**WOJSKOWA AKADEMIA TECHNICZNA**



**HURTOWNIE DANYCH**

Sprawozdanie z projektu

**Temat:** Budowa hurtowni danych na temat transportu lotniczego w USA

**Prowadzący:** dr inż. Marcin Mazurek

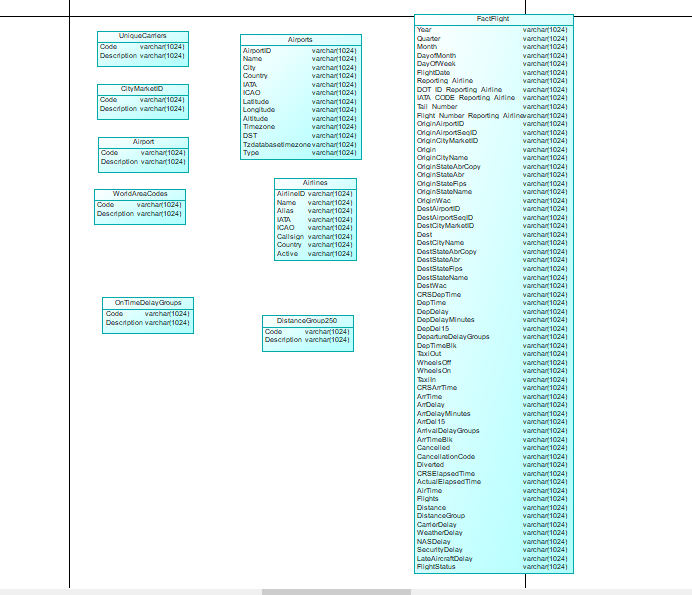
**Imię i nazwisko:** Marek Rośkowicz

**Grupa szkoleniowa:** I6E3S1

**Numer indeksu:** 66183

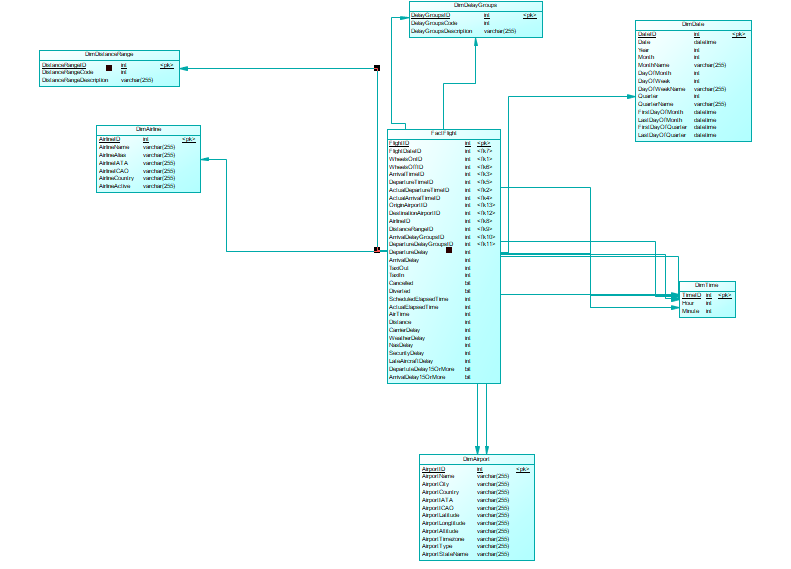
# Modele

* 1. Model fizyczny obszaru stage



Rys. 1 Model fizyczny obszaru stage

* 1. Model fizyczny repozytorium głównego



Rys. 2 Model fizyczny obszaru repozytorium głównego

# ETL

Proces ETL został przeprowadzony za pomocą procedur składowych załadowywanych do obszaru repozytorium głównego hurtowni. Administracja tym procesem odbywa się z poziomu narzędzia SSIS, którego zadaniem jest uruchomienie odpowiednich procedur. Procedury te wykonują następujące czynności:

* Załadowanie danych z plików .csv oraz .dat do tabel w obszarze stage
* Oczyszczenie danych oraz załadowanie danych z obszaru stage to tabel wymiarowych oraz utworzenie kluczy sztucznych dla tabel wymiarów
* Załadowanie tabeli faktów obszaru repozytorium głównego oraz umieszczenie w repozytorium głównym kluczy obcych do tabel wymiarów.

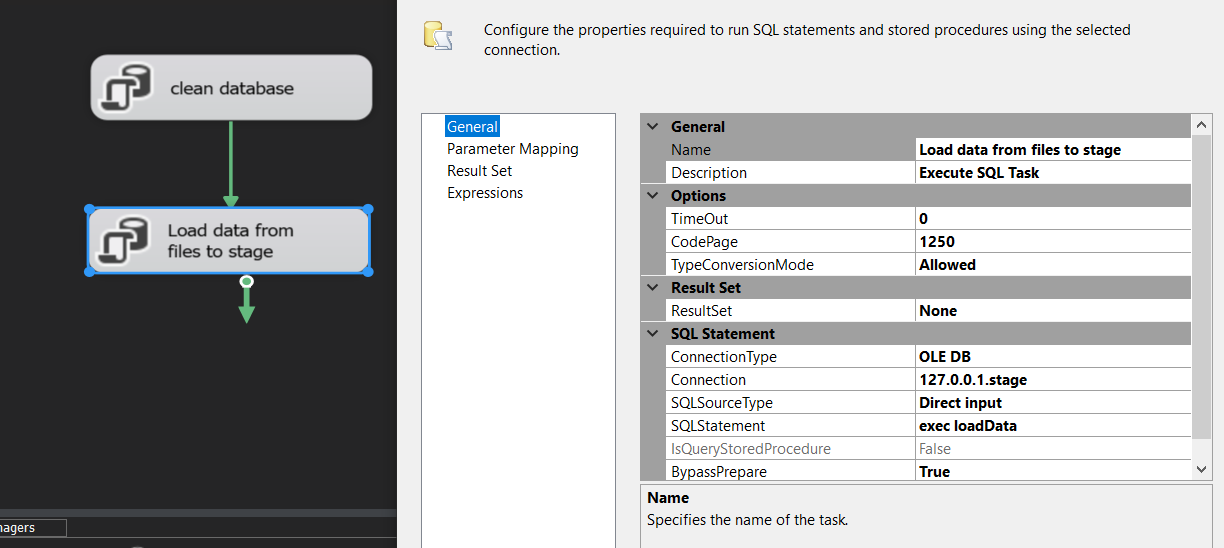
**Procedury użyte podczas procesu ETL:**

* cleanDatabase – procedura czyszcząca obszar stage z danych
* loadData – procedura ładująca dane do obszaru stage z plików .csv za pomocą funkcji bulk insert
* insert\_dimData – procedura ładujące daty z roku 2019 do tabeli wymiaru Daty
* insert\_dimTime – procedura ładująca czas do tabeli wymiaru czasu
* insert\_allDimension – procedura wywołująca podprocedury ładujące dane to tabel wymiarów
* Insert\_FactFlightActivity – procedura zasilająca danymi tabele faktów.
* reset\_warehouse – czyści dane z obszaru repozytorium głównego poprzez na początku usunięcie kluczy obcych usunięcie danych ze wszystkich tabel oraz następnie ponowne połączenie tabel z pomocą kluczy obcych
* reset\_flightIDSEQ – resetuje sekwencje tak aby ponowne ładowani danych do obszaru głównego repozytorium nadawało kluczom głównym wartości rozpoczynające się od wartości 1

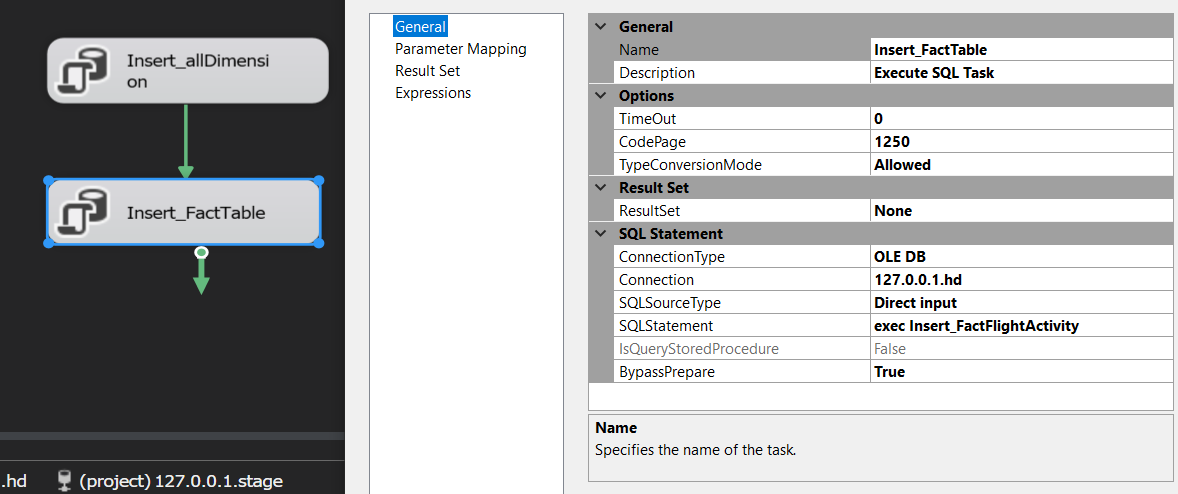
**Podczas procesu ETL, transformacja danych przyjmuje następujące założenia:**

* Dla każdego wiersza w tabelach wymiarów tworzony jest klucz sztuczny
* Dane, które są pustymi wierszami („”) lub wartościami znaku nowego wiersza (\N) są zamieniane na wartość NULL dla danych w tabelach DimAirport i DimAirlines.
* Usuwane są znaki „” dla danych będących typu varchar
* Ponieważ wszystkie dane załadowane do obszaru stage są danymi typu varchar, podczas procesu ETL zostają zamienione typy danych np. FlightDate w tabeli FactFlight z typu varchar zostaje zamieniony na typ date
* Miary które są typu time np. CarrierDelay w tabeli faktów zostają poddane obróbce poprzez uciencie końcówki .00 za pomocą funkcji substring() a następnie konwersje z typu varchar na typ int.

**Wygląd procesu ETL z poziomu narzędzia SSIS**

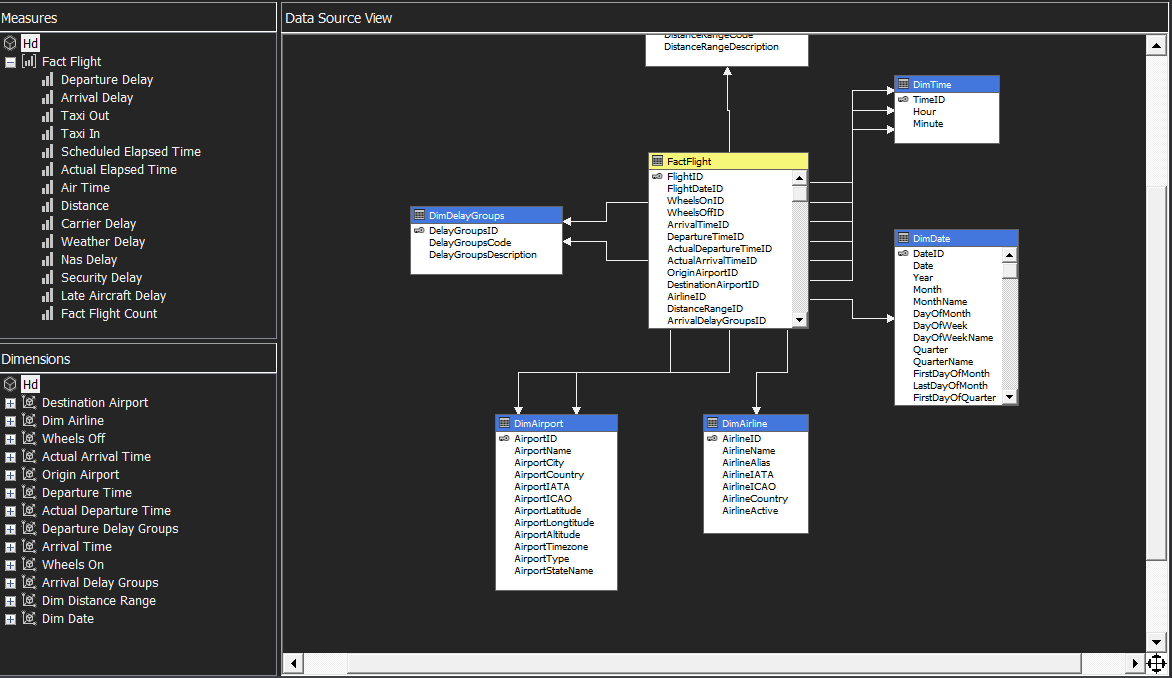
****

Rys. 3 Proces zasilania obszaru stage danymi z plików .csv

****

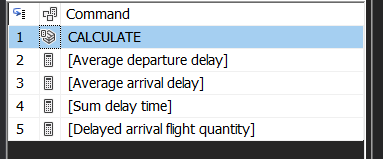
Rys. 4 Proces zasilania obszaru repozytorium głównego(ładowanie tabel wymiarów , następnie tabeli faktów)

# Wielowymiarowa kostka danych

****

Ry**s** 5. Model wielowymiarowej kostki danych. Po lewej stronie zdjęcia możemy zobaczyć dostępne wymiary [Dimensions] oraz miary [Measures].

Dodatkowo podczas tworzenia kostki danych utworzyłem kilka dodatkowych miar analizy wymaganych w projekcie. Są to:



Ry**s** 6. Utworzone miary

# Raporty

