



KATEDRA
INŻYNIERII
OPROGRAMOWANIA

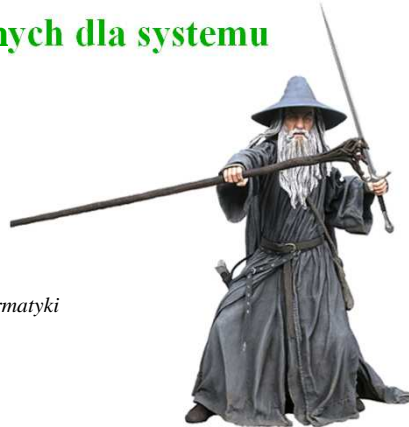
Zasady projektowania hurtowni

Przykład hurtowni danych dla systemu NFZ

Krzysztof Goczyła

Teresa Zawadzka

*Katedra Inżynierii Oprogramowania
Wydział Elektroniki, Telekomunikacji i Informatyki
Politechnika Gdańska
{kris, tegra}@eti.pg.gda.pl*



- 1 -



Szpital - definicja problemu

Zaprojektować hurtownię danych dla Narodowego Funduszu Zdrowia.

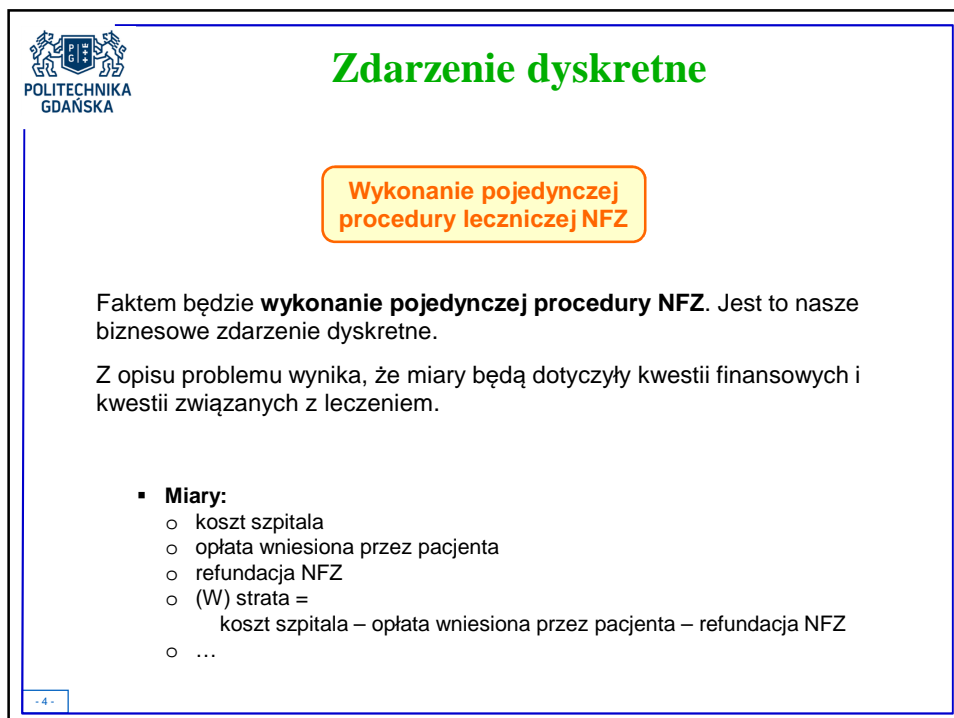
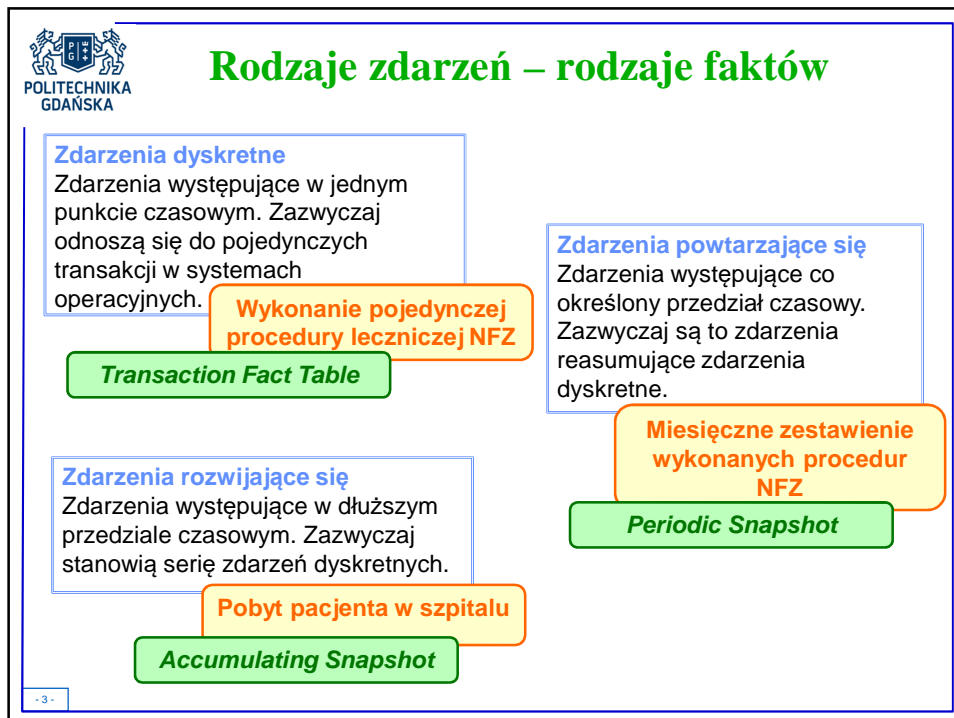
Hurtownia ma umożliwić analizowanie hospitalizacji w polskich szpitalach, które podpisały umowę z NFZ na świadczenie usług medycznych.

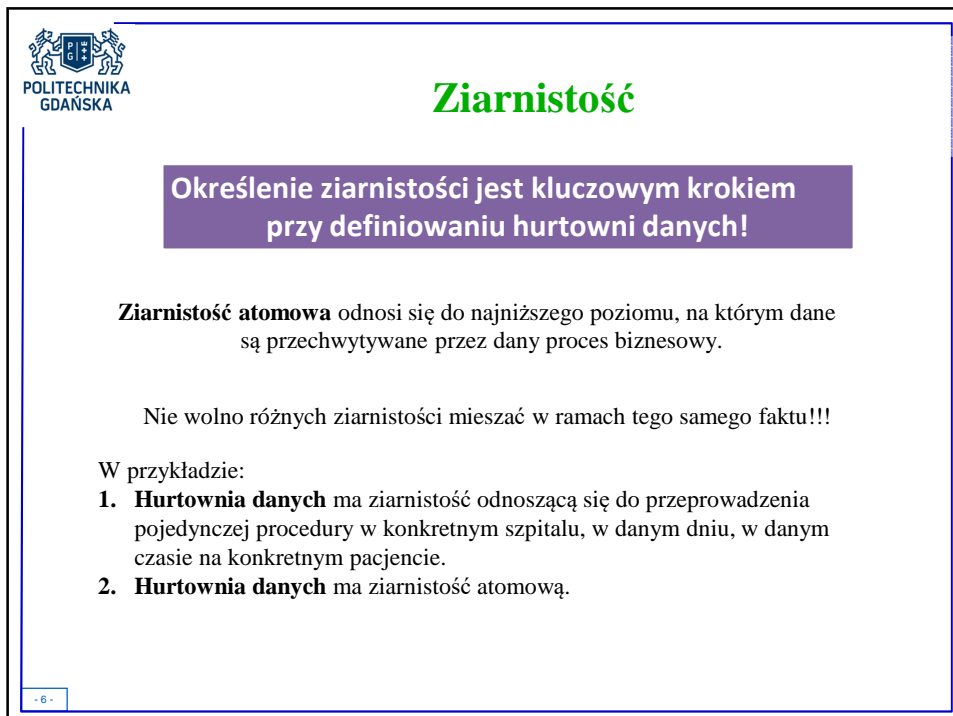
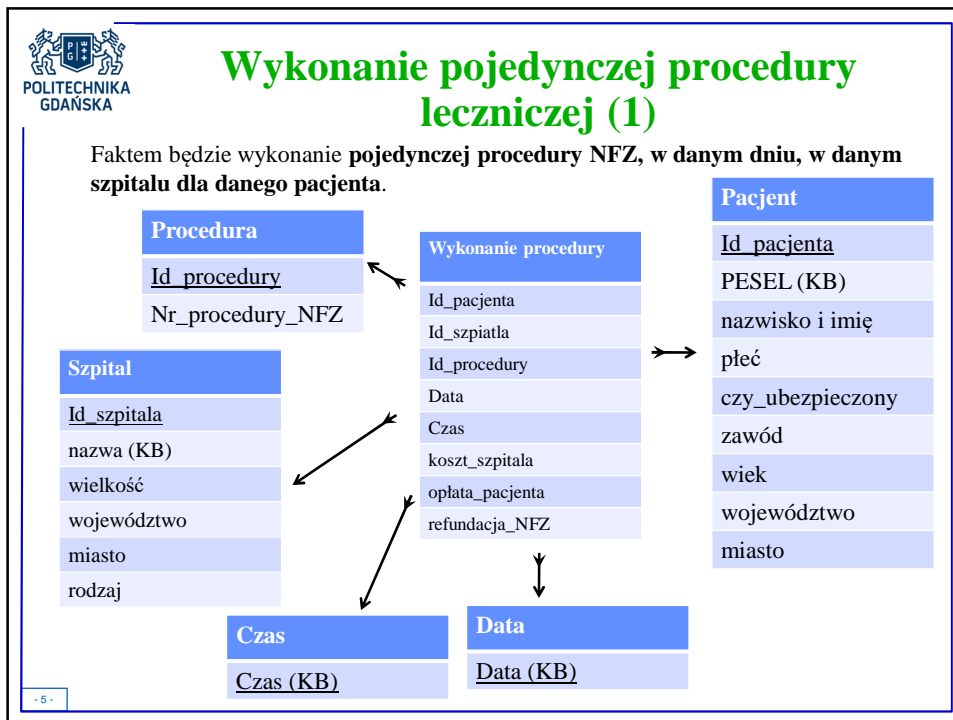
Analizy mają dotyczyć aspektów finansowych oraz zdrowotnych.

Hurtownia ma bazować na danych źródłowych zbieranych przez szpitale oraz na danych zawartych w centralnej bazie NFZ.



- 2 -







Techniki projektowania tabeli faktów (1)

Tabela faktów zawiera ogromną liczbę krotek!
Krotki powinny zajmować mało miejsca!
Każda krotka powinna zawierać tylko wartości numeryczne!

Mamy trzy rodzaje miar:

- Miary **addytywne** (mogą być sumowane po wszystkich wymiarach, w przykładzie **wszystkie miary z przykładu**),
- Miary **pół-addytywne** (po pewnych wymiarach mogą być sumowane, a po pewnych nie, przykład: **stan na rachunku** po wykonaniu transakcji bankowej – nie może być sumowany po czasie, może być sumowany po typie rachunku lub wieku właściciela rachunku),
- Miary **nie-addytywne** (nie mogą być sumowane po żadnych wymiarach, przykład: **zysk procentowy** ze sprzedaży).

-7-



Techniki projektowania tabeli faktów (2)

W tabeli faktów miary **MOGĄ** przyjmować wartości NULL!

Kluczem głównym tabeli faktów jest złożenie kluczy obcych.

W tabeli faktów klucze obce **NIE MOGĄ** przyjmować wartości NULL – naruszenie integralności referencyjnej kluczy.

W przykładzie, gdy nie znamy pacjenta to dodajemy sztuczny wiersz w tabeli wymiaru Pacjent o wartości „Nieznany”.

-8-



Techniki projektowania tabel wymiarów (1)

Tabela wymiarów zawiera liczbę krotek zdecydowanie mniejszą od liczby krotek w tabeli faktów!

Krotki mogą zajmować dużo miejsca!

Każda krotka powinna zawierać poza kluczami tylko wartości opisowe!

W przykładzie, wielkość szpitala i wiek pacjenta nie są wartościami numerycznymi tylko kategoriami – przedziałami liczbowymi określającymi wielkość szpitala i wiek pacjenta.

Co zrobimy jeżeli potrzebujemy odpowiedzi na zapytanie:
Jaka jest średnia wielkość szpitala lub średni wiek pacjenta, który...

- 9 -



Wykonanie pojedynczej procedury leczniczej (2)

Wykonanie procedury
Id_pacjenta
Id_szpitala
data
czas
koszt_szpitala
opłata_pacjenta
refundacja_NFZ
wielkość_szpitala
wiek_pacjenta

- 10 -



Techniki projektowania tabel wymiarów (2)

W każdej tabeli wymiaru znajduje się jeden klucz główny!

Każda tabela wymiaru (ewentualnie poza tabelą daty/czasu) zawiera **klucz główny surogatowy**. Klucz surogatowy jest to klucz generowany od wartości 1, zwiększany o 1.
Przykładowo Id_pacjenta w tabeli wymiaru Pacjent. Klucz główny tabel wymiarów jest przechowywany w tabeli faktów.

Dodatkowo w tabelach wymiaru powinny znajdować się **klucze biznesowe**. Klucze biznesowe nie są kluczami głównymi. Przykładowo PESEL w tabeli Pacjent.

- 11 -



Techniki projektowania tabel wymiarów (3)

Elementy wymiarów mają być wartościami samoopisującymi się.

Nigdy nie kodujemy wartości. W systemach typu BI nie ma dedykowanych pod hurtownię danych aplikacji. Korzysta się z istniejących rozwiązań raportujących.

Kolumna płeć w tabeli wymiaru Pacjent przyjmuje wartości: kobieta lub mężczyzna. Nigdy K lub M albo 0 lub 1.

- 12 -



Wykonanie pojedynczej procedury leczniczej (3)

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

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

- 13 -



Techniki projektowania tabel wymiarów (4)

Dane w tabeli wymiaru Data, Czas zawierają poza datą wartości opisowe opisujące daną datę lub czas.

Dane w tabelach wymiaru Data i Czas są generowane i tabele to powinny być wypełnione przed wykonaniem procesu ETL.

W przykładzie tabela wymiaru Data i Czas.

- 14 -



Wykonanie pojedynczej procedury leczniczej (4)

Jeden pacjent może w szpitalu pojawić się kilkakrotnie. Nie zawsze musi być w tym samym przedziale wiekowym.

Ten sam szpital, nie zawsze będzie miał taką samą wielkość, może się rozbudować...

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

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

- 15 -



Techniki projektowania tabel wymiarów (5)

Wolno zmieniające się wymiary (ang. *Slowly Changing Dimensions*) określają te wymiary które zmieniają się w czasie.

SCD 1 – wartości są nadpisywane, brak historii, analizy mogą być zakłamane.

SCD 2 – nowe krotki jeżeli jakaś wartość ulega zmianie, klucz biznesowy nie zmienia wartości, jest generowany nowy klucz surogatowy. Dodatkowo w tabeli przechowywane są Data wstawienia krotki, Data kiedy krotka straciła aktualność i opcjonalnie pole określające czy krotka jest aktualna, czy też nie.

SCD 3 – obie wartości stara i nowa przechowywane są w tej samej krotce.

W przykładzie SCD 2 jest zaimplementowany w tabeli Szpital i Pacjent.

- 16 -



Wykonanie pojedynczej procedury leczniczej (6)

Dodatkowo:

- procedura medyczna została wykonana w sali o zadanym numerze (nr_sali), sal w szpitalu jest od. 50 – 500 w zależności od szpitala;
- pacjent w czasie wykonania procedury był/lub nie znieczulany (typ_znieczulenie), mamy około 50 różnych typów znieczuleń;
- w czasie wykonywania procedury medycznej była z pacjentem osoba towarzysząca (osoba_towarzysząca);
- procedura wymagała wzywania dodatkowej pomocy (dodatkowa_pomoc);
- ...

Inne
<u>Id_inne</u>
nr_sali
typ_znieczulenia
osoba_towarzysząca
dodakowa_pomoc
...

Liczba możliwych krotek = $500 \times 50 \times 2 \times 2 \times \dots$

- 17 -



Techniki projektowania tabel wymiarów (7)

Tworzymy tabelę wymiarów Inne dla atrybutów wymiarów ze sobą funkcjonalnie nie powiązanych.

Optymalizacja:

- wstawiamy tylko krotki faktycznie występujące, a nie wszystkie możliwe
- w przypadku zbyt dużej liczby krotek w tabeli Inne tworzymy tabelę Inne 1 i Inne 2.

W przykładzie tabela wymiaru Inne.

- 18 -



Wykonanie pojedynczej procedury lecniczej (7)

Procedury medyczne są wykonywane w ramach jednego leczenia/pobytu. Jest nadawany unikalny identyfikator takiego leczenia: [SYMBOL SZPITALA_NR_LECZENIA]. Poza tą informacją nie ma dodatkowych informacji o leczeniu pacjenta (uproszczenie).

Wykonanie procedury
Id_pacjenta
Id_szpitala
data
czas
Id_inne
Nr_leczenia
koszt_szpitala
opłata_pacjenta
wielkość_szpitala
wiek_pacjenta
refundacja_NFZ

- 19 -



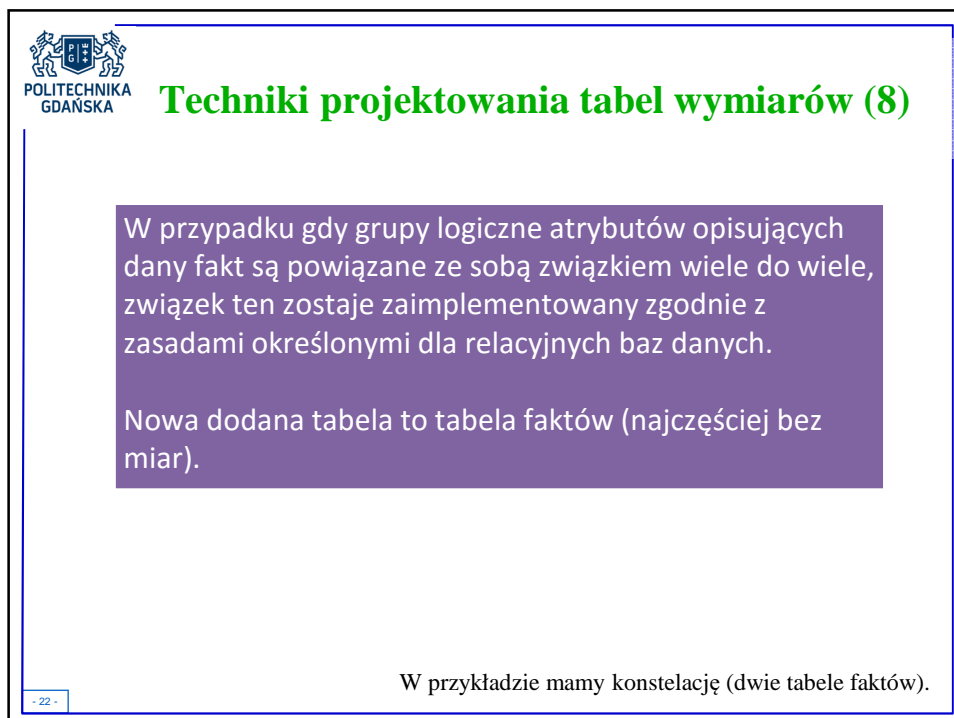
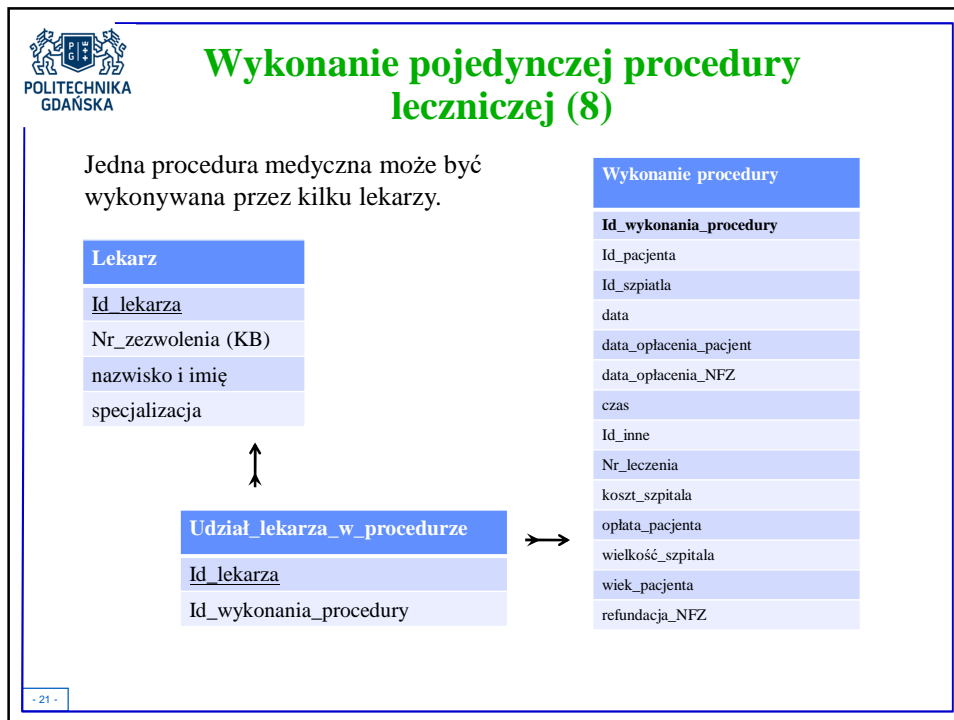
Techniki projektowania tabel wymiarów (7)

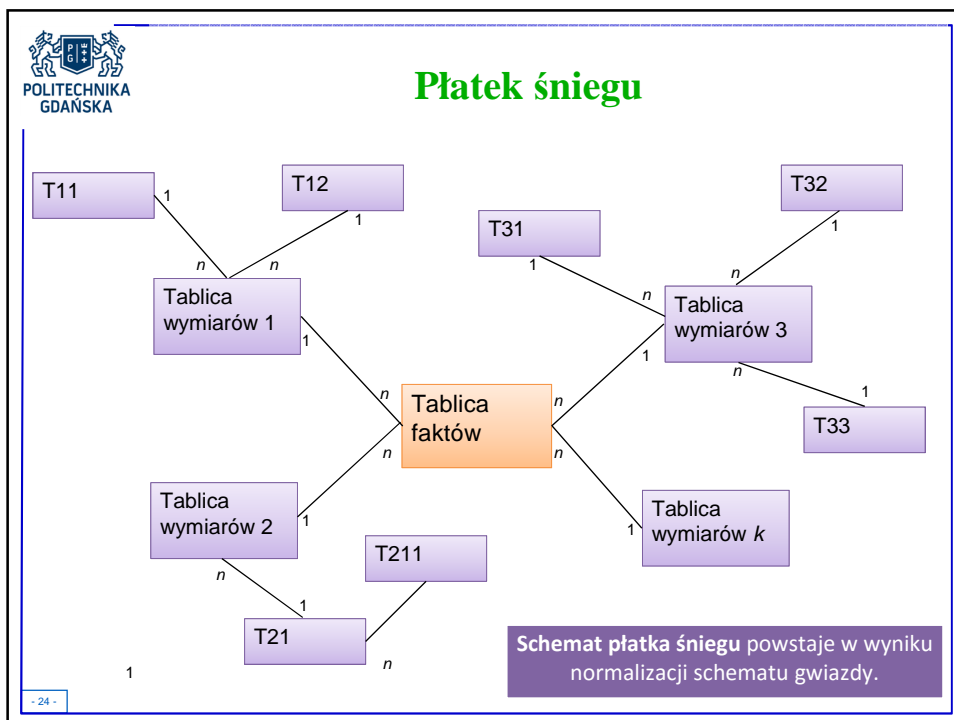
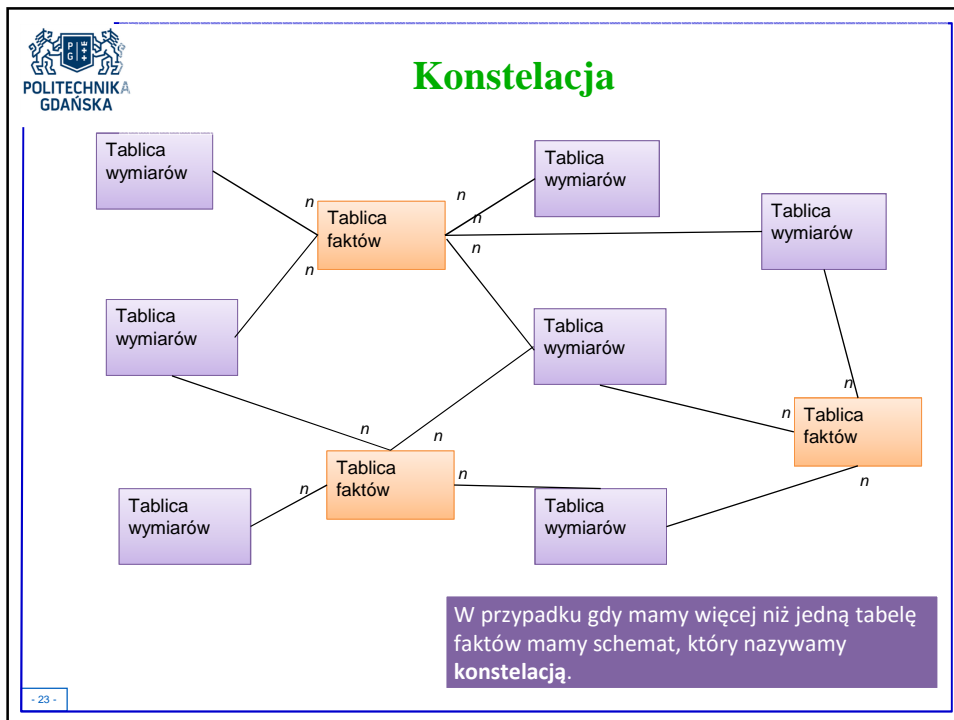
Wymiar zdegenerowany to wymiar, który nie posiada pogrupowanych do tej samej kategorii logicznej innych atrybutów niż klucz główny.

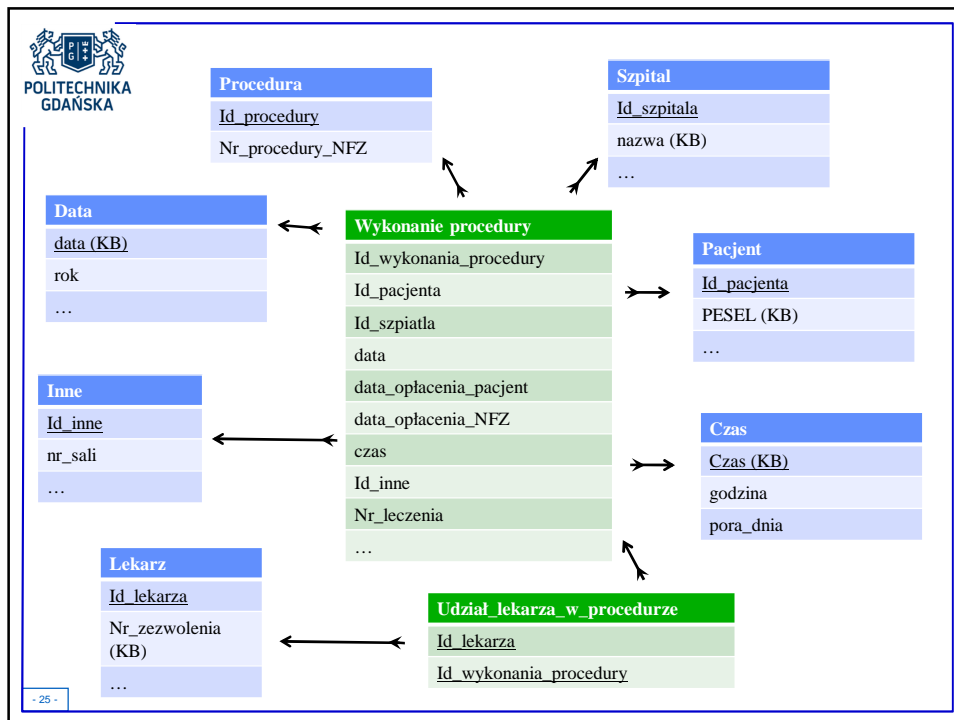
Wymiar zdegenerowany jest umieszczony w tabeli faktów z potwierdzeniem, że nie zawiera skojarzonej z nim tabeli wymiarów.

W przykładzie wymiarem zdegenerowanym jest nr_Leczenia.

- 20 -







POLITECHNIKA GDAŃSKA

Model wielowymiarowy – fakt i miary

Fakt: Wykonanie pojedynczej procedury NFZ, w danym dniu, w danym szpitalu dla danego pacjenta.

Miary i funkcje agregujące:

- Koszt szpitala (koszt_szpitala), SUM
- Oplata pacjenta (opлата_pacjenta), SUM
- Refundacja NFZ, (refundacja_pacjenta) SUM
- Wielkość szpitala (wielkosc_szpitala), AVG, MAX, MIN
- Wiek pacjenta (wiek_pacjenta), AVG, MAX, MIN
- Liczba wykonanych procedur 1, COUNT
- Liczba leceń (nr_leczenia), DISTINCT COUNT
- Strata = Koszt szpitala – Oplata pacjenta – Refundacja NFZ

To jest miara wyliczana!
Nie definiujemy funkcji agregującej!

- 26 -



Model wielowymiarowy – wymiary (1)

Szpital:

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

Czas:

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

Procedura:

- Nr procedury NFZ (Procedura:Nr_procedury_NFZ)

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:województwo)
- Miasto, w którym jest zameldowany pacjent (Pacjent:miasto)

- 27 -



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)

- 28 -

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

- 29 -

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

- 30 -



Model wielowymiarowy – fakt, miary i wymiary

Fakt: Udział danego lekarza w pojedynczej procedurze NFZ.



Miary i funkcje agregujące:

Ile razy lekarz wykonywał procedurę NFZ 1, COUNT

Wymiary jak dla



- 31 -



Co każdy student potrafić powinien...

Zaprojektować hurtownię danych dla zdarzeń dyskretnych.



- 32 -