**Wojskowa Akademia Techniczna im. Jarosława Dąbrowskiego**



Sprawozdanie projektowe z przedmiotu Hurtownie Danych

Grupa : I5B2S1

Wykonawcy : Artur Babkiewicz, Jakub Demidziuk, Patryk Chodoła

Prowadzący : dr inż. Marcin Mazurek

**1) Zadanie projektowe**

Na podstawie danych źródłowych zawierających dane o oglądalności programów zbudować hurtownię danych umożliwiającą analizę oglądalności programów telewizyjnych.

Kamienie milowe:

1. Model bazy danych repozytorium głównego, model obszaru Stage, załadowane dane do Stage

2. Proces ładujący do repozytorium głównego, załadowane dane do repozytorium głównego

3. Baza Analysis Services (wielowymiarowe kostki danych)

4. Raportowanie ad-hoc w narzędziu Power BI.

**2) Uzasadnienie biznesowe realizowanego projektu**

Tworzona hurtownia danych pozwala na przeprowadzenie analizy oglądalności programów telewizyjnych. Dzięki tej analizie użytkownik hurtowni będzie mógł w prosty i przejrzysty sposób przejrzeć dane na temat oglądalności poszczególnych programów.

Korzyści wynikające z realizacji projektu to:

- zwiększenie wglądu na oglądalność programów telewizyjnych

- usprawnienie pracy analityków

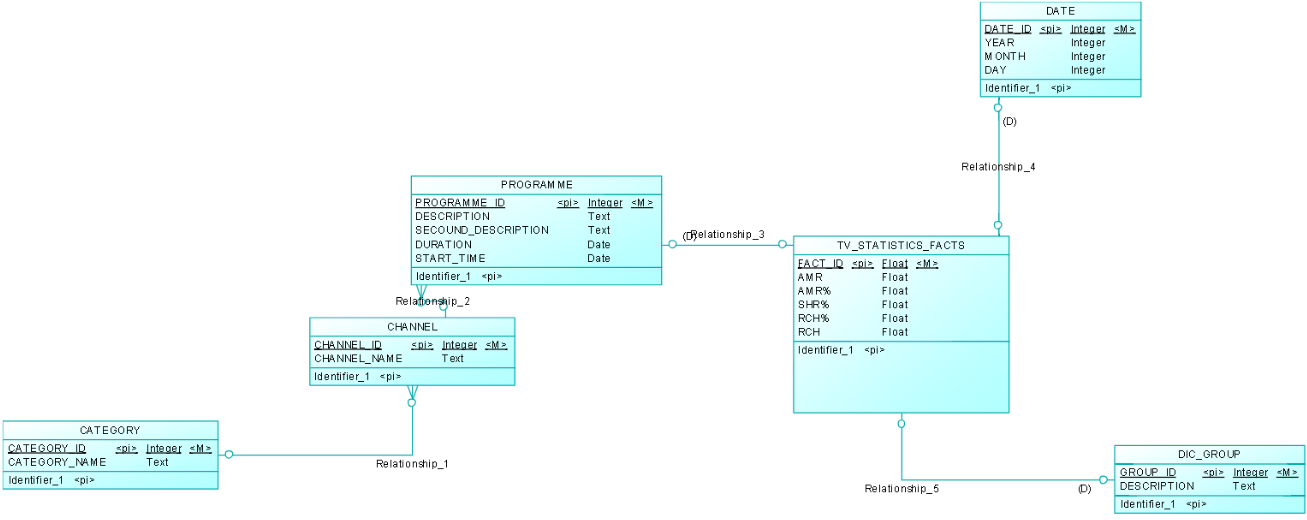
- ocenę strategii rynkowej (reklamy)

- podniesienie jakości świadczonych usług

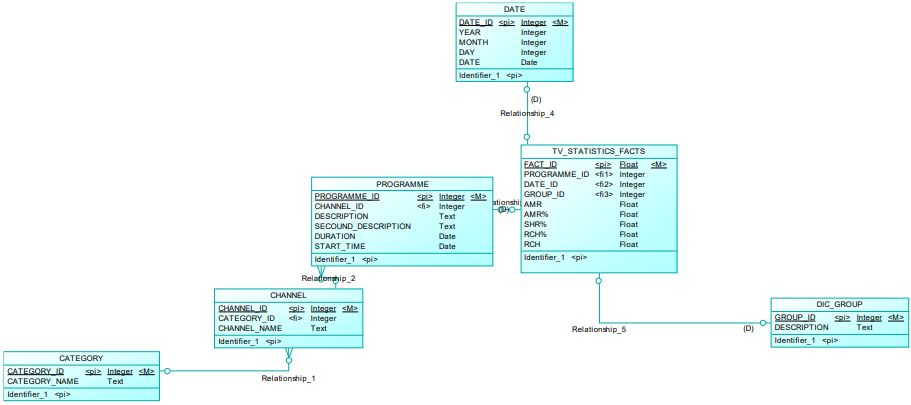
- dopasowanie treści emisji do poszczególnych grup odbiorców

**3) Modele hurtowni danych**

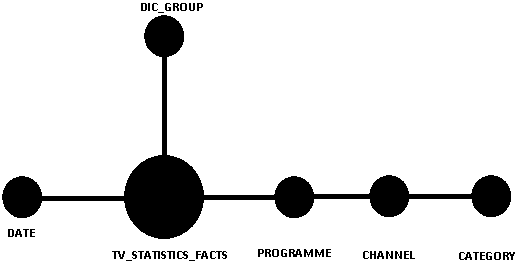
Rysunek 1 - Model wymiarowy hurtowni danych

****

Rysunek 2 - Model logiczny hurtowni danych

****

Rysunek 3 - Model punktowy

****

Fakty :

1 ) TV\_STATISTICS\_FACTS – Zdarzenie główne, które podlega analizie w hurtowni. Zawiera dane odnośnie daty, programów oraz grup odbiorców ze wszystkich tabel hurtowni

Wymiary :

1 ) DATE – pozwala na analizę danych pod kątem dnia, miesiąca i roku.

2) DIC\_GROUP – pozwala na analizę danych pod kątem grup odbiorców (wiek, płeć)

3) PROGRAMME – pozwala na analizę danych pod kątem oglądanego programu

Poziomy szczegółowości :

A ) PROGRAMME

B ) CHANNEL

C ) CATEGORY

**4) Ładowanie danych do repozytorium głównego**

Dane do repozytorium głównego ładujemy przy pomocy skryptów sql. Na początku dane były w odpowiedni sposób przygotowywane np. usunięcie niepotrzebnych spacji. W tym celu również zostały utworzone skrypty sql-owe, które działają dla wszelkich źródeł danych.

**5) Wielowymiarowa kostka danych OLAP**

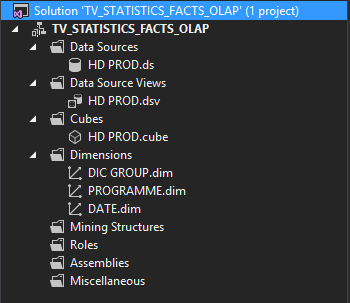
Wielowymiarową kostkę danych OLAP wykonaliśmy w Microsoft Visual Studio 2017. Jako źródło danych podaliśmy naszą utworzoną wcześniej, załadowaną danymi w SQL Server Management Studio hurtownię danych. Utworzyliśmy 3 wymiary na podstawie których możliwa jest analiza interesująych nas danych.

1) Wymiar czasu (Poziomy : Data -> Rok -> Miesiąc -> Dzień)

2) Wymiar programu (Poziomy : Program -> Kanał -> Grupa kanałów)

3) Wymiar grupy odbiorców (Poziomy : Grupa)

Zdjęcie 1 – Widok ogólny projektu kostki



Miary użyte do analizy :

1) AMR

2) AMR%

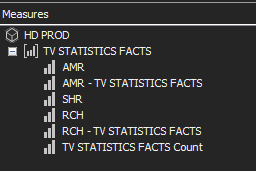
3) RCH

4) RCH %

5) SHR %

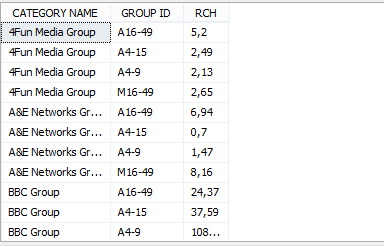
6) Liczba pomiarów

Zdjęcie 2 – Miary analizy kostki



Tak utworzona kostka pozwala nam na szeroki wybór filtrów do analizy danych i uzyskania ciekawych wyników.

Zdjęcie 3 – Przykład analizy kostką OLAP



**6) Raporty Power-BI**

Za pomocą raportów power – BI jesteśmy wstanie tworzyć raporty z wynikami bardzo podobnymi do działania kostki OLAP. Takie raportowanie daje nam jednak więcej możliwości jak dynamiczne filtry do każdego z wymiarów. Utworzyliśmy jeden przykładowy raport ukazujący rankig programów z największą oglądalnością dla poszcególnych grup odbiorców. Dodaliśmy dwa filtry dynamiczne : dla dat oraz grup kanałów.

Zdjęcie 4 – Przykładowy raport AD-HOC. Ranking programów z największą oglądalnością dla 4 grup odbiorców



**7) Wnioski i podsumowanie**

Celem wytworzenia hurtowni danych jest wspomożenie działalności organizacji. Osiąga się to poprzez analizę wycinka rzeczywistości pod kątem interesujących nas aspektów. Dane zewnętrzne potrzebne do utworzenia hurtowni najczęściej pochodzą z wielu innych źródeł (bazy danych, pliki z danymi itp.), więc proces ładowania danych wymaga zintegrowania i zjednolicenia danych pod kątem hurtowni.

W naszym przypadku stworzenie hurtowni danych miało pomóc potencjalnemu analitykowi w analizie oglądalności poszczególnych kanałów z możliwością uzyskania wiedzy o poszczególnych odbiorcach, na dany dzień. Zdobyta przez niego wiedza pozwoliłaby mu na poznanie odbiorców, ich upodobań oraz pewnego rodzaju zachowań. Dawałoby to olbrzymie możliwości, np. zaproponowanie lepszych rozwiązań dopasowanych do danej grupy odbiorców czy doboru reklam. Aby tego dokonać należałoby udostępnić analitykowi raporty zawierające interesujące go dane.

Udało nam się zrealizować wszystkie punkty projektowe dzięki którym zdobyliśmy cenną wiedzę na temat hurtowni danych.