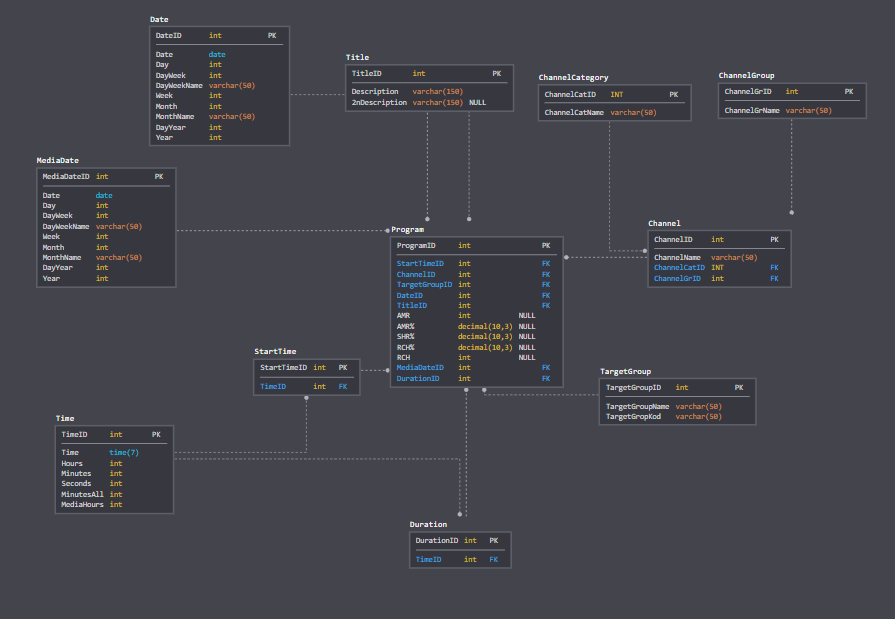
HURTOWNIE DANYCH

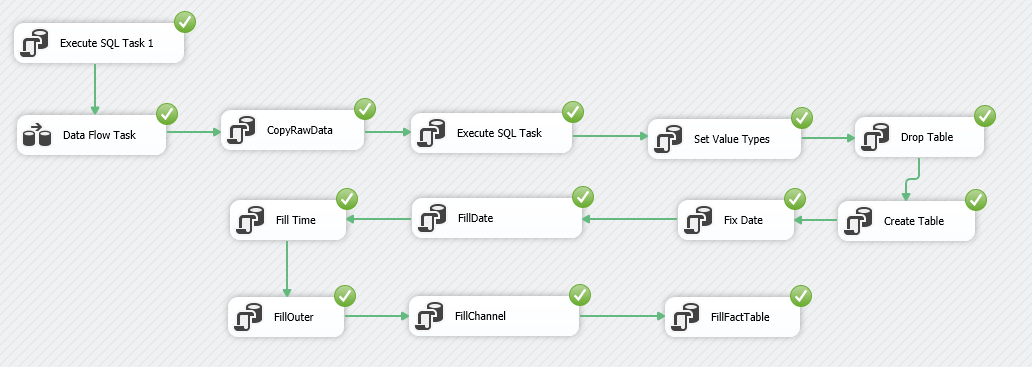
Sprawozdanie z projektu

Prowadzący: dr inż. Marcin Mazurek  
Wykonali: plut. pchor. Angelika Ziółkowska  
 plut. pchor. Adam Zagórski  
Data: 15.06.2018

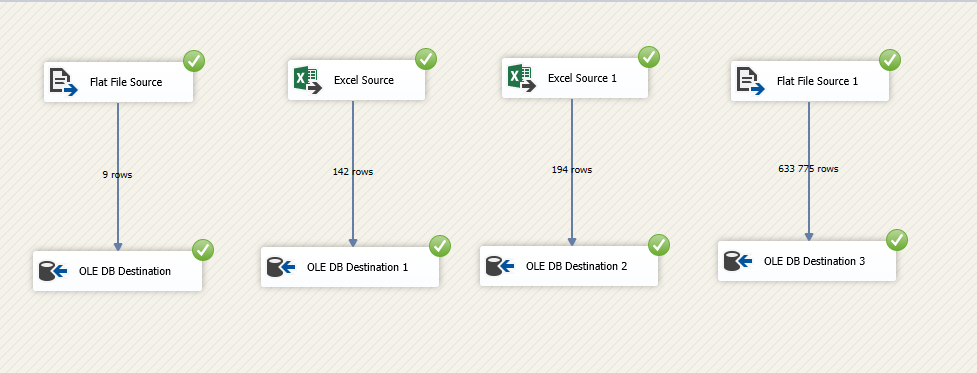
# MODELE I DIAGRAMY



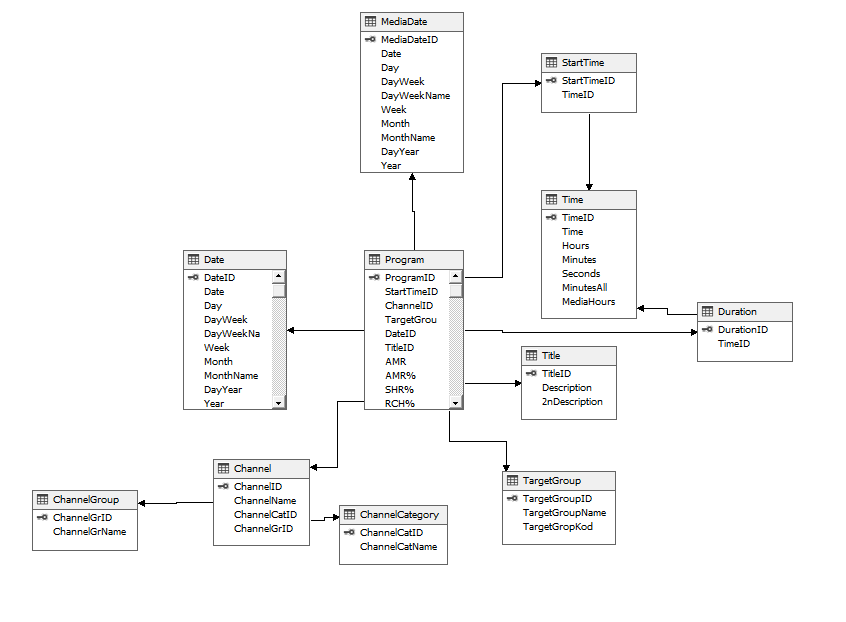
Rys. 1. Model repozytorium głównego



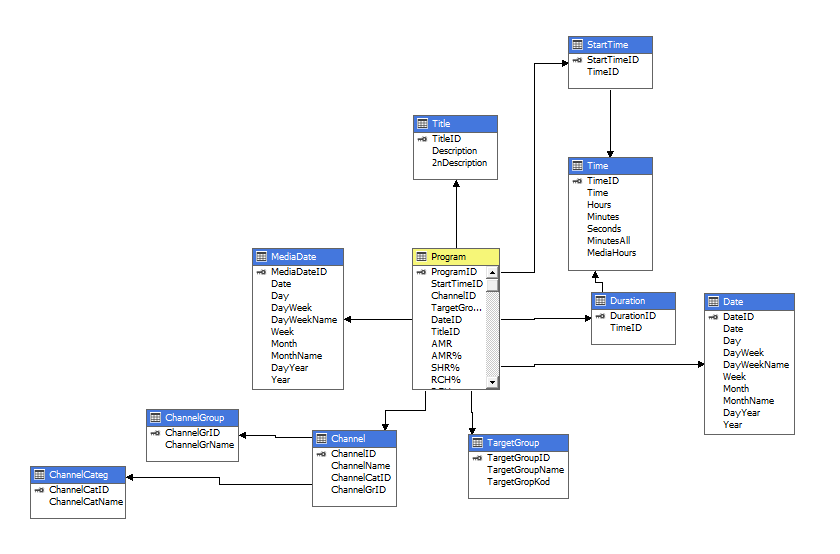
Rys. 2. Diagram sterowania transmisją (Control Flow)



Rys. 3.Diagram przepływu danych (Data Flow)



Rys. 4. Widok źródła danych



Rys. 5. Model kostki danych

# WYMIARY

Wymiar jest to cecha opisująca dany fakt, pozwalająca powiązać go z innymi pojęciami modelu.   
W naszym projekcie występuje łącznie 10 wymiarów. Większość z nich zawiera w sobie hierarchie, które ułatwiają później tworzenie raportów.

Wymiary w projekcie:

* Channel
* Channel Category
* Channel Group
* Date
* Duration
* Media Date
* Program
* Start Time
* Target Group
* Title

Na początku tworzenia projektu popełniliśmy błąd. Nie powinniśmy byli traktować Duration jako wymiar, ale miarę. Byłby on wtedy wyliczany w minutach dla ułatwienia obsługi raportów.

# NAPOTKANE PROBLEMY

Podczas tworzenia projektu napotkaliśmy wiele problemów, m. in.:

* Stworzenie prawidłowego modelu repozytorium głównego

Z każdym kolejnym kamieniem milowym należało zmienić model repozytorium głównego, gdyż był on tworzony zgodnie z aktualnym etapem. Zalecane jest dokładniejsze zapoznanie się z całym zadaniem, aby uniknąć wielokrotnego tworzenia projektu od nowa.

* Konwersja formatu daty 0

Format daty został przez nas źle zrozumiany. Pomyliliśmy dni z miesiącami, przez co raport został błednie wykonany.

* Wypełnianie pustych pól w wierszach, wartością znajdującą się powyżej

Aby uniknąć uzupełniania wypełniania pustych pól kursorami, które wydłużyłyby czas ładowania się repozytorium, użyliśmy TABLOCKX.

Przykład:

UPDATE [HD].[dbo].[Dane] WITH(TABLOCKX)

SET @v = [Description] = CASE WHEN [Description] = '' THEN @v

ELSE [Description] END

OPTION(MAXDOP 1);

* Czas ładowania pełnego repozytorium

Podczas wypełniania tabel repozytorium musieliśmy zrzygnować z używania kursorów, które znacznie wydłużały czas ładowania się pełnego repozytorium

# OPIS BLOKÓW SQL TASK

**Execute SQL Task 1** - tworzy tabele, do których przenoszone będą dane z plików.

**CopyRawData** - Przenoszenie danych wejściowych do tabeli roboczej.

**Execute** **SQL** **Task** - Formatowanie danych i wypełnianie pustych pól w tabeli roboczej.

**Set** **Value** **Types** - Zmiana typów danych w tabeli roboczej.

**Drop** **table** - Usuwanie repozytorium, jeżeli istnieje.

**Create** **Table** - Tworzenie tabel nowego repozytorium.

**Fix** **Date** - Zmiana czasu z mediowego, na astronomiczny (wraz ze zmianą daty).

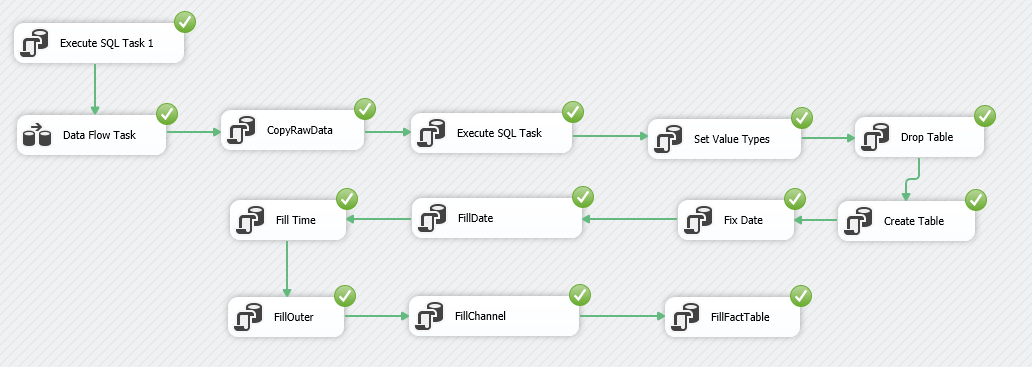
**FillDate** - Uzupełnianie tabeli Date i MediaDate w repozytorium.

**Fill** **Time** - Uzupełnianie tabel Time, StartTime i Duration.

**Fill** **Outer** - Uzupełnianie innych tabel, nie kożystających z kluczy obcych.

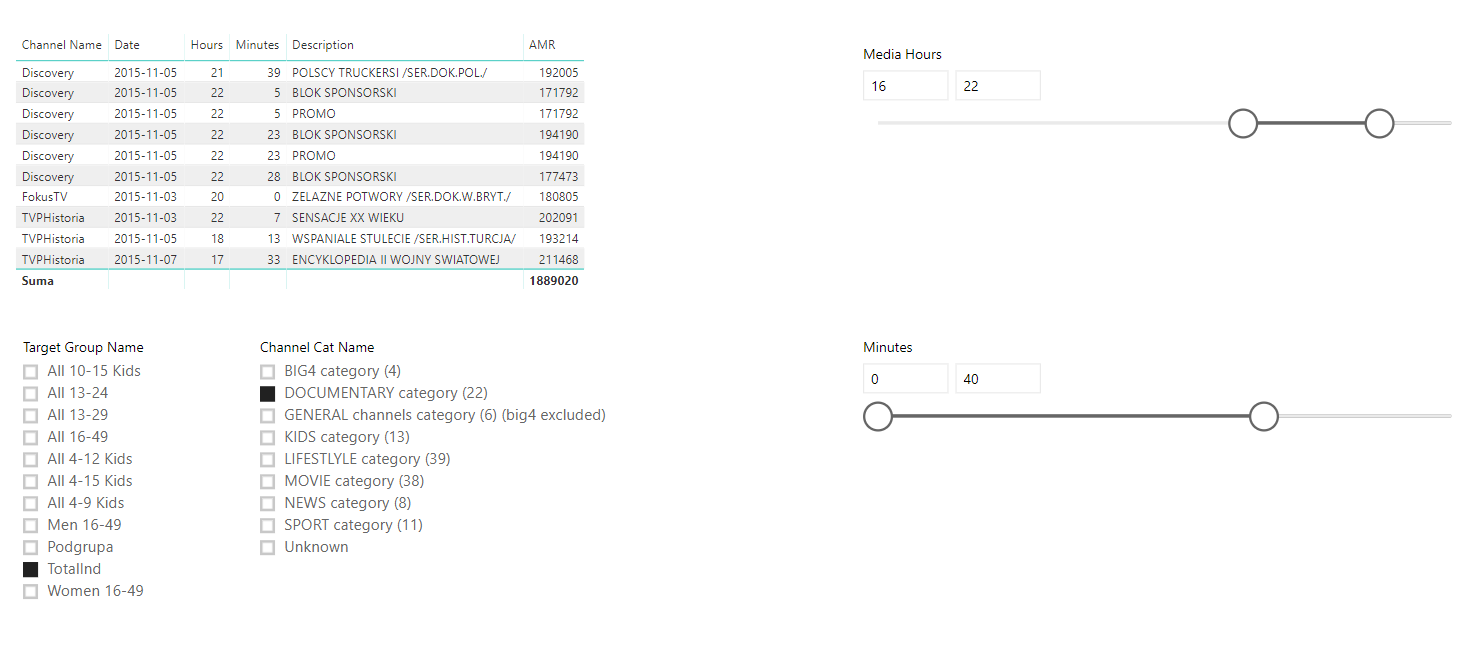
**FillChannel** - Uzupełnianie tabeli Channel.

**FillFactTable** - Uzupełnianie tablei „Program” (Tabeli faktów).

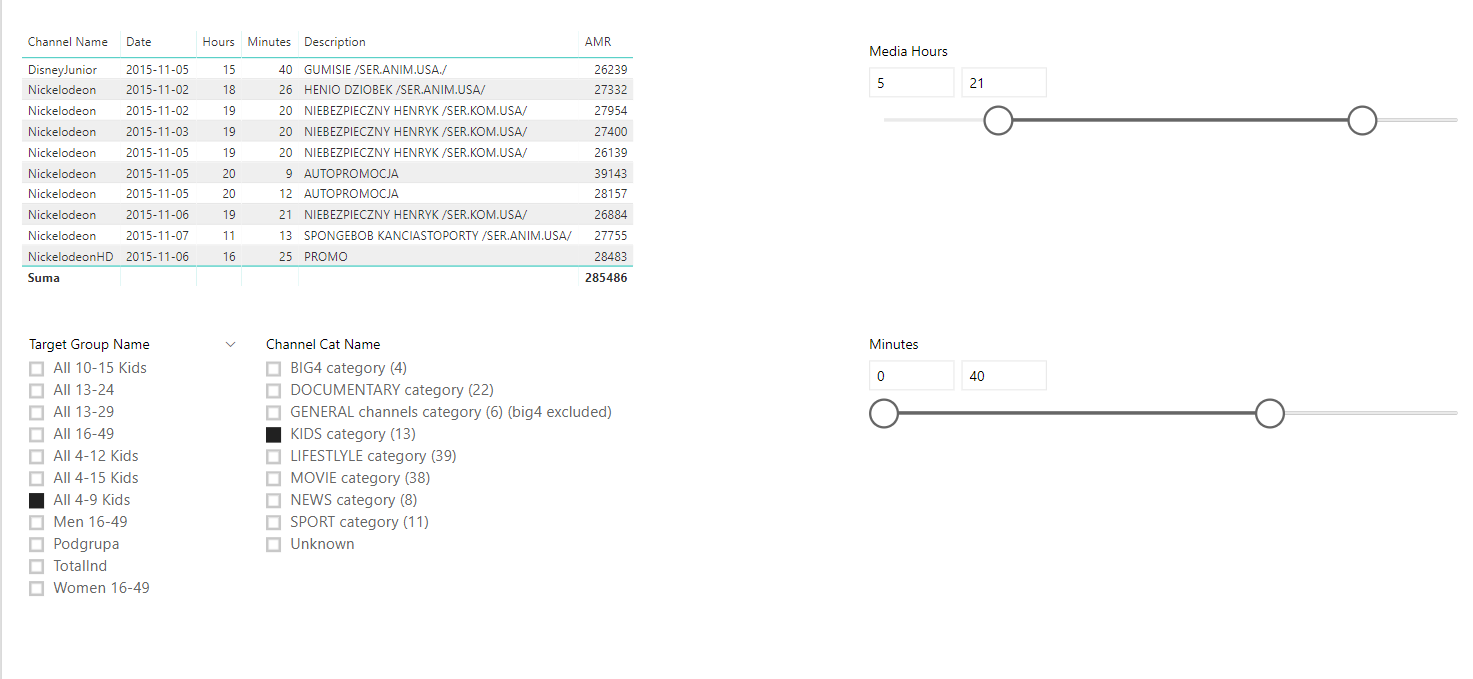


Rys. 6. Diagram sterowania transmisją (Control Flow) – ponowne ukazanie w celu ułatwienia zapoznania się z opisami bloków SQL Task

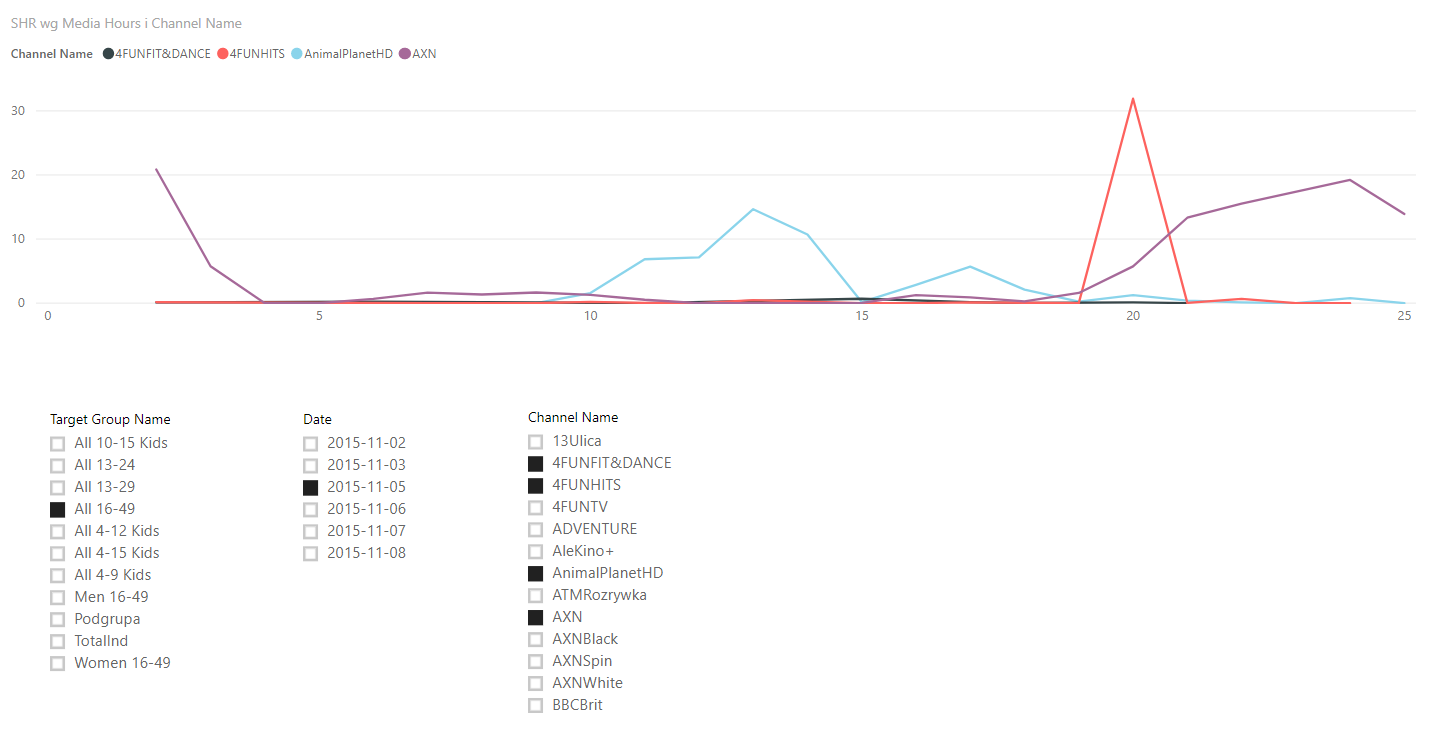
# PRZYKŁADY DZIAŁANIA RAPORTÓW:



Rys. 7. Raport, ukazujący 10 programów o największej oglądalności AMR dla kanałów dokumentalnych, oglądanych przez wszystkie kategorie osób



Rys. 8. Raport, ukazujący 10 programów o największej oglądalności AMR dla kanałów dziecięcych, oglądanych przez dzieci w wieku 4-9 lat



Rys. 9. Raport, ukazujący wykres oglądalności SHR (%) w ciągu doby, liczonej czasem medialnym, dla wybranych kanałów, oglądanych przez wszystkie osoby przedziale wiekowym 16-49