

# Projekt Hurtowni danych

INŻYNIERIA INTEGRACJI USŁUG INFORMATYCZNYCH  
PRZEMYSŁAW FOLTYN

## Spis treści

1.	Uzasadnienie biznesowe .....	2
2.	Wymagania.....	2
2.1.	Wymagania funkcjonalne .....	2
2.2.	Wymagania pozafunkcjonalne .....	2
3.	Struktura baz danych.....	3
4.	Proces ETL (Extract-Transform-Load).....	5
5.	Hurtownia danych .....	8
6.	Raportowanie .....	9

## 1. Uzasadnienie biznesowe

Hurtownia danych daje bardzo duże możliwości analityczne. Głównym celem stworzenia hurtowni danych mogą być:

- Wykonywanie analiz biznesowych bez ingerencji w systemy transakcyjne
- Wspomaganie decyzji
- Całościowy wgląd w dane firmy
- Dostęp do danych historycznych
- Ujednolicenie posiadanych informacji

W tym projekcie skupimy się na analizach biznesowych. Pozostałe cele mogą zostać wykorzystane w przyszłości. Aplikacja będzie umożliwiać:

- **Analiza trendów i zachowań** - analiza jakie rodzaje leków są najbardziej popularne i jakie należy bardziej promować, ustalić plany sprzedażowe, porównania z tym samym okresem w roku poprzednim.
- **Ukierunkowany marketing** – na podstawie danych można ustalić jakie promocje i na jakie produkty będą prowadzone kampanie reklamowe

## 2. Wymagania

### 2.1. Wymagania funkcjonalne

- *Logowanie*
- *Zarządzanie użytkownikami*
- *Dane sprzedażowe*
- *Produkty*

### 2.2. Wymagania pozafunkcjonalne

Oprogramowanie będzie tworzone na potrzeby kierownictwa hurtowni leków w Warszawie. Dane będą pobierane z baz danych hurtowni na terenie polski.

System będzie dostępny z poziomu przeglądarki internetowej wyłącznie w sieci lokalnej. Musi być kompatybilny z przeglądarka Mozilla Firefox(wersja 18 wzwyż) lub Google Chrome(wersja 30 wzwyż).

Obsługiwanyymi systemami operacyjnymi będą wersje Windows 7 i wyższe z uwagi na infrastrukturę przedsiębiorstwa. Urządzenia mobilne nie będą miały dostępu do aplikacji.

Aplikacja będzie napisana w języku C# .NET 4.0 w technologii ASP.NET MVC 5.0.

Baza danych będzie wykonana w oparciu o MS SQL 2014 Standard.

Serwer bazy danych i serwer aplikacji będzie znajdował się na serwerze lokalnym przedsiębiorstwa.

Dane będą kopiowane na zapasowy serwer znajdujący się poza siedzibą firmy. Kopia bazy danych będzie robiona:

- codziennie przyrostowo
- raz w tygodniu pełna.

Obsługa serwerów zapasowych pozostaje w gestii użytkownika końcowego. Dostawca oprogramowania nie odpowiada za utracone dane w przypadku awarii maszyny fizycznej.

### 3. Struktura baz danych

Do dyspozycji posiadam transakcyjne bazy danych(OLTP) dystrybutorów produktów farmaceutycznych – SuperFarm(SF) oraz Grupy Farmaceutycznej(GF). Dane będą gromadzone w bazie centralnej (OLAP) w siedzibie głównej firmy.

#### 3.1. Schemat bazy Superfarm

Baza danych składa się z tablicy transakcyjnej:

- SF\_Sales – przechowuje identyfikator produktu, regionu, ilość, dzień, miesiąc i rok

Do SF\_Sales podłączone są tablice słownikowe:

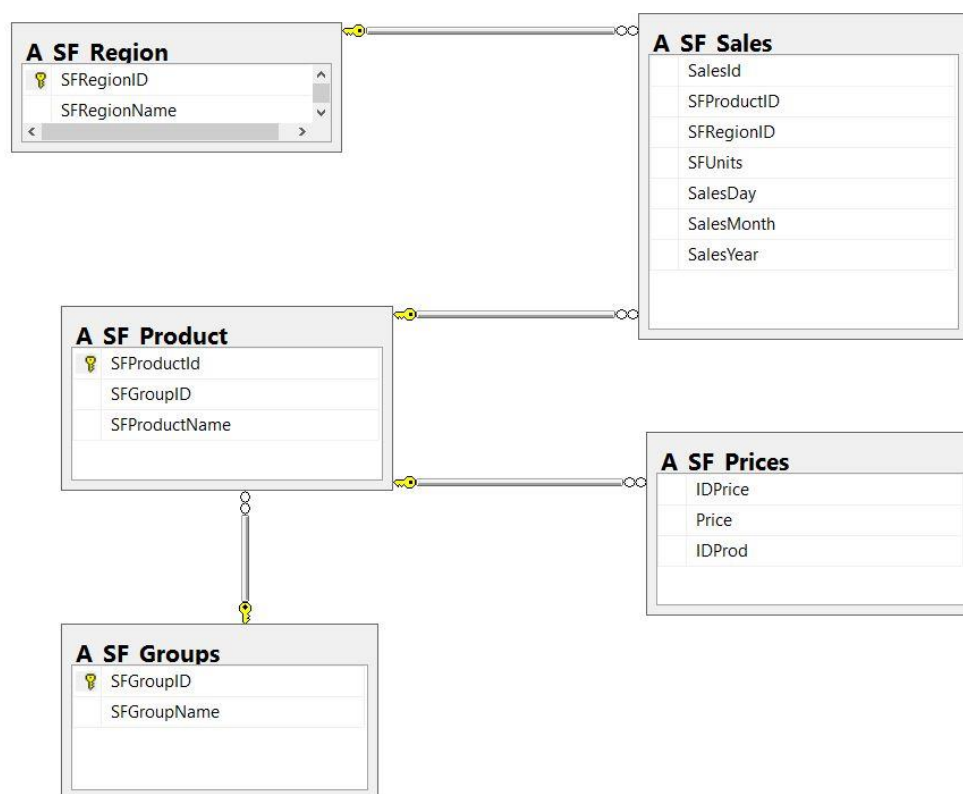
- SF\_Region – tablica posiada 5 regionów na które jest podzielona Polska

- SF\_Product – tablica posiadająca produkty które są podzielone na grupy i nazwy(dawki leku)

Do tablicy SF\_Product podłączone są tablice :

- SF\_Groups – tablica przechowuje nazwy grup produktów

- SF\_Proces – tablica przechowuje ceny produktów

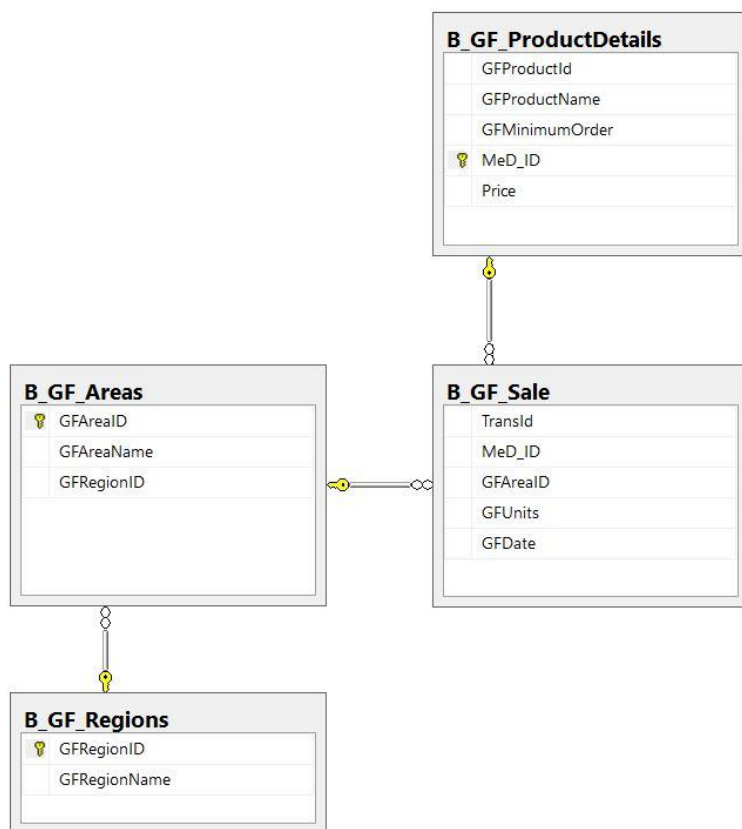


#### 3.2. Schemat bazy Grupy Farmaceutycznej

Baza składa się z tabeli transakcyjnej:

- GF\_Sale – przechowuje identyfikator medyczny leku, identyfikator obszaru, ilość oraz datę transakcji. Do GF\_Sale podłączone są:

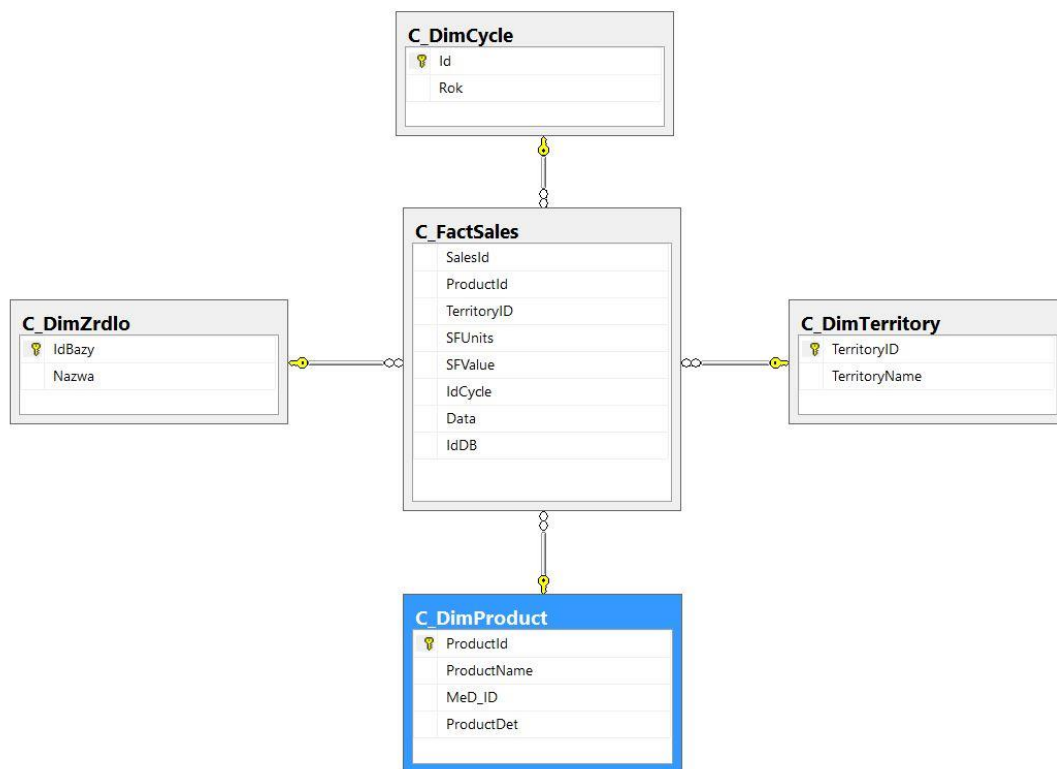
- GF\_ProductDetails – przechowuje Id produktu, nazwę, minimalną ilość zamówienia, identyfikator medyczny oraz cenę.
  - GFAreas – przechowuje identyfikator obszaru, nazwę oraz id regionu.
- Do GF\_Areas podłączona jest tablica:
- GF\_Regions – przechowuje ona identyfikator i nazwę regionu.



### 3.3. Schemat bazy przedsiębiorstwa

Tabelą transakcyjną i tym samym tabelą faktów jest:

- C\_FactSales – przechowuje id produktu, id terytorium, ilość jednostek, wartość zakupu, cykl sprzedażowy, Data sprzedaży oraz id bazy z której pochodzą dane. Do C\_FacSales podłączone są:
- C\_DimCycle – przechowuje id cyklu oraz rok
- C\_DimTerritory – id terytorium oraz nazwę
- C\_DimProduct – przechowuje identyfikator produktu, nazwę, identyfikator medyczny, dawki leku
- C\_DimZrdl – identyfikator źródła oraz nazwa



### 3.4. Różnice między bazami danych

Baza A - transakcja przechowuje ID regionu, datę rozbity na 3 składowe i ilość sprzedanych jednostek

Baza B – transakcja przechowuje ID obszaru, identyfikator medyczny, ilość i datę transakcji.

Baza C – transakcji przechowuje ID produktu, ID Terytorium(Regionu), ilość, wartość, cykl sprzedaży, datę oraz identyfikator źródła.

Dane które trzeba przekształcać tak żeby pasowały do bazy C:

Baza A- Identyfikatory produktów z dawkami produktu, produkt musi pobierać cenę z dodatkowej tabeli oraz składać datę z 3 pól,

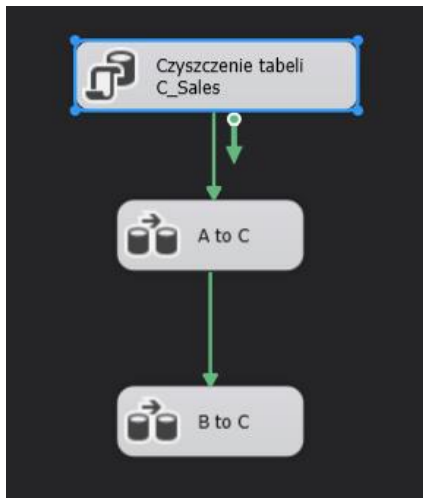
Baza B- Obszary muszą być parowane z regionami i identyfikator medyczny musi być parowany z identyfikatorem produktu.

## 4. Proces ETL (Extract-Transform-Load)

Proces ETL został podzielony na 3 etapy:

1. Czyszczenie tablicy docelowej – C\_FactSales
2. Przesłanie danych z bazy A do bazy C
3. Przesłanie danych z bazy B do bazy C

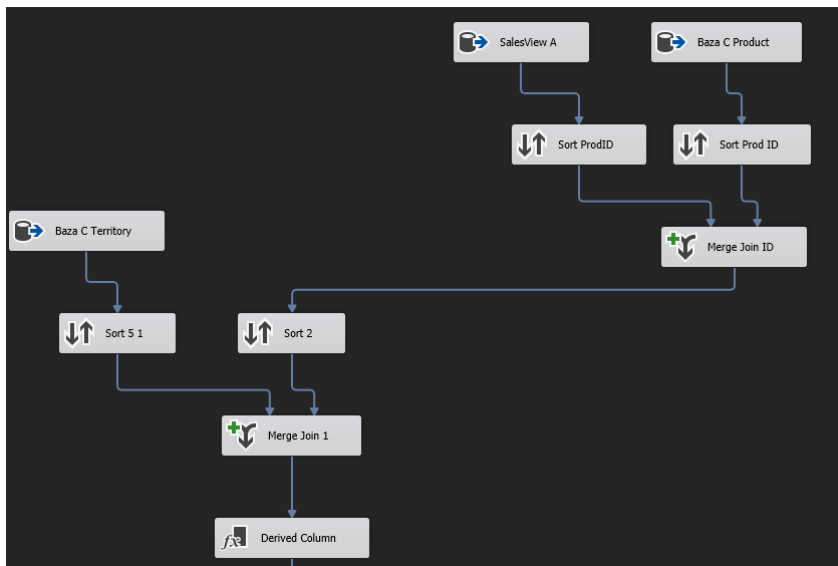
Ad.1 Czyszczenie tabeli jest potrzebne do utrzymania tabeli docelowej w rozmiarze ułatwiającym analizę danych.

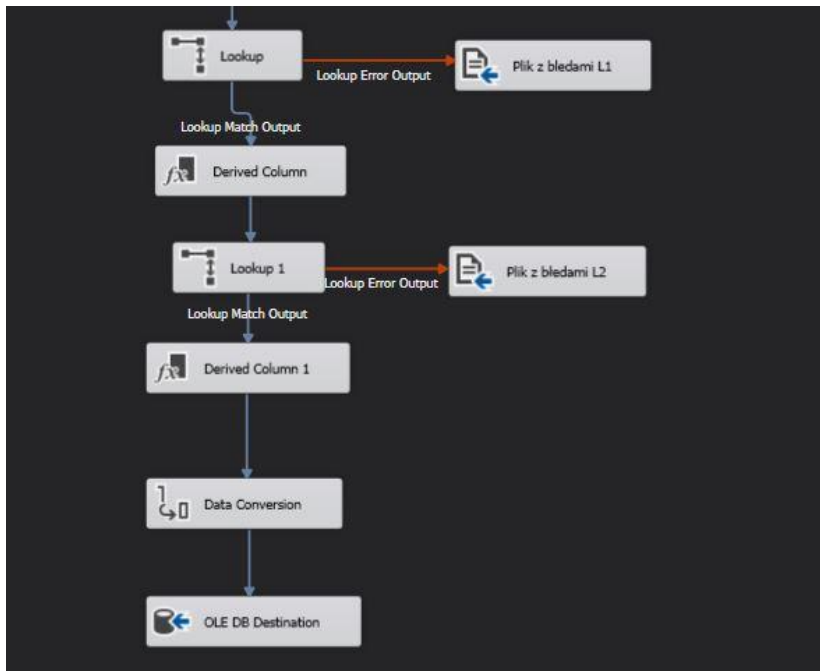


Ad.2 Przesyłanie danych z tabeli A do tabeli C odbywa się w 4 etapach.

1. Mapowanie Id produktu w bazie A na Id produktu w bazie C na podstawie nazwy.
2. Przypisanie odpowiedniego terytorium.
3. Złożenie daty sprzedaży z 3 pól – rok, miesiąc, dzień
4. Kontrola danych i wysłanie danych do bazy docelowej.

Poniższy zrzut ekranu przedstawia cały proces.



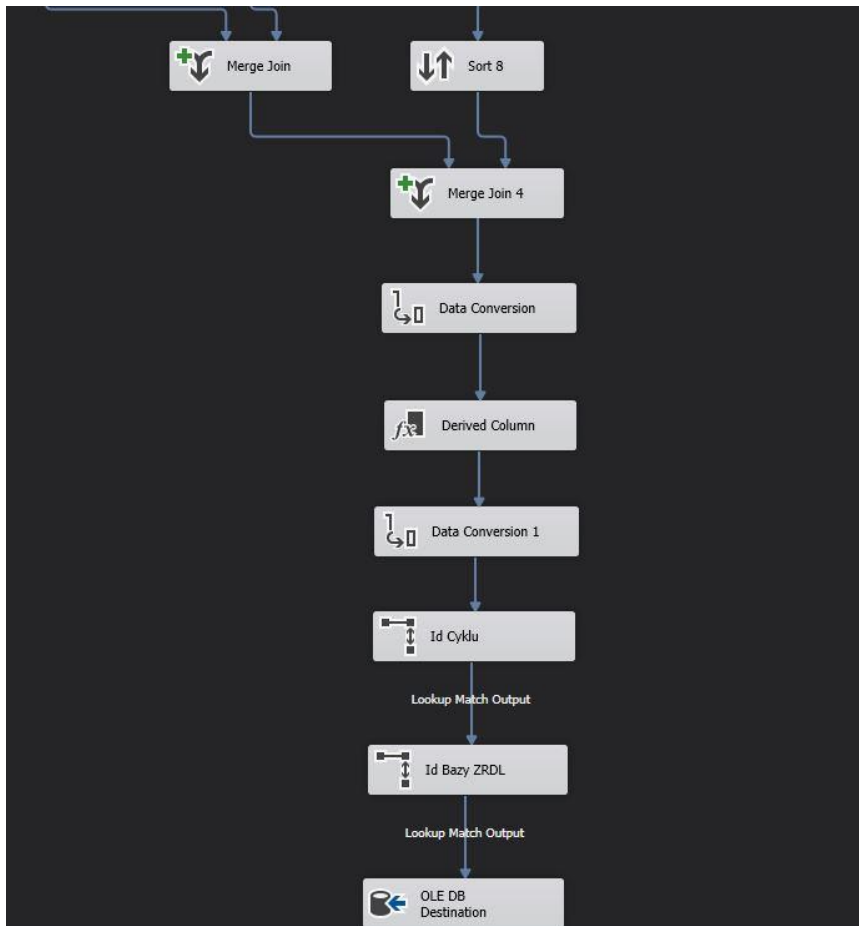


Ad.3 Przesyłanie danych z tablicy B do tablicy C odbywa się w 5 etapach.

1. Zamiana obszarów na regiony w samej bazie B
2. Mapowanie Id obszaru na Id regionu w bazie B
3. Mapowanie Id terytorium bazy B na terytorium bazy C
4. Mapowanie Id produktu bazy B na Id produktu bazy C na podstawie identyfikatora produktu zgodnie z polską normą.
5. Konwersja danych, sprawdzenie i przesył do bazy docelowej







## 5. Hurtownia danych

Hurtownia posiada jedną kostkę danych która posiada 4 wymiary i jeden fakt.

### 1. Wymiary

- C\_Dim\_Zrdlo – posiada wprowadzoną hierarchię pól gdzie pierwszym polem jest nazwa hurtowni z której pochodzą dane. Daje to obraz udziału w sprzedaży przy wybranych kryteriach

- C\_Dim\_Regiony – obszar Polski został podzielony na 5 regionów i na podstawie nazwy regionów została wyznaczona hierarchia.

- C\_Dim\_Produkty – wymiar posiada komplet informacji na temat podziału produktów – grupa produktowa, dawka, identyfikator - na tej podstawie zostały utworzone 2 hierarchie – grupa produktowa i dawkowanie

- C\_Dim\_Cykle – wymiar posiada informacje na temat cykli sprzedażowych podzielonych na lata. Aktualne dane są wprowadzane od roku 2010 do roku 2014.

### 2. Fakt

- C\_Fact\_Sales – Ilości sprzedanych jednostek, wartość sprzedanych jednostek, liczba transakcji sprzedażowych. Na podstawie tych informacji tworzone są kategorie w wykresach i tabelach Excela.

## 6. Raportowanie

Hurtownia będzie miała dostępne raporty umożliwiające analizę danych pod względem ilościowym jak i wartościowym. Raporty będą obejmować wyniki w postaci tabeli przestawnej, jak również wykresów w formie pliku Excel. Dodatkowo będzie stworzony raport SSRS który umożliwi filtrowanie danych pod kątem okresu i źródła danych.

### 6.1 Raporty Excel

- Tabela z wartościami sprzedażowymi z filtrami:
  - Nazwy grup produktu
  - Dawki
  - Regionu
  - Cyklu sprzedażowego
- Wykres sprzedażowy – **wartości** z filtrami regionu i cyklu porównujący wyniki SF i GF z filtrem okresu sprzedażowego.
- Tabela z ilościami sprzedażowymi z filtrami:
  - Nazwy grup produktu
  - Dawki
  - Regionu
  - Cyklu sprzedażowego
- Wykres sprzedażowy – **ilościowy** z filtrami regionu i cyklu porównujący wyniki SF i GF z filtrem okresu sprzedażowego.

### 6.2 Raport SSRS

Pokazuje wyniki sprzedaży w latach 2010-2014 na wykresie oraz pokazuje dane w postaci tabeli. Tabela posiada dodatkowo wyniki sprzedaży wartościowo z podziałem na regiony.