Projekt Hurtowni danych

Inżynieria Integracji usług Informatycznych

Przemysław Foltyn

2016

# 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

# Wymagania

## Wymagania funkcjonalne

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

## 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 iternetowej wyłącznie w sieci lokalnej. Musi być kompatybilny z przeglądarka Mozilla Firefox(wersja 18 wzwyż) lub Google Chrome(wersja 30 wzwyż).

Obsługiwanymi 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.

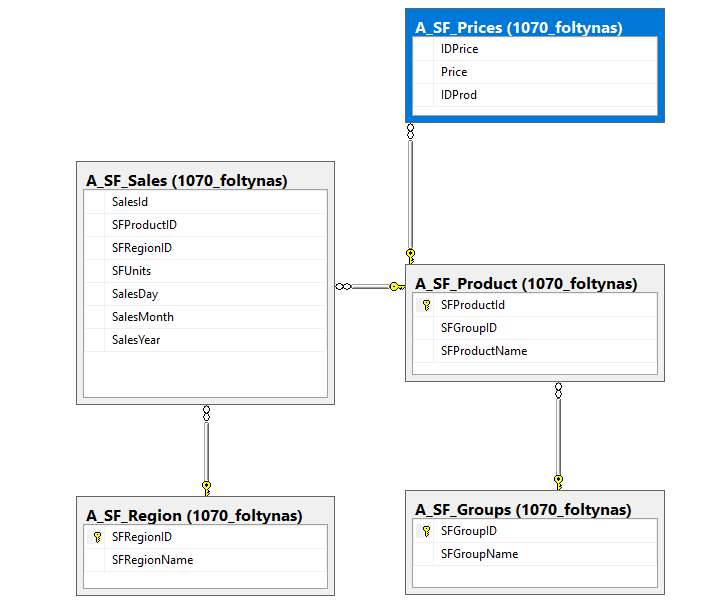
Dane będą kopiowane na zapasowy serwer znajdujący się poza siedzibą firmy. Kopia bazy danych będzie robiona:

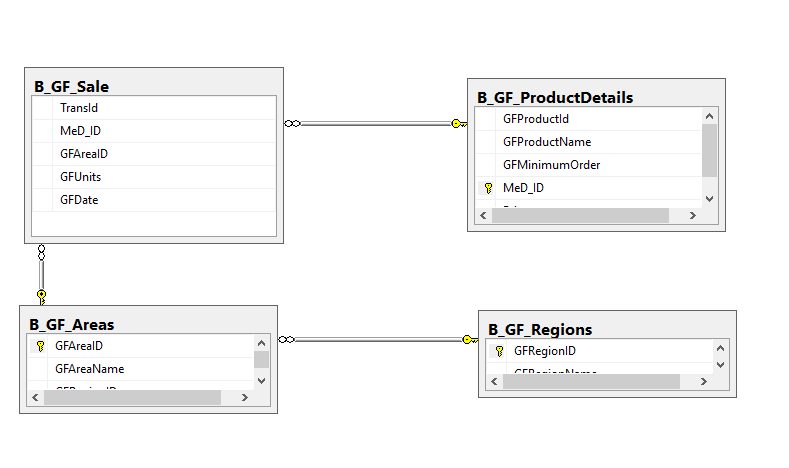
* codziennie przyrostowo
* raz w tygodniu pełna.

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

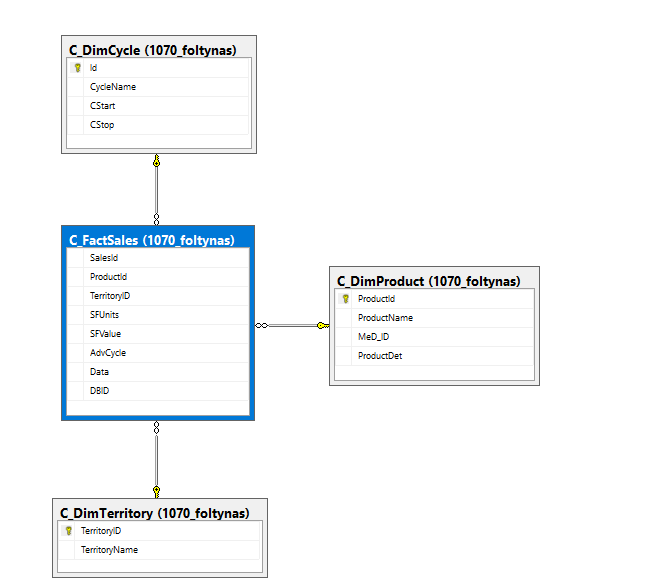
# Struktura baz danych

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

1. 
   1. Schemat bazy Superfarm
   2. Schemat bazy Grupy Farmaceutycznej



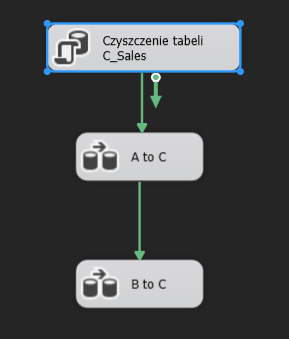
* 1. Schemat bazy przedsiębiorstwa



# 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

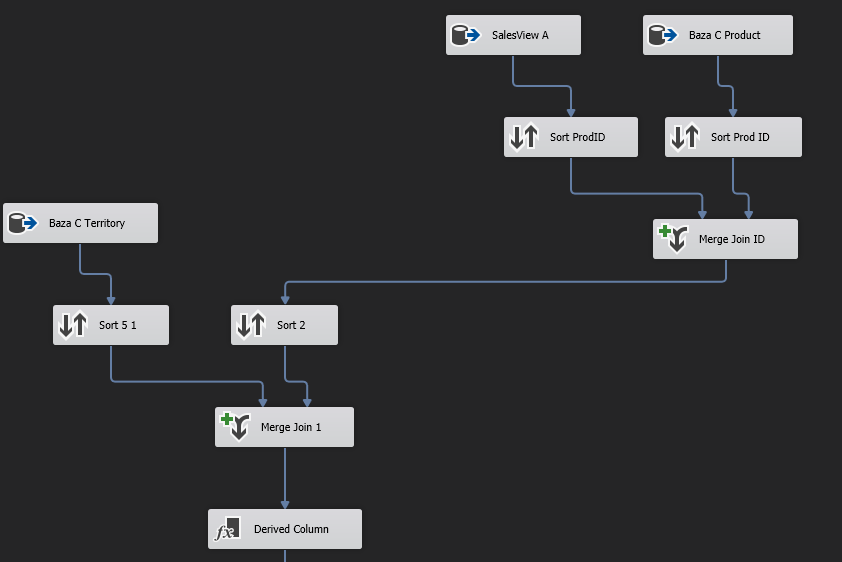


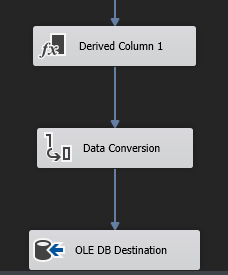
Ad.1 Czyszczenie tablicy jest potrzebne do utrzymania tablicy docelowej w rozmiarze ułatwiającym analizę danych.

Ad.2 Przesyłanie danych z tablicy A do tablicy 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.

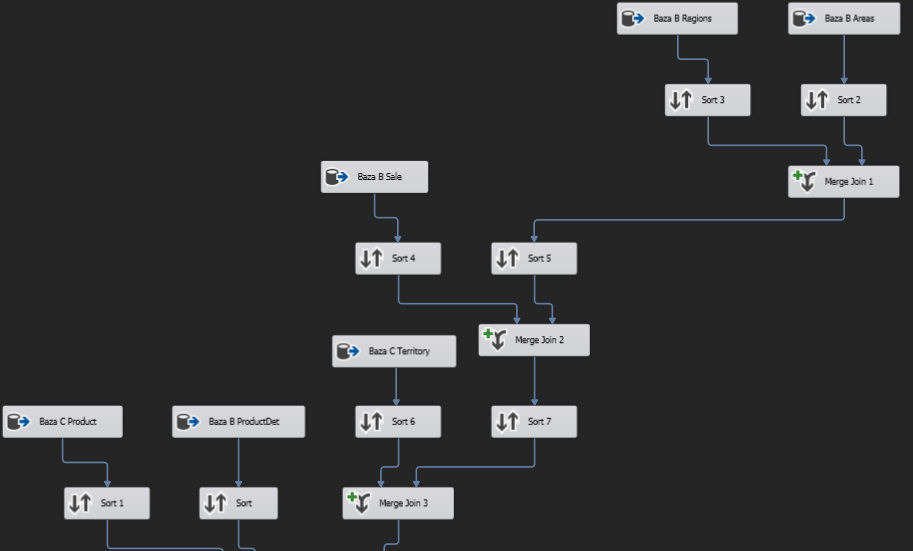


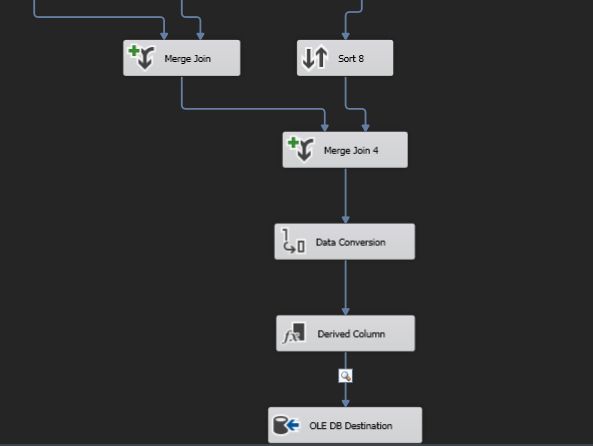


ZMIENIĆ!!!!!~!

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





ZMIENIĆ!!!!!!!!!

# 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 cyklów 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.