**WOJSKOWA AKADEMIA TECHNICZNA**

Wydział Cybernetyki



Laboratorium z przedmiotu:

Hurtownie Danych

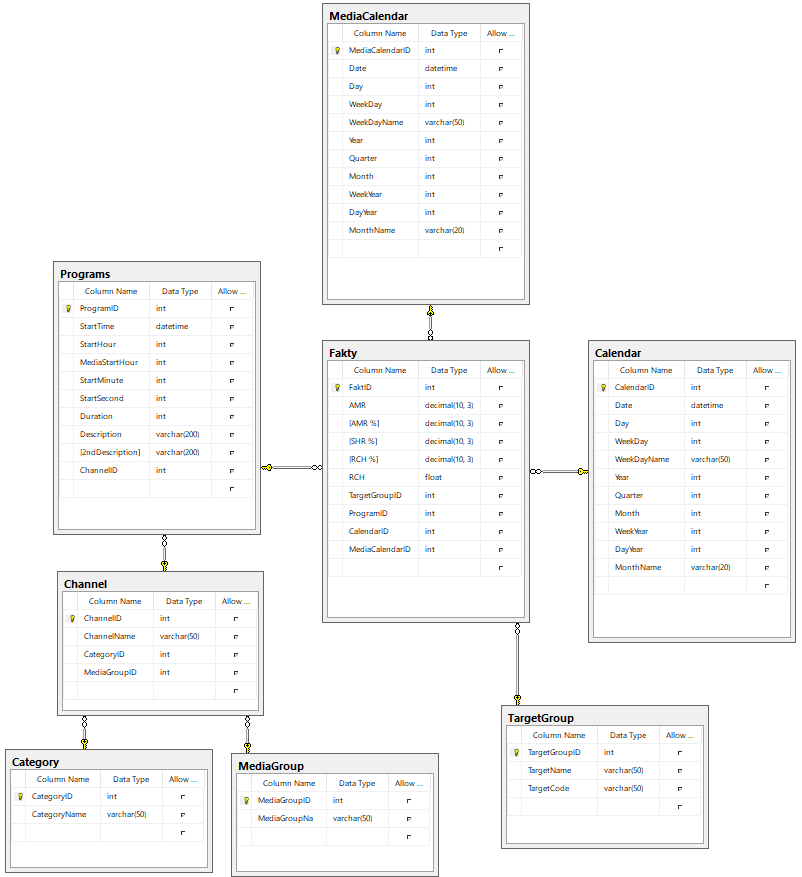
**Prowadzący**: dr inż. Marcin Mazurek

**Autorzy**: sierż. pchor. Paweł Wójcik

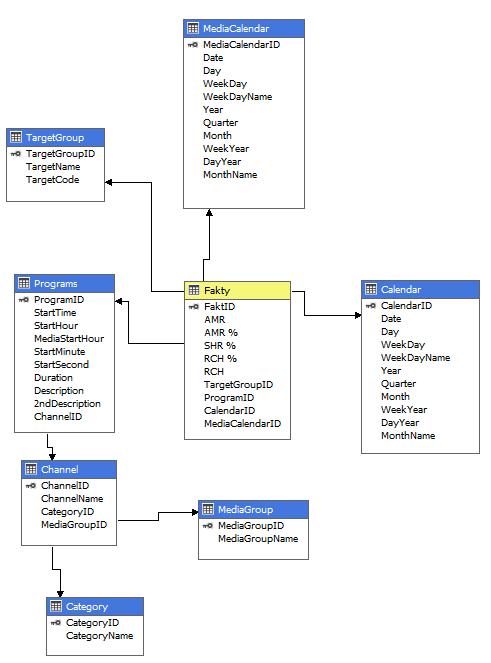
plut. pchor. Gabriel Wrzosek

**Grupa**: I5B1S1

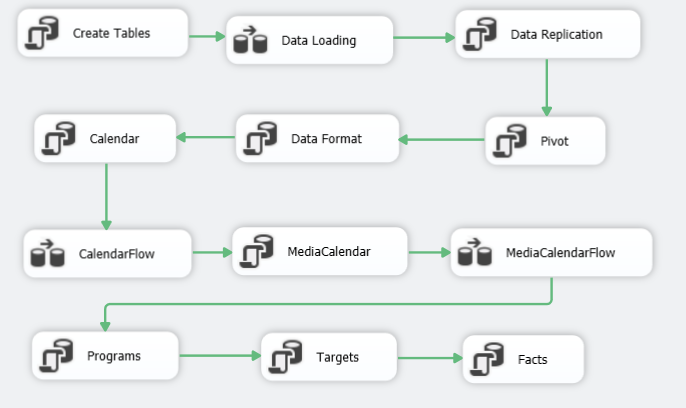
1. Model danych repozytorium głównego



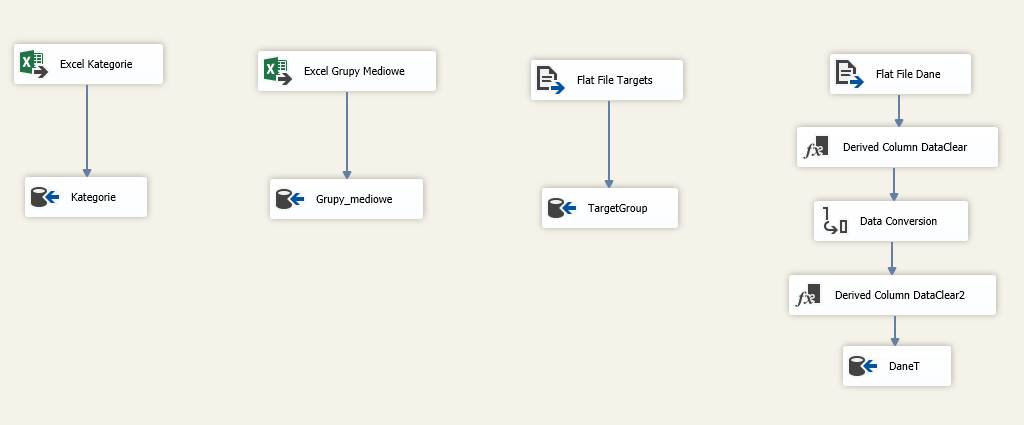
1. Model „Cube”



1. Model procesu ETL



1. Elementy DataFlow



1. Wnioski

- Początkowa koncepcja formatowania danych opierała się na sporej ilości skryptów i częstym używaniu kursorów – koncepcja okazała się bardzo NIE trafna głównie pod względem czasochłonności wykonywania a także pod względem ilości pisanego kodu skryptowego

- Do redundancji ilości skryptów i pozbycia się jak największej ilości kursorów ( które znacznie opóźniają wykonanie procesu) użyliśmy zatem pakietu SSIS z Visual Studio. Użycie takich komponentów jak Derived Column czy Data Conversation, nie tylko przyśpieszyło w dużym stopniu wykonanie całości, ale także zastąpiło większość używanych dotychczas skryptów (nareszcie mogliśmy się swobodnie poruszać po napisanych przez nas kodach, bez konieczności zastanawiania się – gdzie my to wcześniej umieściliśmy)

- Większość tabel próbowaliśmy uzupełniać SELECT’ami, zamiast używać do tego kursora. Zastosowanie takiego rozwiązania pozwoliło przyśpieszyć wykonywanie procesu uzupełniania tabel około pięciokrotnie.

- Podczas tworzenia modelu ETL umieszczanie kolejno wykonywanych części procesu w oddzielnych komponentach i skryptach umożliwiało nam szybkie zlokalizowanie miejsca wystąpienia problemu

- W komponentach ‘DataFlow’ wykorzystanie ‘Data Viever’ pozwoliło na dokładnie kontrolowanie poprawności wykonywania wczytywania i transformowania danych nawet w momencie samego wykonywania Task’a

- Użycie DATEPART w SSIS było bardzo dobrym pomysłem jeśli chodzi o ekstraktowanie poszczególnych elementów daty. Jednak w odróżnieniu od Transact SQL, SSIS w VS nie obsługuje DATENAME zatem wyciągniecie z daty takich elementów jak nazwa dnia tygodnia lub miesiąca musiało zostać zatem zastąpione dość długim skryptem z użyciem DATEPART np:

(DATEPART("dw",Date) == 1 ? "Monday" : DATEPART("dw",Date) == 2 ? "Tuesday" : DATEPART("dw",Date) == 3 ? "Wednesday" : DATEPART("dw",Date) == 4 ? "Thursay" : DATEPART("dw",Date) == 5 ? "Friday" : DATEPART("dw",Date) == 6 ? "Saturday" : DATEPART("dw",Date) == 7 ? "Sunday" : "")