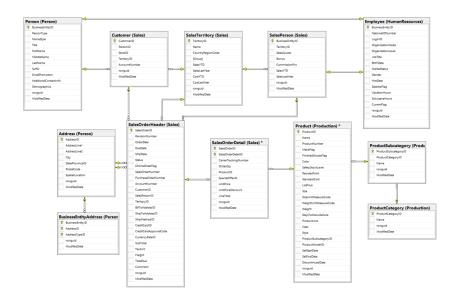
2 Zadanie 2

2.1 Opis zadania

(1, '2025-03-07', '10:30:00', 1, 1),

Do wykonania zadania użyłem bazy danych "AdventureWorks2014" (https://github.com/Microsoft/sql-server-samples/releases/download/adventureworks/AdventureWorks2014.bak)

Tymi kwerendami sprawdzone zostały wszystkie ograniczenia.



Rysunek 5: Wykorzystane w rozwiązaniu tabele

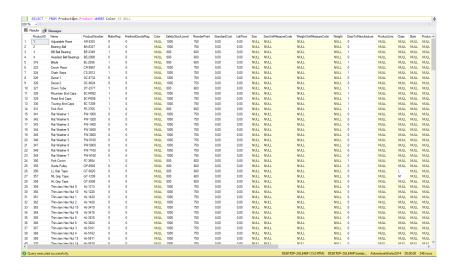
Kwerendy do wykonania:

- 1. Ile jest produktów w bazie? Ile kategorii i podkategorii?
- 2. Wypisz produkty, które nie mają zdefiniowanego koloru.
- 3. Podaj roczną kwotę transakcji (SalesOrderHeader.TotalDue) w poszczególnych latach.
- 4. Ilu jest klientów, a ilu sprzedawców w sklepie? Ilu w poszczególnych regionach?
- 5. Ile było wykonanych transakcji w poszczególnych latach?
- 6. Podaj produkty, które nie zostały kupione przez żadnego klienta. Zestawienie pogrupuj według kategorii i podkategorii.
- 7. Oblicz minimalną i maksymalną kwotę rabatu udzielonego na produkty w poszczególnych podkategoriach.
- 8. Podaj produkty, których cena jest wyższa od średniej ceny produktów w sklepie.
- 9. Ile średnio produktów w każdej kategorii sprzedaje się w poszczególnych miesiącach?
- 10. Ile średnio czasu klient czeka na dostawę zamówionych produktów? Przygotuj zestawienie w zależności od kodu regionu (Sales Territory. Country Region Code).

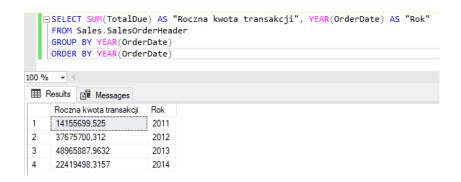
2.2 Kwerendy



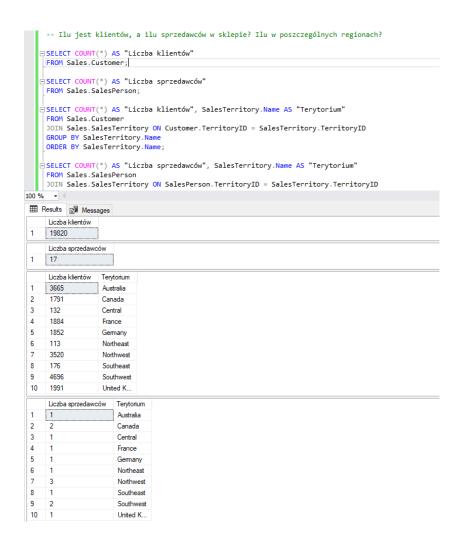
Rysunek 6: Wynik kwerendy 1



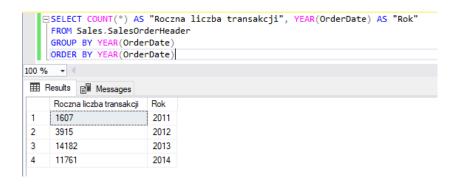
Rysunek 7: Wynik kwerendy 2



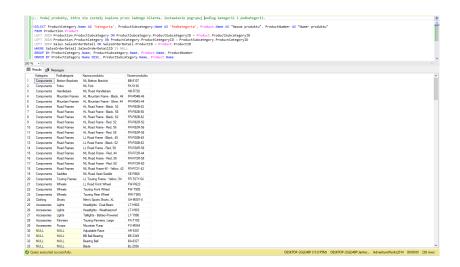
Rysunek 8: Wynik kwerendy 3



Rysunek 9: Wynik kwerendy 4



Rysunek 10: Wynik kwerendy 5



Rysunek 11: Wynik kwerendy 6

```
SEELECT Productionalisms is nately maked by the control of the con
```

Rysunek 12: Wynik kwerendy 7



Rysunek 13: Wynik kwerendy 8

```
The Freshin produktion w kalley battegerii sprzedeje się w pozaczególnych wiesiącach?

(SOUTH Pothbylogie SG (Produktickagory, mane SG Estagory, SUDMICEALES, Salachderrederic Ordenter), SS Paulos, SUMICEALES, Salachderrederial S
```

Rysunek 14: Wynik kwerendy 9



Rysunek 15: Wynik kwerendy 10

3 Wnioski

3.1 Wnioski z zadania 1

Zadanie okazało się znacznie bardziej skomplikowane, niż mogło by się wydawać na pierwszy rzut oka. Im więcej czasu spędzi się nad takim z pozoru prostym diagramem, tym więcej można dodać reguł do uściślenia zasad. Bez jasnych wytycznych klienta można też było przyjąć zupełnie inne założenia wobec bazy danych. Aktualna wersja bazy danych jest dość abstrakcyjna -

nie jest jasne, czy dotyczy tylko sklepów internetowych, czy stacjonarnych, czy obu naraz. Moim zamysłem było nie ograniczać za bardzo możliwości, tak więc aktualną tabelę można by wykorzystać do obu rodzajów sklepów.

W przykładowych danych, którymi zasiałem bazę, widać już pewne usprawnienie - prawdopodobnie w przyszłości można by wyciągnąć kategorię sklepu do osobnej tabeli, aby, na przykład, mieć pewność, że wszystkie sklepy konkretnej marki będą łatwo dostępne. Wtedy każdy z tych sklepów byłby np. kategorii "Biedronka". Można by przekształcić też kategorię w kierunku opisu innych właściwości sklepu, np. czy jest internetowy, czy stacjonarny.

Podsumowując, liczba reguł znacznie wzrosła, ale bez jasnych wytycznych klienta warto ograniczyć się w ich ilości, żeby zezwalać na więcej możliwości w przyszłości. Oczywiście, jedyną pewną rzeczą w oprogramowaniu jest ciągła zmiana, ale nowoczesne narzędzia bazodanowe są w stanie bardzo wspomóc migracje do nowych wytycznych. Tak więc nie trzeba przesadnie "bać" się różnych pomysłów w początkowej fazie projektowania, byleby je dobrze i ciągle konsultować z klientem.

3.2 Wnioski z zadania 2

Pewne jest to, że te 10 kwerend to było zdecydowanie za mało, by dowiedzieć się wszystkiego o stanie omawianej firmy. To też za mało, żeby próbować zrozumieć czynniki wpływające na taką, a nie inną, sytuację. Ale mimo tego, można i tak wywnioskować ciekawe informacje.

Oferta sklepu wydaje się dość różnorodna i logicznie podzielona na kategorie i podkategorie. Sklep z każdym rokiem wydaje się zarabiać coraz więcej (mogłoby się wydawać, że w roku 2014 jest spadek, ale zapisy w bazie danych kończą się w połowie roku).

Liczba klientów jest dość spora, zwłaszcza, jak na liczbę sprzedawców. Korelacja między liczbą klientów a liczbą sprzedawców w danym regionie wydaje się istnieć (więcej klientów -> więcej sprzedawców). Wyjątkiem jest Kanada - w innych regionach jest tylko 1 sprzedawca dla tylu klientów. Rozkład klientów na pewno nie jest równomiernie podzielony na regiony - region Southeast i Northeast wydają się radzić sobie z nimi najgorzej (być może z dalszych badań udałoby się odkryć przyczynę).

Liczba transakcji zdecydowanie wzrasta z każdym rokiem. Zarobki wzrastają jednak wolniej, co oznacza, że średnia wartość transakcji spadła w późniejszych latach.

Aż połowa produktów nigdy nie została sprzedana. Być może oferta jest zbyt różnorodna, być może zostały wprowadzone bardzo niedawno - bez dalszych badań nie wiadomo.

Sklep nie zawsze udziela rabatów. Maksymalne wartości rabatów wydają się umiarkowane. Przydałoby się poznać średnią wartość rabatu, bo maksimum i minimum daje nam za mało informacji.

Próg średniej ceny przekracza ponad 1/4 produktów. Wydaje się to być standardową sytuacją.

Zauważalny jest wzrost sprzedaży w miesiącach letnich, co było do przewidzenia dla sklepu rowerowego.

Średni czas oczekiwania klienta w zależności od regionu jest podejrzany. W aż 5 z 6 regionów średni czas wynosi dokładnie 7 dni.

Podsumowując, udało się odkryć część informacji dotyczących sklepu. Bez dalszych badań można jedynie przypuszczać, co mogło być powodem takich, a nie innych, wyników.