

Zasady projektowania hurtowni

Przykład modelu wielowymiarowego dla hurtowni danych NFZ



- 1 -

W poprzednim wykładzie wyjaśniliśmy ogólne zasady budowania schematu gwiazdy i schematu płatka śniegu. Wykład, który teraz rozpoczynamy przedstawia szczegółowe zasady projektowania modelu wielowymiarowego na podstawie schematu relacyjnego.

Wykonanie procedury leczniczej

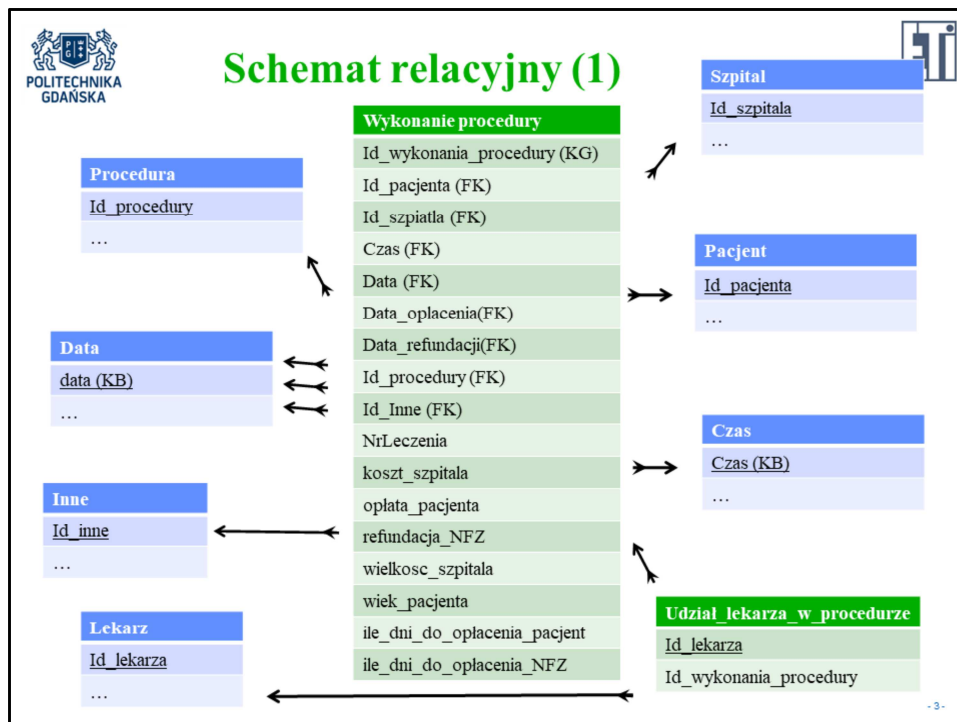


Faktem jest wykonanie **pojedynczej procedury NFZ**, w danym dniu, w danym szpitalu dla danego pacjenta, przez grupę lekarzy. Pojedyncza procedura jest opłacona w danym dniu i w danym dniu zrefundowana przez NFZ.



- 2 -

Przeddefiniujemy fakt wykonania procedury leczniczej raz jeszcze. O jakim typie zdarzenia mówimy?



Dla tak zdefiniowanego faktu przedstawiam schemat relacyjny (jest to schemat gwiazdy i schemat płatka śniegu). W schemacie gwiazdy główną tabelą faktów jest *Wykonanie procedury*. W schemacie płatka śniegu tabelą faktów (jedyną) jest *Udział lekarza_w_procedurze*.

Schemat relacyjny (2)



Procedura

<u>Id_procedury</u>
Nr_procedury_NFZ (KB)
Typ_procedury

Szpital

<u>Id_szpitala</u>
nazwa (KB)
wielkość
województwo
miasto
rodzaj

Data

<u>Data (KB)</u>
rok
miesiąc
dzień
sezon
dzień_pracujący
dzień_tygodnia
święta

Lekarz

<u>Id_lekarza</u>
Nr_zezwolenia (KB)
nazwisko i imię
specjalizacja

Pacjent

<u>Id_pacjenta</u>
PESEL (KB)
nazwisko i imię
pleć
czy_ubezpieczony
zawód
wiek
województwo
miasto

Czas

<u>czas (KB)</u>
pora_dnia
godzina

- 4 -

Na slajdzie przedstawione są tabele wymiarów ze wszystkimi atrybutami. Ze względu na fakt, że schemat relacyjny definiuje dwa schematy tworzymy dwie kostki.

Model wielowymiarowy – fakt i miary



Fakt: wykonanie pojedynczej procedury NFZ, w danym dniu, w danym szpitalu dla danego pacjenta, przez grupę lekarzy. Pojedyncza procedura jest opłacona w danym dniu i w danym dniu zrefundowana przez NFZ.



Miary i funkcje agregujące:

Koszt szpitala (koszt_szpitala), SUM

Oplata pacjenta (opлата_pacjenta), SUM

Refundacja NFZ, (refundacja_pacjenta) SUM

Wielkość szpitala (wielkość_szpitala), AVG, MAX, MIN

Wiek pacjenta (wiek_pacjenta), AVG, MAX, MIN

Ile dni do opłacenia przez pacjenta (ile_dni_do_opłacenia_pacjent), AVG, MAX, MIN

Ile dni do opłacenia przez NFZ (ile_dni_do_opłacenia_NFZ), AVG, MAX, MIN

Liczba wykonanych procedur 1, COUNT

Liczba leceń (nr_leczenia), DISTINCT COUNT

Strata = Koszt szpitala – Oplata pacjenta – Refundacja NFZ

- 5 -

Kostka pierwsza odnosi się do faktu głównego (tak go nazwijmy, ze względu na to, że cały proces projektowania został uruchomiony dla tego faktu). Definiujemy miary i funkcje agregujące. Miara odnosi się do wartości z tabeli faktów zapisanej w nawiasie. Po niej są wypisane funkcje agregujące.

Już raz definiowaliśmy miarę jako wartość liczbową generowaną przez fakt. Wiemy ponadto, że z miarą skojarzona jest funkcja agregująca. Jeżeli jednak chcemy, aby z miarą była skojarzona więcej niż jedna funkcja agregująca dochodzimy do pewnej sprzeczności. Mianowicie nie wiadomo, czym miar jest: wartością liczbową generowaną przez fakt, czy parą: wartość liczbową generowaną przez fakt ze skojarzoną funkcją agregującą.

Pojęcie miara jest używane w obu tych znaczeniach. Stąd należy być uważnym. W tym drugim znaczeniu przykładowo **Wielkość szpitala (wielkość_szpitala), AVG, MAX, MIN** jest definicją trzech miar, a nie jednej.

Zauważamy, że występuje również miara: **liczba wykonanych procedur**. Nie odpowiada ona żadnej wartości w tabeli faktów. Dla każdej krotki reprezentującej fakt przyjmuje ona zawsze wartość 1. Zliczając te jedynki możemy policzyć liczbę faktów. Ten typ miary występuje dla każdej tablicy faktów (dla każdego modelu wielowymiarowego).

Warto zwrócić uwagę, że możliwe jest również zbudowanie miary w odniesieniu do nienumerycznej wartości zapisanej w tabeli faktów (przykład: **nr_leczenia**). Takiej mierze najczęściej przypisujemy funkcję agregującą DISTINCT COUNT, która w naszym przykładzie umożliwia zliczenie różnych „leceń”. W tej sytuacji zauważmy, że **nr_leczenia** jest wymiarem (co prawda zdegenerowanym, ale zawsze wymiarem) i miarą jednocześnie.

Kiedyś (pewnie już Państwo nie pamiętają kiedy) obiecałam, że wprowadzę miarę, która nie ma przypisanej funkcji agregującej. Miara taka jest wyliczana na podstawie wartości innych miar i nazywamy ją **miarą wyliczaną**. Dlaczego miara ta nie ma przypisanej funkcji agregującej? Aby to wytłumaczyć posłużymy się przykładem ze slajdu: miarą wyliczaną **Strata**.

Zadajemy pytanie analityczne:

Jaka jest strata w 2016 roku na procedurach medycznych wykonywanych przez lekarzy bez specjalizacji?

Jak wyliczana jest taka strata? Wyliczane są odpowiedzi na trzy zapytania:

Jaki jest koszt szpitala w 2016 roku za wykonanie procedur medycznych przez lekarzy bez specjalizacji? (tutaj zwracany jest koszt zagregowany k)

Jaka jest wniesiona przez pacjentów opłata za wykonanie procedur medycznych wykonanych w 2016 roku przez lekarzy bez specjalizacji? (tutaj zwracana jest zagregowana opłata pacjenta o)

Jaką kwotę zrefundował NFZ za procedury medyczne wykonane w 2016 roku przez lekarzy bez specjalizacji? (tutaj zwracana jest zagregowana kwota refundacji r)

Strata wyliczana jest jako k-o-r.

Czyli funkcje agregujące przypisane są do miar **Koszt szpitala**, **Oplata pacjenta** i **Refundacja NFZ**, a nie do miary wyliczanej **Strata**.

Model wielowymiarowy – wymiary (1)

Szpital:

- Nazwa szpitala (Szpital:nazwa)
- Wielkość (Szpital:wielkosc)
- Województwo (Szpital:wojewodztwo)
- Miasto (Szpital:miasto)
- Rodzaj (Szpital: rodzaj)

Czas wykonania procedury:

- Godzina wykonania procedury (Czas:godzina)
- Pora dnia wykonania procedury (Czas:pora_dnia)

Procedura:

- Nr procedury NFZ (Procedura:Nr_procedury_NFZ)
- Typ procedury (Typ_procedury)

Pacjent:

- PESEL pacjenta(Pacjent:PESEL)
- Nazwisko i imię pacjenta (Pacjent:nazwisko i imię)
- Czy pacjent jest ubezpieczony (Pacjent:czy_ubezpieczony)
- Zawód pacjenta(Pacjent:zawód)
- Wiek pacjenta (Pacjent:wiek)
- Województwo, w którym jest zameldowany pacjent (Pacjent:wojewodztwo)
- Miasto, w którym jest zameldowany pacjent (Pacjent:miasto)
- Płeć pacjenta (Pacjent: płeć)

Na slajdzie widoczna jest definicja wymiarów *Szpital*, *Czas wykonania procedury*, *Procedura* i *Pacjent*. W nawiasach znajdują się pola ze schematu relacyjnego odpowiadające poszczególnym wymiarom.

Model wielowymiarowy – wymiary (2)



Lekarz wykonujący procedurę:

- Nr zezwolenia (Lekarz:Nr_zezwolenia)
- Nazwisko i imię lekarza (Lekarz: nazwisko_i_imię)
- Specjalizacja (Lekarz: specjalizacja)

Data opłacenia procedury przez pacjenta:

- Rok (Data:rok)
- Miesiąc (Data: miesiąc)
- Dzień (Data: dzień)
- Sezon (Data: sezon)
- Czy dzień pracujący (Data: dzień_pracujący)
- Dzień tygodnia (Data: dzień_tygodnia)
- Okres świąteczny (Data: święta)

Data wykonania procedury:

- Rok (Data:rok)
- Miesiąc (Data: miesiąc)
- Dzień (Data: dzień)
- Sezon (Data: sezon)
- Czy dzień pracujący (Data: dzień_pracujący)
- Dzień tygodnia (Data: dzień_tygodnia)
- Okres świąteczny (Data: święta)

- 7 -

Kolejny prezentowany wymiar to lekarz wykonujący procedurę. Pamiętajmy, że jest to wymiar dla tej kostki – pomimo zależności wiele do wiele z główną tabelą faktów.

Na slajdzie są jeszcze dwa wymiary, dla których dane znajdują się w tej samej tabeli tj. tabeli dat. Należy pamiętać, że jedna tabela może generować więcej niż jeden wymiar. Tak jest w tym przypadku. Tabela ta (ze względu na obecność trzech kluczy obcych w tabeli faktów wskazujących na tą samą tabelę) generuje aż trzy wymiary (dwa przedstawione na tym slajdzie: *Data wykonania procedury* i *Data opłacenia procedury przez pacjenta* i jeden ze slajdu następnego tj. *Data opłacenia procedury przez NFZ*).

Model wielowymiarowy – wymiary (3)



Data opłacenia procedury przez NFZ:

- Rok (Data:rok)
- Miesiąc (Data: miesiąc)
- Dzień (Data: dzień)
- Sezon (Data: sezon)
- Czy dzień pracujący (Data: dzień_pracujący)
- Dzień tygodnia (Data: dzień_tygodnia)
- Okres świąteczny (Data: święta)

Data wykonania procedury od roku do dnia:

- Rok
- • Sezon
- • • Miesiąc
- • • • Dzień

Lekarze wg specjalizacji:

- Specjalizacja
- • Nazwisko i imię lekarza

Miejsce zamieszkania pacjenta:

- Województwo, w którym jest zameldowany pacjent
- • Miasto, w którym jest zameldowany pacjent

- 8 -

Zostało zdefiniowanych również sześć wymiarów hierarchicznych, z czego trzy znajdują się na następnym slajdzie.

Model wielowymiarowy – wymiary (4)



Lokalizacja szpitala:

- Województwo
- • Miasto
- • • Nazwa szpitala

Charakterystyka szpitala:

- Wielkość szpitala
- • Rodzaj szpitala
- • • Nazwa szpitala



Charakterystyka szpitala:

- Rodzaj szpitala
- • Wielkość szpitala
- • • Nazwa szpitala

Wymiar hierarchiczny *Charakterystyka szpitala* może być zbudowany w dwojaki sposób. Można również zdefiniować dwa wymiary hierarchiczne opisujące charakterystykę szpitala.

Model wielowymiarowy – fakt, miary i wymiary



Fakt: **Udział danego lekarza w pojedynczej procedurze NFZ.**



Miary i funkcje agregujące:

Ile razy lekarz brał udział w wykonywaniu procedury NFZ 1, COUNT

Wymiary jak dla



Druga kostka jest definiowana dla faktu udział danego lekarza w pojedynczej procedurze NFZ. Jest to fakt, dla którego zdefiniowana jest tylko 1 miara określająca udział lekarza w wykonaniu procedury i przyjmująca wartość 1.

Wymiary są identyczne jak dla kostki 1.

Należy pamiętać, że tej kostce odpowiada schemat relacyjnych będący schematem płatką śniegu.

Co każdy student potrafić powinien...



Zaprojektować model wielowymiarowy dla
poprawnego schematu relacyjnego.

