

Wojskowa Akademia Techniczna



Przedmiot: ***Hurtownie danych***
Zadanie projektowe

Prowadzący: dr inż. Marcin Mazurek

Studenci: Dobrowolski Filip, Jaremko Krzysztof

Spis treści

Dokumentacja techniczna.....	3
Uzasadnienie biznesowe realizowanego projektu.....	4
Modele danych.....	4
Zamodelowane procesy w Visual Studio	5
Kostki danych	7
Raporty Power BI.....	8
Wnioski i podsumowanie	10

Dokumentacja techniczna

1. Microsoft SQL Server (MS SQL)

System zarządzania bazą danych, wspierany i rozpowszechniany przez korporację Microsoft.

2. Power Designer

Narzędzie typu CASE (*Computer Aided System Engineering*) firmy Sybase służące do modelowania systemów. Posiada funkcje umożliwiające modelowanie różnych domen architektonicznych.

Program posłużył nam do stworzenia modelu wymiarowego oraz logicznego hurtowni danych.

3. SQL Server Management Studio

Zintegrowane środowisko do zarządzania wszystkimi komponentami (baza danych, usługi analityczne, usługi raportowe itd.), wchodzącymi w skład Microsoft SQL Server. Zawiera narzędzia do konfiguracji, monitorowania i administrowania instancjami SQL Server. Umożliwia budowę zapytań i skryptów.

Program ułatwił nam zarządzanie hurtownią danych. Posłużył nam także w celu załadowania danych do STAGE, a następnie wypełnienia danymi Hurtowni danych.

4. Microsoft Visual Studio Community 2015 + SQL Server Data Tools

Zintegrowane środowisko programistyczne firmy Microsoft.

SSDT- to nowoczesne narzędzie programistyczne do budowania relacyjnych baz danych SQL Server, baz danych SQL Azure, modeli danych Analysis Services (AS), pakietów Integration Services (IS) i raportów Reporting Services (RS).

5. Power BI

Analityczne narzędzie wykorzystywane w biznesie służące do wizualizacji danych i tworzenia raportów .

Uzasadnienie biznesowe realizowanego projektu

Stworzona hurtownia danych umożliwia przeprowadzenie analizy transportów lotniczych w USA.

Korzyści biznesowe z realizowanego projektu mogą być następujące:

- Ocenę sprawności połączeń między lotniskami
- Ocenę jakości świadczonych usług transportowych
- Umożliwienie szerokiego wglądu w transporty lotnicze na terenie USA
- Usprawnienie połączeń między lotniskami
- Usprawnienie jakości świadczonych usług przewozowych

Modele danych

Model obszaru Stage

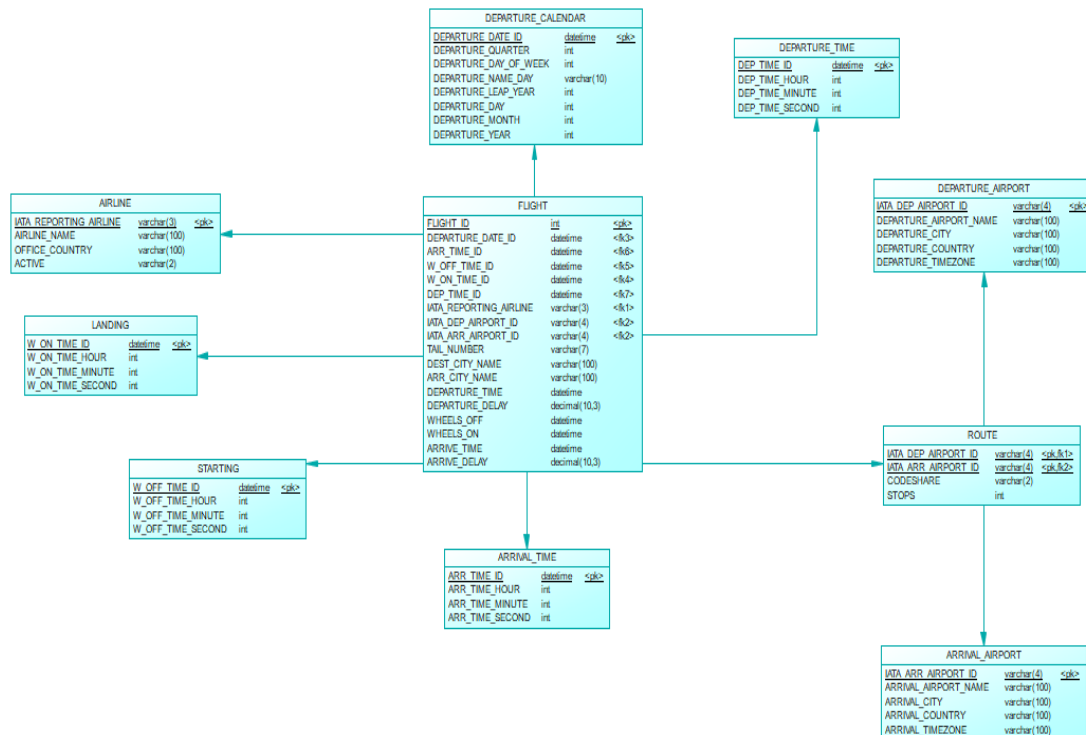
DIM_AIRLINE	
OF_AIRLINE_ID	varchar(MAX)
AIRLINE_NAME	varchar(MAX)
ALIAS	varchar(MAX)
AIRLINE_IATA	varchar(MAX)
AIRLINE_ICAO	varchar(MAX)
CALLSIGN	varchar(MAX)
AIRLINE_COUNTRY	varchar(MAX)
ACTIVE	varchar(MAX)

DIM_ROUTE	
AIRLINE_CODE	varchar(MAX)
AIRLINE_OF_ID	varchar(MAX)
SOURCE_AIRPORT	varchar(MAX)
SOURCE_AIRPORT_ID	varchar(MAX)
DESTINATION_AIRPORT	varchar(MAX)
DESTINATION_AIRPORT_ID	varchar(MAX)
CODESHARE	varchar(MAX)
STOPS	varchar(MAX)
EQUIPMENT	varchar(MAX)

DIM_AIRPORT	
OF_AIRPORT_ID	varchar(MAX)
AIRPORT_NAME	varchar(MAX)
CITY	varchar(MAX)
AIRPORT_COUNTRY	varchar(MAX)
AIRPORT_IATA	varchar(MAX)
AIRPORT_ICAO	varchar(MAX)
LATITUDE	varchar(MAX)
LONGITUDE	varchar(MAX)
ALTITUDE	varchar(MAX)
TIMEZONE	varchar(MAX)
DST	varchar(MAX)
TZ_TIMEZONE	varchar(MAX)
TYPE	varchar(MAX)
SOURCE	varchar(MAX)

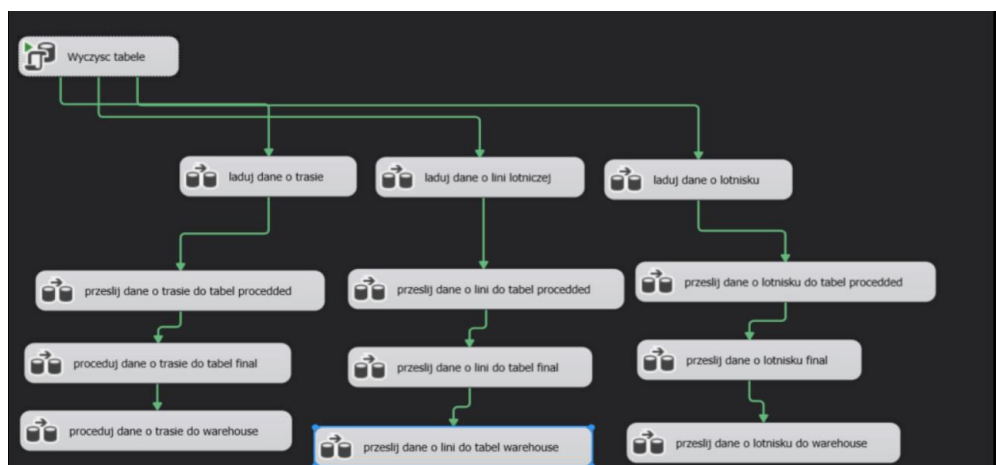
FACT_FLIGHT	
FL_DATE	varchar(MAX)
OP_CARRIER	varchar(MAX)
TAIL_NUMBER	varchar(MAX)
ORIGIN	varchar(MAX)
ORIGIN_CITY_NAME	varchar(MAX)
DEST	varchar(MAX)
DEST_CITY_NAME	varchar(MAX)
DEP_TIME	varchar(MAX)
DEP_DELAY	varchar(MAX)
WHEELS_OFF	varchar(MAX)
WHEELS_ON	varchar(MAX)
ARR_TIME	varchar(MAX)

Model repozytorium głównego bazy danych

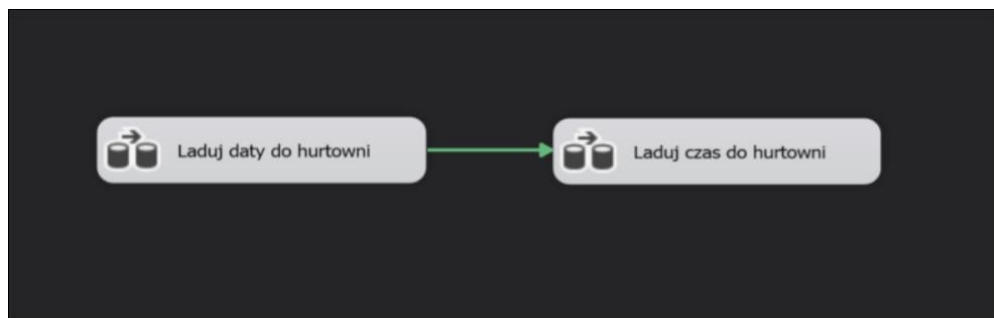


Zamodelowane procesy w Visual Studio

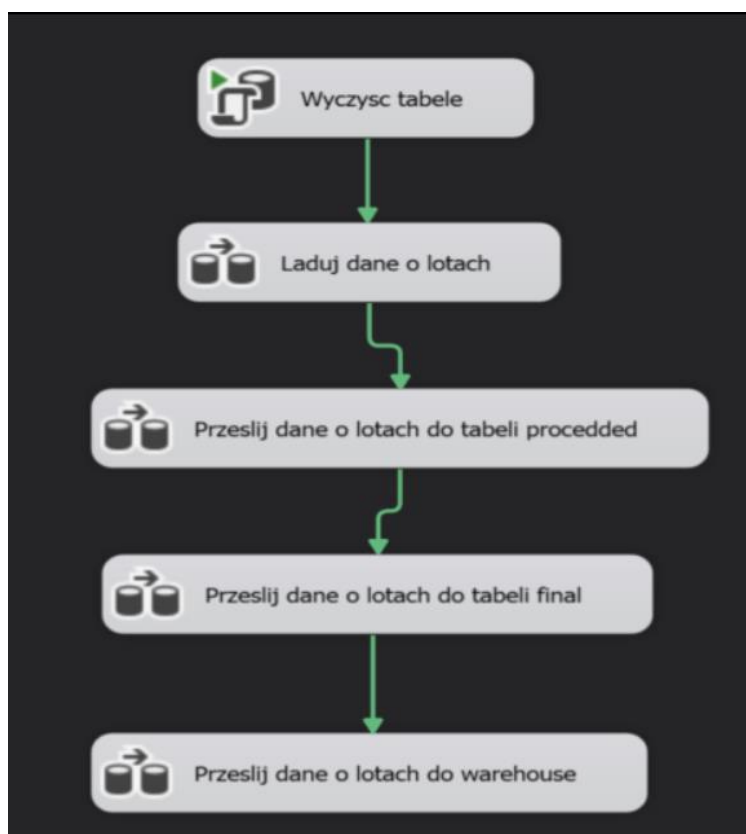
Ładowanie tabel wymiarów



Ładowanie tabeli daty i czasu



Ładowanie tabeli faktów



Kostki danych

Poniżej załączamy przykładowe analizy wykonane z użyciem kostki OLAP

AIRLINE NAME	DEPARTURE YEAR	Laczny_czas_opoznien
AirTran Airways	2008	50626
Alaska Airlines	2008	76715
Aloha Airlines	2008	-12670
American Airl...	2008	465389
American Eagl...	2008	565412
Atlantic South...	2008	215628
Comair	2008	91050
Continental Al...	2008	115448
Delta Air Lines	2008	121784
ExpressJet	2008	206087
Frontier Airlines	2008	41120
Hawaiian Airlines	2008	-21480
JetBlue Airways	2008	30095
Mesa Airlines	2008	269283
Northwest Airl...	2008	222141
Pinnacle Airlines	2008	211507
SkyWest	2008	671643
Southwest Air...	2008	575916
United Airlines	2008	516699
US Airways	2008	81540

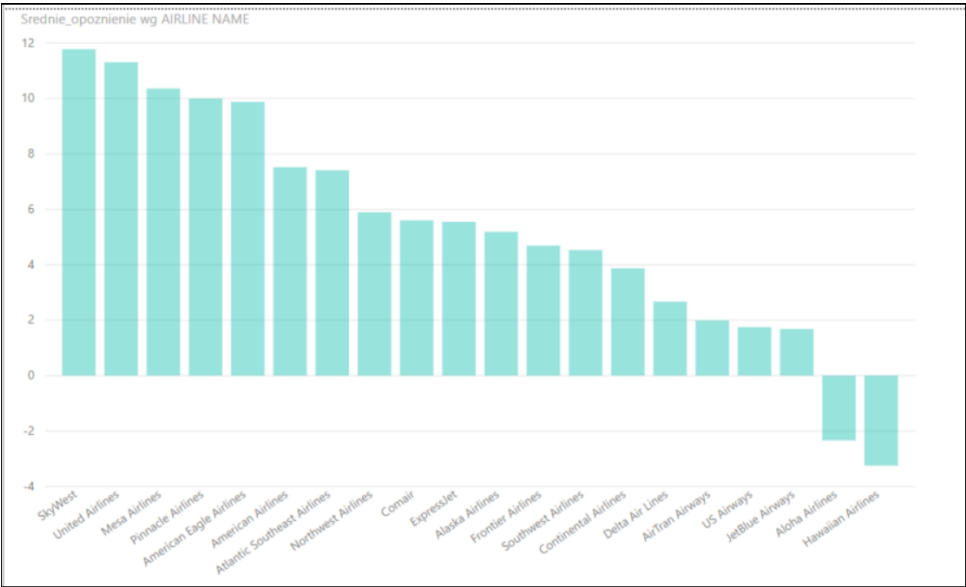
AIRLINE NAME	DEP COUNTRY	Srednie_opoznienie
AirTran Airways	United States	1,99
Alaska Airlines	United States	5,19
Aloha Airlines	United States	-2,34
American Airl...	Puerto Rico	8,98
American Airl...	United States	7,53
American Airl...	Virgin Islands	-2,37
American Eagl...	United States	9,88
Atlantic South...	United States	7,41
Comair	United States	5,6
Continental Al...	Puerto Rico	1,06
Continental Al...	United States	3,9
Continental Al...	Virgin Islands	-18
Delta Air Lines	Puerto Rico	-0,46
Delta Air Lines	United States	2,68
Delta Air Lines	Virgin Islands	13,5
ExpressJet	United States	5,55
Frontier Airlines	United States	4,69
Hawaiian Airlines	United States	-3,25
JetBlue Airways	Puerto Rico	-2,16
JetBlue Airways	United States	1,81
Mesa Airlines	United States	10,36
Northwest Airl...	Puerto Rico	3,3
Northwest Airl...	United States	5,89

DEPARTURE CITY	AIRLINE NAME	FACT FLIGHT Count
Ablene	American Eagl...	204
Aguadilla	Continental Al...	31
Aguadilla	JetBlue Airways	93
Akron	AirTran Airways	216
Akron	Atlantic South...	157
Akron	Frontier Airlines	30
Akron	Mesa Airlines	53
Akron	Pinnacle Airlines	57
Akron	SkyWest	16
Albany	American Eagl...	81
Albany	Atlantic South...	163
Albany	ExpressJet	145
Albany	Mesa Airlines	22
Albany	Northwest Airl...	46
Albany	Pinnacle Airlines	26
Albany	Southwest Air...	306
Albany	United Airlines	37
Albany	US Airways	90
Albuquerque	American Airl...	210
Albuquerque	Continental Al...	68
Albuquerque	Delta Air Lines	88
Albuquerque	ExpressJet	74
Albuquerque	Frontier Airlines	5

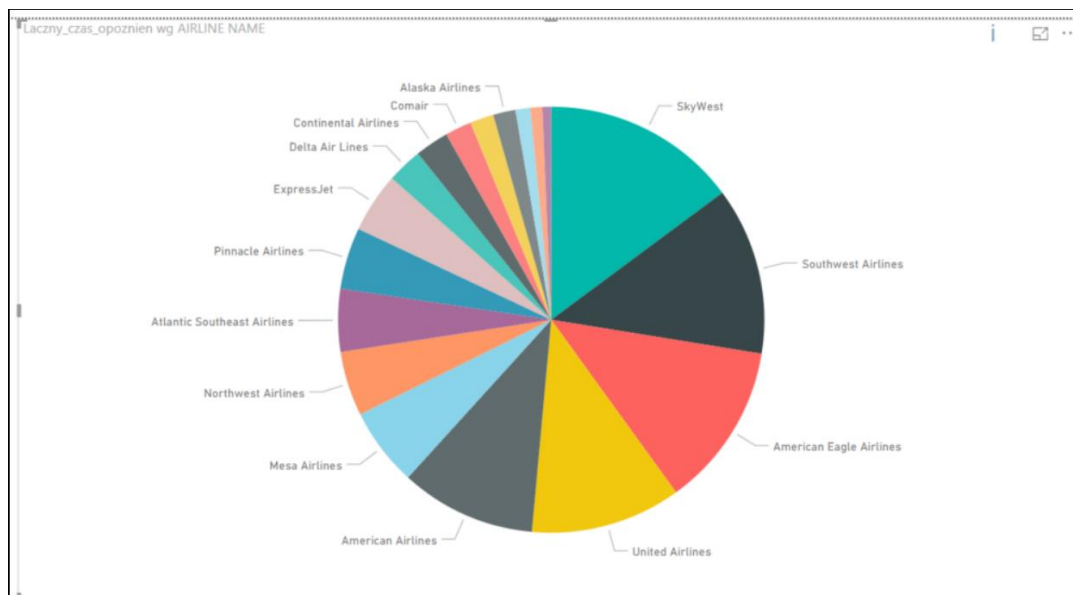
Raporty Power BI

Poniżej załączamy zrzuty ekranu przykładowych raportów wygenerowanych z użyciem narzędzia Power BI

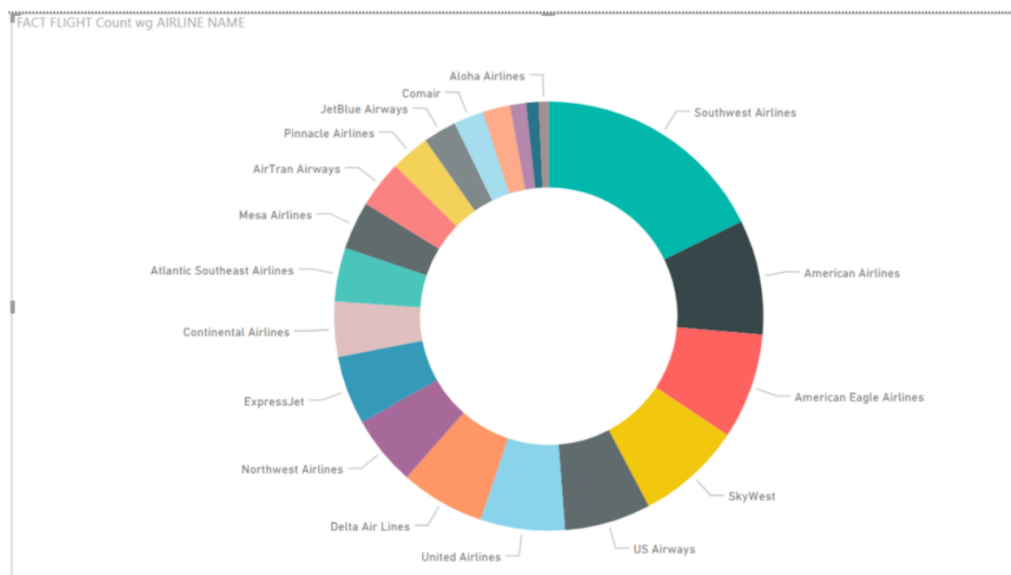
Wykres zależności średniego opóźnienia od linii lotniczej



Wykres przedstawiający łączny czas opóźnień dla każdej linii lotniczej



Wykres ukazujący zależność łącznej liczby lotów dla każdej z linii lotniczych



Wnioski i podsumowanie

Celem stworzenia hurtowni danych jest wspomoczenie działalności organizacji. Ów efekt osiąga się poprzez analizę wycinka rzeczywistości pod kątem interesujących nas aspektów. Dane zewnętrzne potrzebne do utworzenia hurtowni najczęściej pochodzą z wielu różnych źródeł (bazy danych, plików z danymi itp.), więc proces ładowania danych wymaga zintegrowania i ujednolicenia danych pod kątem hurtowni.

W naszym przypadku stworzenie hurtowni danych miało pomóc potencjalnemu klientowi w analizie transportów lotniczych w USA. Zdobyta przez niego wiedza pozwoliłaby na ocenę jakości i sprawności świadczonych przez poszczególnych przewoźników usług transportowych. Dawałoby to olbrzymie możliwości, np. usprawnienie połączeń między lotniskami czy też poprawę jakości świadczonych przez danego przewoźnika usług. Aby tego dokonać należałoby udostępnić klientowi raporty zawierające požądane przez niego dane.