

Politechnika Warszawska
Wydział Elektroniki i Technik Informacyjnych

ANALITYCZNE BAZY DANYCH
PROJEKT CZĘŚĆ 2

HURTOWNIA DANYCH

Warszawa 2024

Spis treści

1. Cel projektu	2
2. Fakt 1	2
2.1. Modelowanie hurtowni danych	2
2.1.1. Schemat	2
2.1.2. Schemat w Toad Data Modeler	3
2.1.3. Wymiary	3
2.1.4. Tabela faktów	5
2.2. Modelowanie ODS	5
2.3. Opracowanie mechanizmu przyrostowego względem źródeł danych	6
2.4. Opracowanie ETL do ODS	9
2.5. Opracowanie ETL do DW	11
3. Fakt 2 - akcja marketingowa	12
3.1. Modelowanie hurtowni danych	12
3.1.1. Schemat	12
3.1.2. Schemat w Toad Modeler	13
3.1.3. Wymiary	13
3.1.4. Tabela faktów	15
3.2. Modelowanie ODS i stałego wymiaru czasu	16
3.2.1. Schemat w Toad Modeler	16
3.3. Opracowanie mechanizmu przyrostowego względem źródeł danych	16
3.4. Opracowanie ETL do ODS	18
3.5. Opracowanie ETL do DW	21
3.5.1. Przykład ładowania hurtowni	22

1. Cel projektu

Celem projektu było opracowanie oraz zaimplementowanie hurtowni danych przy użyciu narzędzi takich jak:

- Qesta Toad Data Modeler 7.2,
- SQL Server Management Studio 19,
- Visual Studio 2019 plus SSDT.

Hurtownia ma zapewnić co najmniej możliwość gromadzenia i analizowania danych dotyczących sprzedaży i obłożenia atrakcji, jak też kosztów pracowniczych i przeprowadzanych akcji marketingowych, w tym ich kosztów.

2. Fakt 1

Jako temat hurtowni danych obrany został bilet, a konkretnie - sprzedaż biletów, dlatego faktem 1. został bilet (FaktBilet). Dobranymi jego wymiarami są wymiary: atrakcji (dimAtrakcje) oraz czasu (dimCzas). Poziom ziarnistości zbieranych danych:

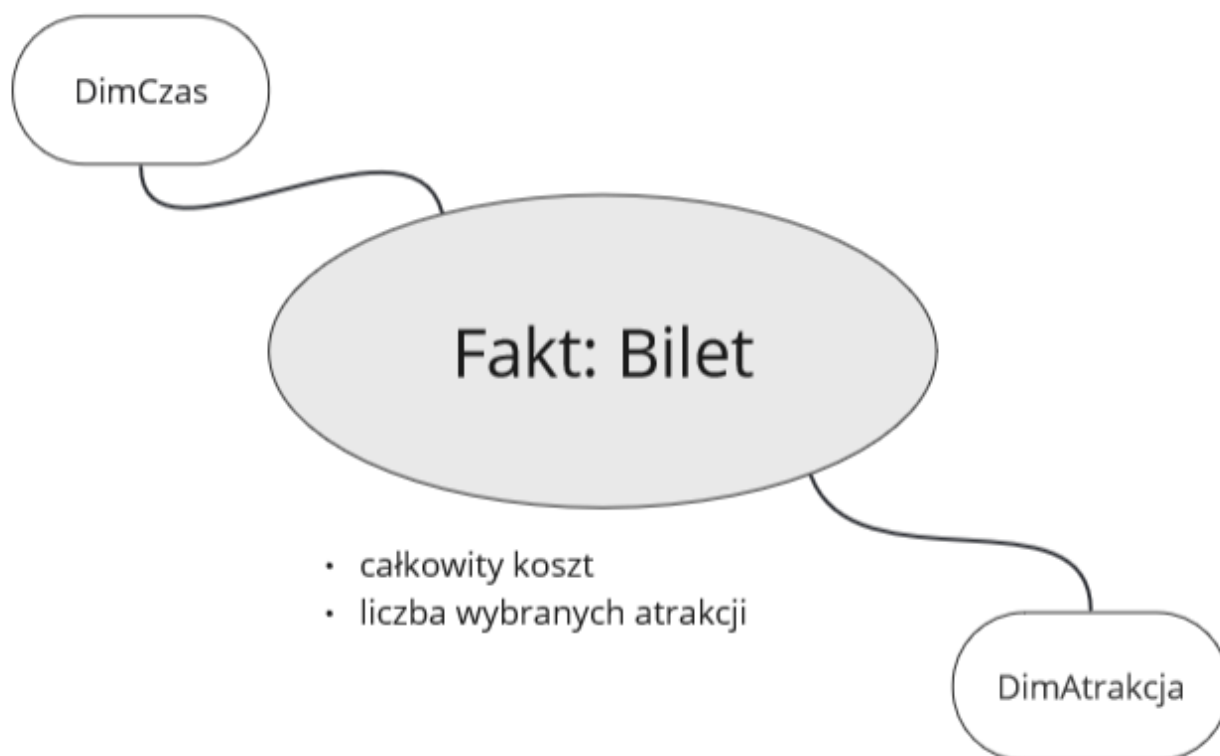
- atrakcje - na poziomie pojedynczej atrakcji w konkretnym parku,
- czas - na poziomie dziennym.

Wymiary mają atrybuty o hierarchiach opisanych w [tabelach 1-3](#). Zdefiniowanymi miarami są całkowity koszt biletów oraz liczba atrakcji wybranych na bilecie.

2.1. Modelowanie hurtowni danych

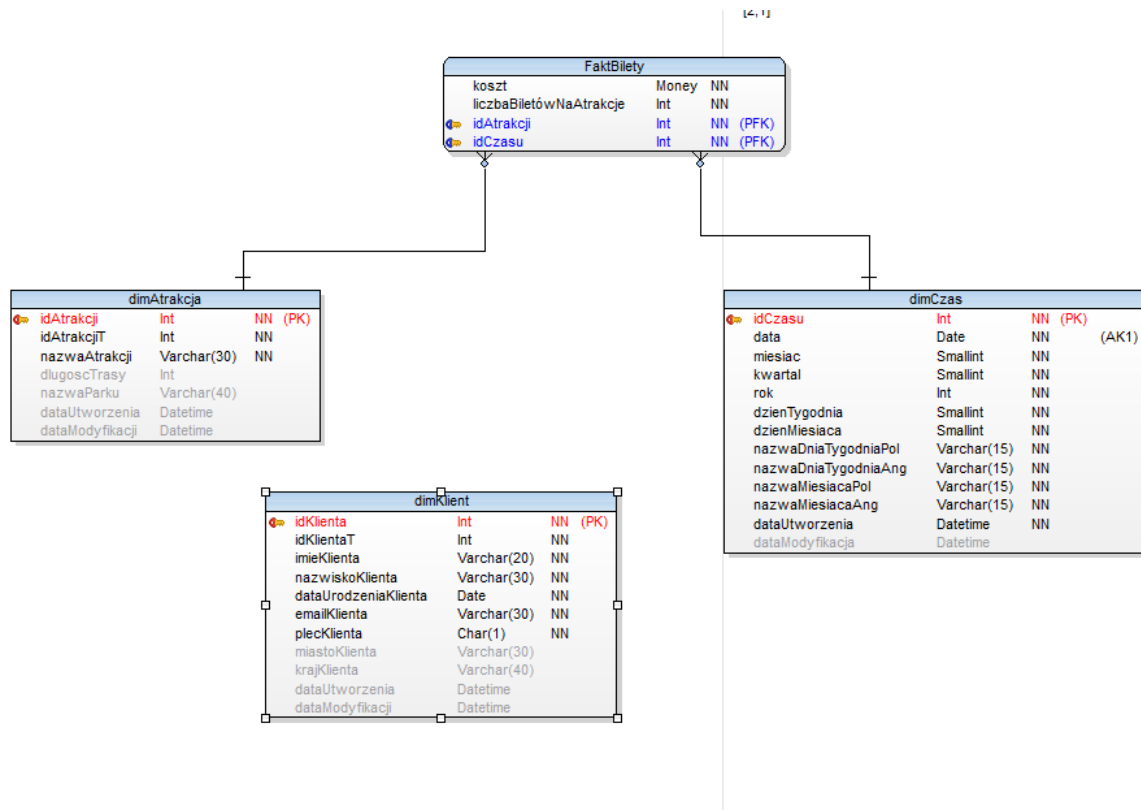
Na początkowym etapie projektowania hurtowni założono, że jednym z wymiarów tabeli faktów Bilet będzie dimKlient. Takie ustawienie jednak okazało się bardzo trudne w implementacji, zdecydowaliśmy się więc zostawić tylko 2 wymiary - dimAtrakcje oraz dimCzas.

2.1.1. Schemat



Rysunek 1: Schemat tematu.

2.1.2. Schemat w Toad Data Modeler



Rysunek 2: Schemat tabeli faktów Bilet.

2.1.3. Wymiary

Atrybut	Typ i dziedzina	Czy obowiązkowy?	Opis
idKlienta	Int	Tak	Unikatowy numer klienta
idKlientaT	Int	Tak	Pomocniczy numer klienta
imieKlienta	Varchar(20)	Tak	Imię klienta
nazwiskoKlienta	Varchar(30)	Tak	Nazwisko klienta
dataUrodzeniaKlienta	Date	Tak	Data urodzenia klienta
emailKlienta	Varchar(30)	Tak	Adres E-mail klienta
plecKlienta	plecD	Nie	Płeć klienta (wartość ze zbioru: {'K', 'M'})
miastoKlienta	Varchar(20)	Tak	Miasto zamieszkania klienta
krajKlienta	Varchar(20)	Tak	Kraj zamieszkania klienta

Tabela 1: dimKlient (nieużyte)

Atrybut	Typ i dziedzina	Czy obowiązkowy?	Opis
idAtrakcji	Int	Tak	Unikatowy numer atrakcji
idAtrakcjiT	Int	Tak	Pomocniczy numer atrakcji
nazwaAtrakcji	Varchar(30)	Tak	Nazwa atrakcji
dlugoscTrasy	Int	Nie	Długość trasy (jeśli atrakcja to np. <i>rollercoaster</i>)
nazwaParku	Varchar(30)	Tak	Nazwa parku, do którego należy atrakcja
dataUtworzenia	Datetime	Tak	Data utworzenia rekordu
dataModyfikacji	Datetime	Tak	Data ostatniej modyfikacji rekordu

Tabela 2: dimAtrakcja

Atrybut	Typ i dziedzina	Czy obowiązkowy?	Opis
idCzasu	Integer	Tak	Unikatowy numer czasu
data	Date	Tak	Data
miesiac	SmallInt	Tak	Numer miesiąca
kwartal	SmallInt	Tak	Numer kwartału
rok	Int	Tak	Rok
dzienTygodnia	SmallInt	Tak	Numer dnia tygodnia
dzienMiesiaca	SmallInt	Tak	Numer dnia w miesiącu
nazwaDniaTygodniaPol	Varchar(15)	Tak	Nazwa dnia tygodnia (po polsku)
nazwaDniaTygodniaAng	Varchar(15)	Tak	Nazwa dnia tygodnia (po angielsku)
nazwaMiesiacaPol	VarChar(15)	Tak	Nazwa miesiąca (po polsku)
nazwaMiesiacaAng	VarChar(15)	Tak	Nazwa miesiąca (po angielsku)
dataUtworzenia	Datetime	Tak	Data utworzenia rekordu
dataModyfikacji	Datetime	Nie	Data ostatniej modyfikacji rekordu

Tabela 3: dimCzas

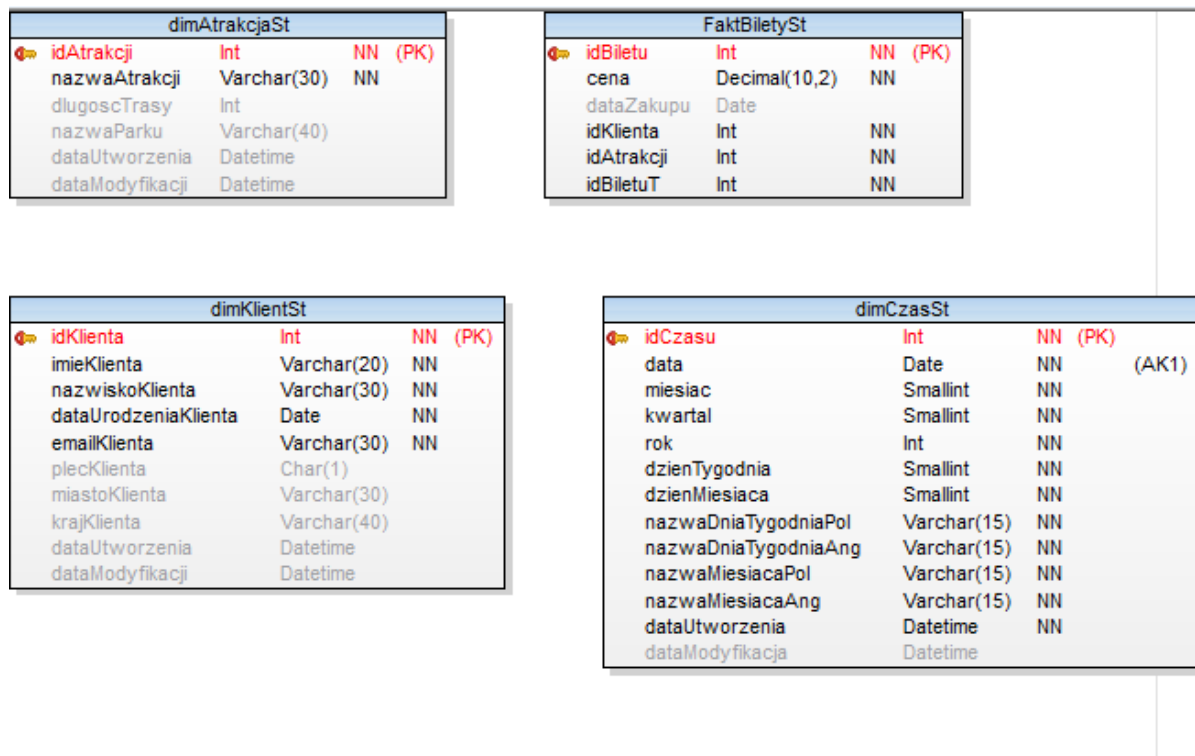
2.1.4. Tabela faktów

Atrybut	Typ i dziedzina	Czy obowiązkowy?	Opis
idCzasu	Int	Tak	Unikatowy identyfikator czasu
idKlienta	Int	Tak	Unikatowy identyfikator klienta
idAtrakcji	Int	Tak	Unikatowy identyfikator atrakcji
koszt	Money	Tak	Koszt biletu
liczbaWybranychAtrakcji	Int	Tak	Liczba atrakcji wybranych na bilecie
dataZakupu	Datetime	Tak	Data zakupu biletów

Tabela 4: FaktBilety

2.2. Modelowanie ODS

ODS - magazyn danych operacyjnych, warstwa znajdująca się pomiędzy źródłami a centralną hurtownią danych. To do niej najpierw ładowane są dane, zanim załadowane zostaną do hurtowni. Warstwa ta umożliwia "odciążenie" głównej hurtowni danych z części zadań związanych z aktualizacją danych.



Rysunek 3: Schemat ODS.

Stały wymiar czasu został załadowany za pomocą skryptu SQL na okres 10 lat (2022-2032) i jest to tabela wymiaru która jest współdzielona między dwoma faktami.

SQLQuery1.sql - 1...DS (ABDGR06 (144))

```

SELECT TOP (1000) [idCzasu]
, [data]
, [miesiac]
, [kwartal]
, [rok]
, [dzienTygodnia]
, [dzienMiesiaca]
, [nazwaDniaTygodniaPol]
, [nazwaDniaTygodniaAng]
, [nazwaMiesiacaPol]
, [nazwaMiesiacaAng]
, [dataUtworzenia]
, [dataModyfikacja]
FROM [ABD_GR06_ODS].[dbo].[dimCzasSt]

```

	idCzasu	data	miesiac	kwartal	rok	dzienTygodnia	dzienMiesiaca	nazwaDniaTygodniaPol	nazwaDniaTygodniaAng	nazwaMiesiacaPol	nazwaMiesiacaAng	dataUtworzenia	dataModyfikacja
1	20220101	2022-01-01	1	1	2022	6	1	Sobota	Saturday	Styczeń	January	2024-04-03 18:00:24.500	NULL
2	20220102	2022-01-02	1	1	2022	7	2	Niedziela	Sunday	Styczeń	January	2024-04-03 18:00:24.500	NULL
3	20220103	2022-01-03	1	1	2022	1	3	Poniedziałek	Monday	Styczeń	January	2024-04-03 18:00:24.503	NULL
4	20220104	2022-01-04	1	1	2022	2	4	Wtorek	Tuesday	Styczeń	January	2024-04-03 18:00:24.507	NULL
5	20220105	2022-01-05	1	1	2022	3	5	Środa	Wednesday	Styczeń	January	2024-04-03 18:00:24.507	NULL
6	20220106	2022-01-06	1	1	2022	4	6	Czwartek	Thursday	Styczeń	January	2024-04-03 18:00:24.510	NULL
7	20220107	2022-01-07	1	1	2022	5	7	Piątek	Friday	Styczeń	January	2024-04-03 18:00:24.510	NULL
8	20220108	2022-01-08	1	1	2022	6	8	Sobota	Saturday	Styczeń	January	2024-04-03 18:00:24.513	NULL
9	20220109	2022-01-09	1	1	2022	7	9	Niedziela	Sunday	Styczeń	January	2024-04-03 18:00:24.517	NULL
10	20220110	2022-01-10	1	1	2022	1	10	Poniedziałek	Monday	Styczeń	January	2024-04-03 18:00:24.520	NULL
11	20220111	2022-01-11	1	1	2022	2	11	Wtorek	Tuesday	Styczeń	January	2024-04-03 18:00:24.520	NULL
12	20220112	2022-01-12	1	1	2022	3	12	Środa	Wednesday	Styczeń	January	2024-04-03 18:00:24.523	NULL

Rysunek 4: Tabela wymiaru stałego Czas

2.3. Opracowanie mechanizmu przyrostowego względem źródeł danych

Mechanizm przyrostowy (ang. *incremental mechanism*) to podejście w systemach przetwarzania danych, które polega na identyfikowaniu i ładowaniu tylko tych rekordów, które uległy zmianie od czasu ostatniego przetwarzania, zamiast ładować całą zawartość źródłowej bazy danych lub tabeli przy każdym cyklu przetwarzania. Mechanizm przyrostowy analizuje zmiany (np. nowe lub zaktualizowane rekordy) i w odpowiedni sposób aktualizuje docelową bazę danych, magazyn danych lub inną formę systemu.

Mechanizm przyrostowy ma na celu zmniejszenie obciążenia systemu poprzez minimalizację zbędnego przetwarzania danych, szczególnie w przypadku dużych zbiorów danych. Dzięki temu można oszczędzić zasoby obliczeniowe, czas i przepustowość sieci, minimalizując jednocześnie opóźnienia i czas potrzebny na aktualizację danych.

W celu utworzenia mechanizmu przyrostowego względem źródeł danych, do każdego wymiaru zostały dodane pola `dataUtworzenia` oraz `dataModyfikacji` typu `Datetime`. Pole `dataUtworzenia` przechowuje informacje o czasie utworzenia danego rekordu w bazie danych. Można dzięki temu śledzić, kiedy dane zostały wprowadzone do systemu, co może być przydatne do analizy danych historycznych. Analogicznie, `dataModyfikacji` wskazuje na czas ostatniej modyfikacji danego rekordu. Umożliwia to monitorowanie zmian w danych oraz identyfikację czasu aktualizacji danych. Dzięki temu możemy dowiedzieć się, kiedy dane zostały zmienione i załadować je od nowa.

Dodanie tych atrybutów w bazie SRC1:

SQL Query:

```
SELECT TOP (1000) [idAtrakcji]
, [nazwaAtrakcji]
, [dlugoscTrasy]
, [nazwaParku]
, [dataUtworzenia]
, [dataModyfikacji]
FROM [ABD_GR06_ODS].[dbo].[dimAtrakcjaSt]
```

100 %

Results Messages

	idAtrakcji	nazwaAtrakcji	dlugoscTrasy	nazwaParku	dataUtworzenia	dataModyfikacji
1	1	Karuzela	NULL	Park Rozrywki AA	2024-04-13 20:32:29.720	2024-04-14 00:29:58.370
2	2	Wodny Park Rozrywki	NULL	Park Rozrywki B	2024-04-13 20:32:29.737	2024-04-14 00:29:58.370
3	3	Roller Coaster	900	Park Rozrywki C	2024-04-13 20:32:29.750	2024-04-14 20:53:19.850
4	4	Strzała Wodna	NULL	Park Rozrywki D	2024-04-13 20:32:29.763	2024-04-14 00:29:58.370
5	5	Koło Widokowe	NULL	Park Rozrywki AA	2024-04-13 20:32:29.773	2024-04-14 00:29:58.370
6	6	Tajemniczy Labirynt	NULL	Park Rozrywki B	2024-04-13 20:32:29.787	2024-04-14 00:29:58.370
7	7	Ekstremalny Zjazd	555	Park Rozrywki C	2024-04-13 20:32:29.803	2024-04-14 00:29:58.370
8	8	Karaoke Bar	NULL	Park Rozrywki D	2024-04-13 20:32:29.817	2024-04-14 00:29:58.370
9	9	Skocznia Bungee	NULL	Park Rozrywki AA	2024-04-13 20:32:29.840	2024-04-14 00:29:58.370
10	10	Kraj Dinozaurów	NULL	Park Rozrywki B	2024-04-13 20:32:29.857	2024-04-14 00:29:58.370
11	11	Wahadło	NULL	Park Rozrywki C	2024-04-13 20:32:29.867	2024-04-14 00:29:58.370
12	12	Plac BBQ	NULL	Park Rozrywki D	2024-04-13 20:32:29.880	2024-04-14 00:29:58.370
13	13	Quad Safari	500	Park Rozrywki AA	2024-04-13 20:32:29.897	2024-04-14 00:29:58.370
14	14	Pokazy Akrobatyczne	NULL	Park Rozrywki B	2024-04-13 20:32:29.910	2024-04-14 00:29:58.370
15	15	Symulator Lotu	NULL	Park Rozrywki C	2024-04-13 20:32:29.923	2024-04-14 00:29:58.370
16	16	Park Linowy	NULL	Park Rozrywki D	2024-04-13 20:32:29.933	2024-04-14 00:29:58.370
17	17	Teatr na Żywo	NULL	Park Rozrywki AA	2024-04-13 20:32:29.947	2024-04-14 00:29:58.370
18	18	Mini ZOO	NULL	Park Rozrywki B	2024-04-13 20:32:29.967	2024-04-14 00:29:58.370
19	19	Paintball	NULL	Park Rozrywki C	2024-04-13 20:32:29.990	2024-04-14 00:29:58.370
20	20	Dmuchany Zamek	NULL	Park Rozrywki D	2024-04-13 20:32:30.000	2024-04-14 00:29:58.370
21	21	Safari Afrykańskie	NULL	Park Rozrywki AA	2024-04-13 20:32:30.013	2024-04-14 00:29:58.370
22	22	Wodna Rozrywka	NULL	Park Rozrywki B	2024-04-13 20:32:30.033	2024-04-14 00:29:58.370
23	23	Strzelnica	NULL	Park Rozrywki C	2024-04-13 20:32:30.047	2024-04-14 00:29:58.370
24	24	Park Edukacyjny	NULL	Park Rozrywki D	2024-04-13 20:32:30.060	2024-04-14 00:29:58.370

Rysunek 5: Data utworzenia i data modyfikacji.

Przykład działania mechanizmu przyrostu względem źródeł danych:

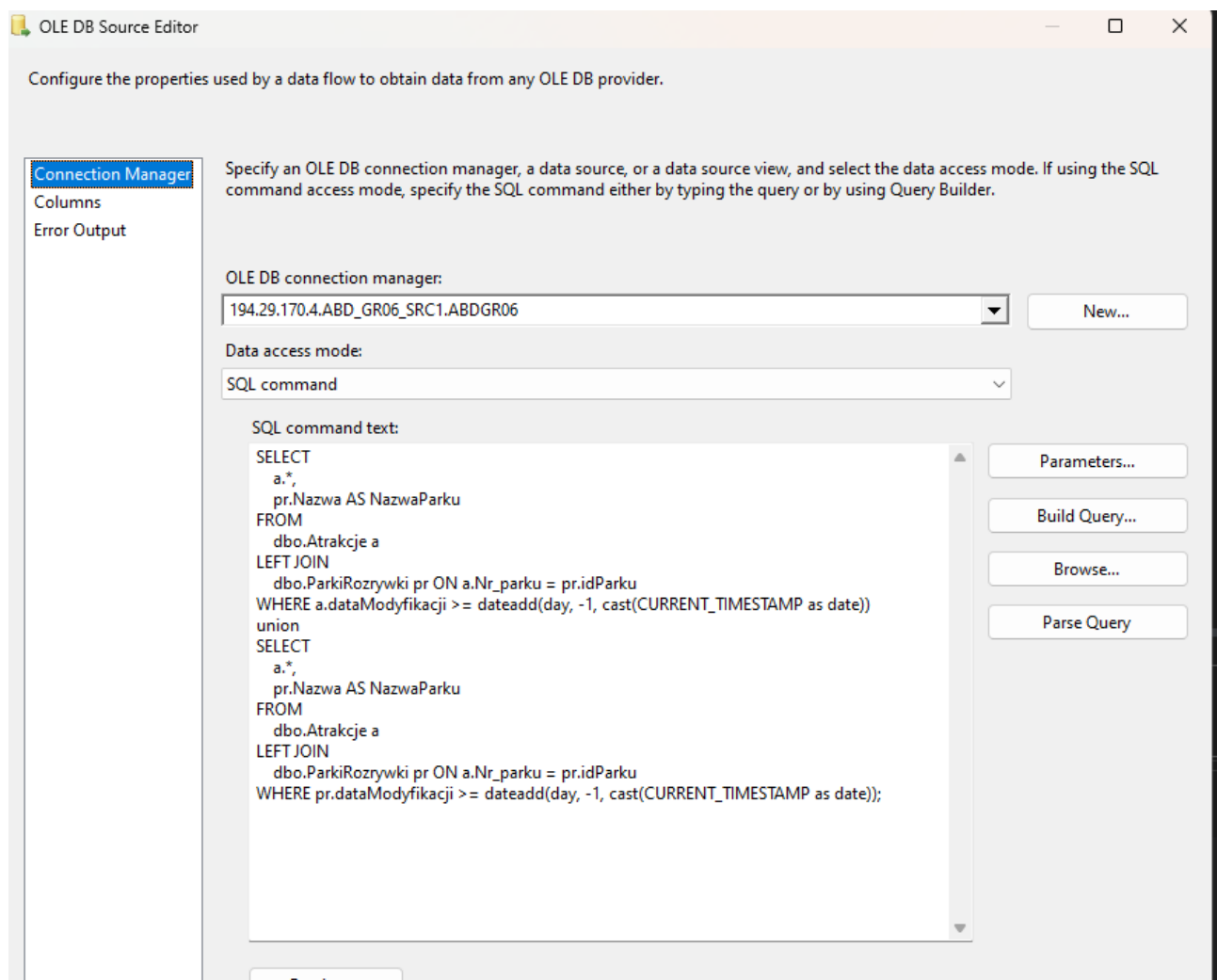
SELECT TOP (1000) [idCzasu],
[idAtrakcji],
[liczbaBiletówNaAtrakcje],
[koszt]
FROM [ABD_GR06_DW].[dbo].[FaktBilety]

	idCzasu	idAtrakcji	liczbaBiletówNaAtrakcje	koszt
1	20230701	15	4	603,00
2	20230701	24	1	150,75
3	20230701	29	1	150,75
4	20230701	35	1	150,75
5	20230701	37	1	150,75
6	20230701	38	1	150,75
7	20230810	7	1	120,25
8	20230810	11	1	120,25
9	20230810	18	1	120,25
10	20230810	21	2	240,50
11	20230810	31	2	240,50
12	20230810	32	2	240,50

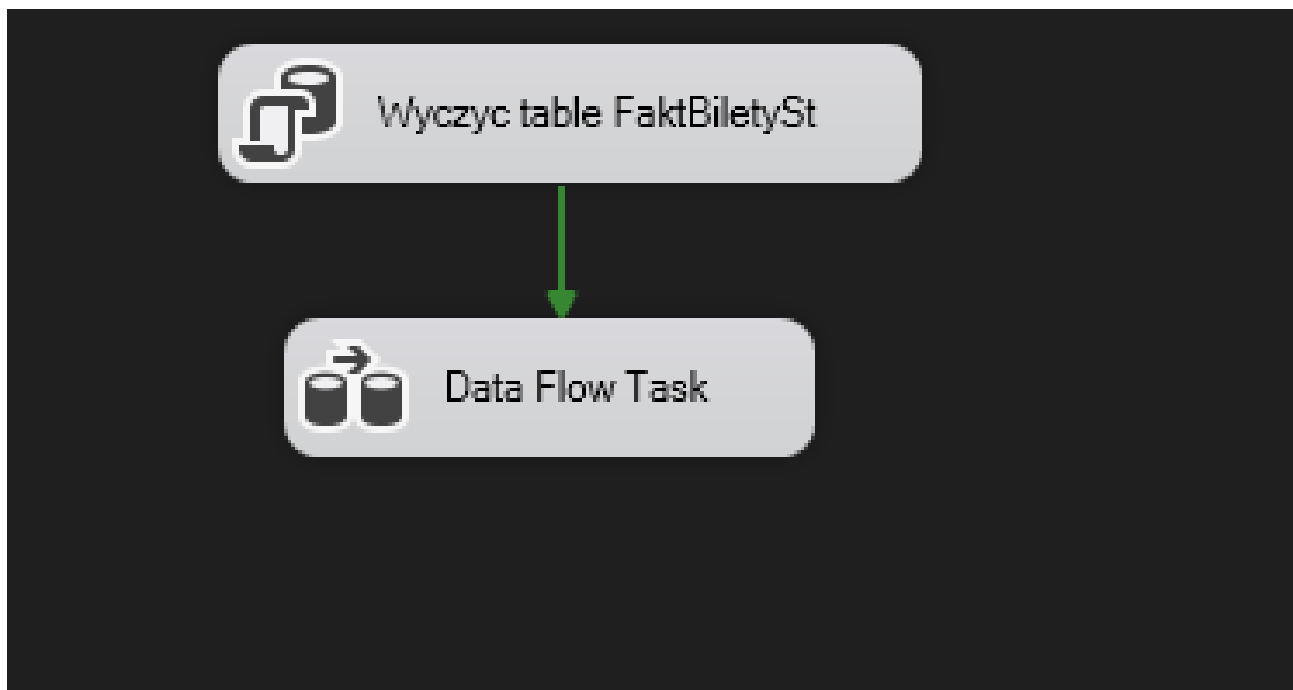
Rysunek 6: Przykład działania na tabeli faktów Bilet - przed zmianą rekordu nr 2 (po lewej) oraz po zmianie ilości sprzedanych biletów (po prawej).

2.4. Opracowanie ETL do ODS

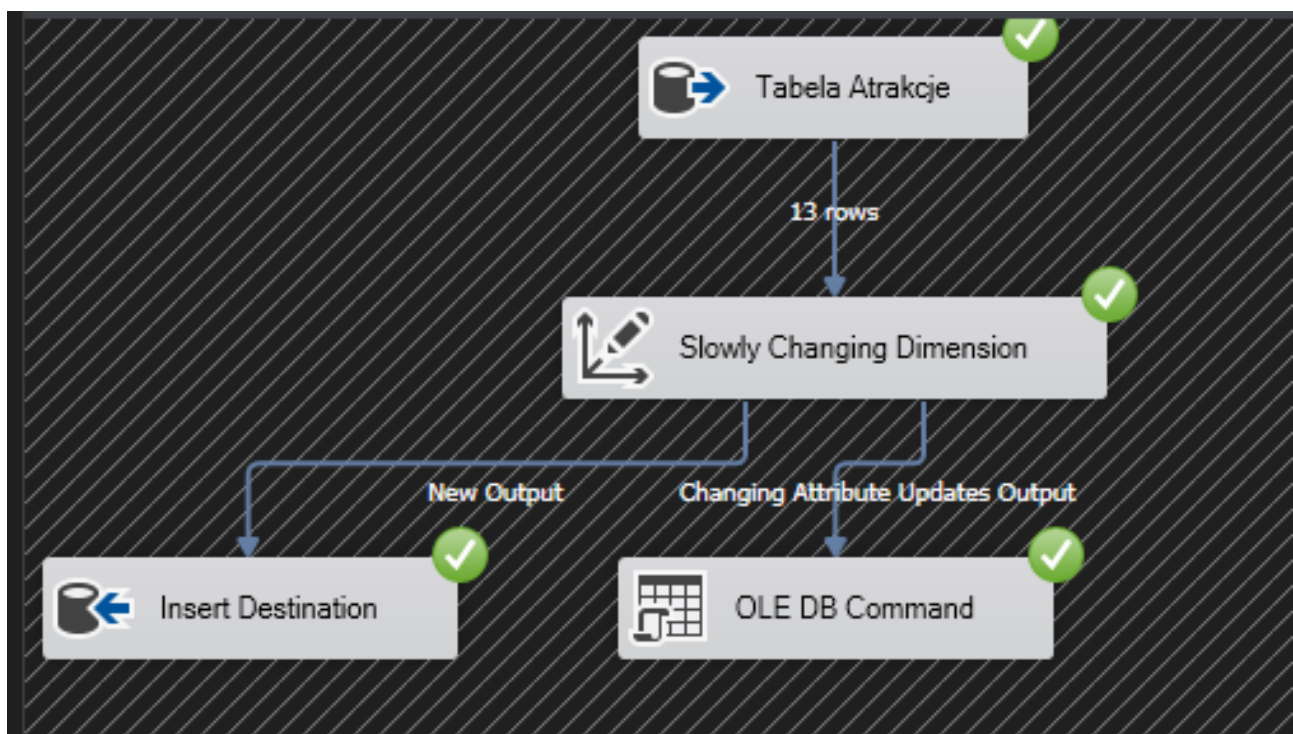
ODS znajduje się w warstwie pomiędzy źródłem a hurtownią danych, zatem dane do niego ładowane są bezpośrednio ze źródła. Za ich ładowanie odpowiada ETL, którego zaprojektowanie jest bardzo ważnym krokiem w projektowaniu systemów hurtowni danych.



Rysunek 7: Ładowanie danych do ODS - polecenie SQL.



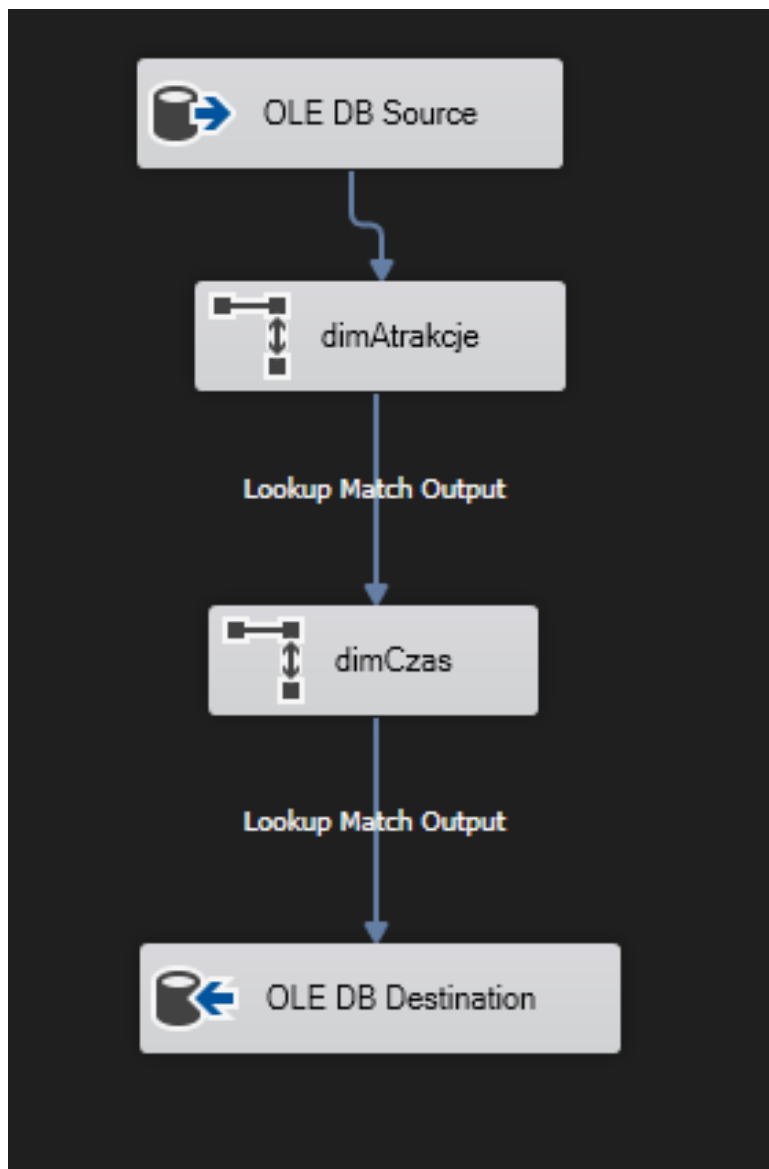
Rysunek 8: Czyszczenie tabeli faktów Bilet przed przesłaniem zmian.



Rysunek 9: Proces ładowania ODS w Visual Studio - poprawne załadowanie.

2.5. Opracowanie ETL do DW

DW, czyli sama hurtownia danych, także musi skądś otrzymywać dane. Do tego celu ponownie używa się narzędzia ETL, aby transportowało dane z warstwy pośredniej ODS do DW. Poniższy rysunek obrazuje schemat działania opracowanego narzędzia:



Rysunek 10: Schemat działania ETL do DW.

3. Fakt 2 - akcja marketingowa

W przypadku Faktu dotyczącego bazy danych Akcji Marketingowej uznano za tabele faktów samą Akcję Marketingową i stworzono 4 tabele wymiarów: dimKanal, dimCzas, dimUczestnik oraz dimAkcja. Głównymi miarami do analizy w tabeli faktów jest całkowity koszt oraz liczba uczestników jako że są to atrybuty sumowalne i posłużą nam pod kątem analizy w kolejnych etapach. Analiza ma dotyczyć kosztów akcji oraz też liczby uczestników w zależności od pory roku czy innych czynników.

3.1. Modelowanie hurtowni danych

3.1.1. Schemat



Rysunek 11: Schemat tematu.

3.1.2. Schemat w Toad Modeler



Rysunek 12: Schemat hurtowni danych

3.1.3. Wymiary

Tabela 5: dimUczestnik

Atrybut	Typ i dziedzina	Czy obowiązkowy?	Opis
idUczestnika	Integer	Tak	Unikatowy numer uczestnika akcji
idUczestnikaT	Integer	Tak	Pomocniczy numer uczestnika akcji
imieUczestnika	VarChar(20)	Tak	Imię uczestnika akcji
nazwiskoUczestnika	VarChar(30)	Tak	Nazwisko uczestnika akcji
emailUczestnika	VarChar(50)	Tak	Email uczestnika
dataUrodzenia	Date	Tak	Data urodzenia uczestnika
dataUtworzenia	DateTime	Tak	Data utworzenia wiersza
dataModyfikacji	DateTime	Nie	Data modyfikacji wiersza

Tabela 6: dimAkcja

Atrybut	Typ i dziedzina	Czy obowiązkowy?	Opis
idAkcji	Integer	Tak	Unikatowy numer akcji
idAkcjiT	Integer	Tak	Pomocniczy numer akcji
statusAkcji	VarChar(20)	Tak	Aktualny status akcji
calkowityKoszt	Decimal(20,6)	Tak	Całkowity koszt akcji
nazwaAkcji	Varchar(25)	Tak	Nazwa akcji
dataRozpoczecia	Date	Tak	Data rozpoczęcia akcji
dataZakonczenia	Date	Tak	Data zakończenia akcji
opisAkcji	Varchar(200)	Tak	Opis akcji
targetWiekowy	Varchar(15)	Nie	Target wiekowy akcji
formaAkcji	Varchar(20)	Nie	Forma akcji
sponsorzy	Varchar(50)	Nie	Sponsorzy akcji
dataUtworzenia	DateTime	Tak	Data utworzenia wiersza
dataModyfikacji	DateTime	Nie	Data modyfikacji wiersza

Tabela 7: dimKanal

Atrybut	Typ i dziedzina	Czy obowiązkowy?	Opis
idKanału	Integer	Tak	Unikatowy numer kanału
idKanałuT	Integer	Tak	Pomocniczy numer kanału
typKanału	VarChar(20)	Tak	Typ kanału
opisKanału	Varchar(150)	Tak	Opis kanału
dataUtworzeniaKanału	Date	Tak	Data utworzenia kanału
liczbaOdbiorcow	Integer	Tak	Liczba odbiorców kanału
kosztKanału	Decimal(20,6)	Tak	Całkowity koszt kanału
formaKanału	Varchar(20)	Tak	Forma kanału
nazwaKanału	Varchar(20)	Tak	Nazwa kanału
dataUtworzenia	DateTime	Tak	Data utworzenia wiersza
dataModyfikacji	DateTime	Nie	Data modyfikacji wiersza

Tabela 8: dimCzas

Atrybut	Typ i dziedzina	Czy obowiązkowy?	Opis
idCzasu	Integer	Tak	Unikatowy numer czasu
data	Date	Tak	Data
miesiac	Smallint	Tak	Miesiąc
kwartal	Smallint	Tak	Kwartał
rok	Integer	Tak	Rok
dzienTygodnia	Smallint	Tak	Dzień tygodnia
dzienMiesiaca	Smallint	Tak	Dzień miesiąca
nazwaDniaTygodniaPol	Varchar(15)	Tak	Nazwa dnia tygodnia po polsku
nazwaDniaTygodniaAng	Varchar(15)	Tak	Nazwa dnia tygodnia po angielsku
nazwaMiesiacaPol	Varchar(15)	Tak	Nazwa miesiąca po polsku
nazwaMiesiacaAng	Varchar(15)	Tak	Nazwa miesiąca po angielsku
dataUtworzenia	DateTime	Tak	Data utworzenia wiersza
dataModyfikacji	DateTime	Nie	Data modyfikacji wiersza

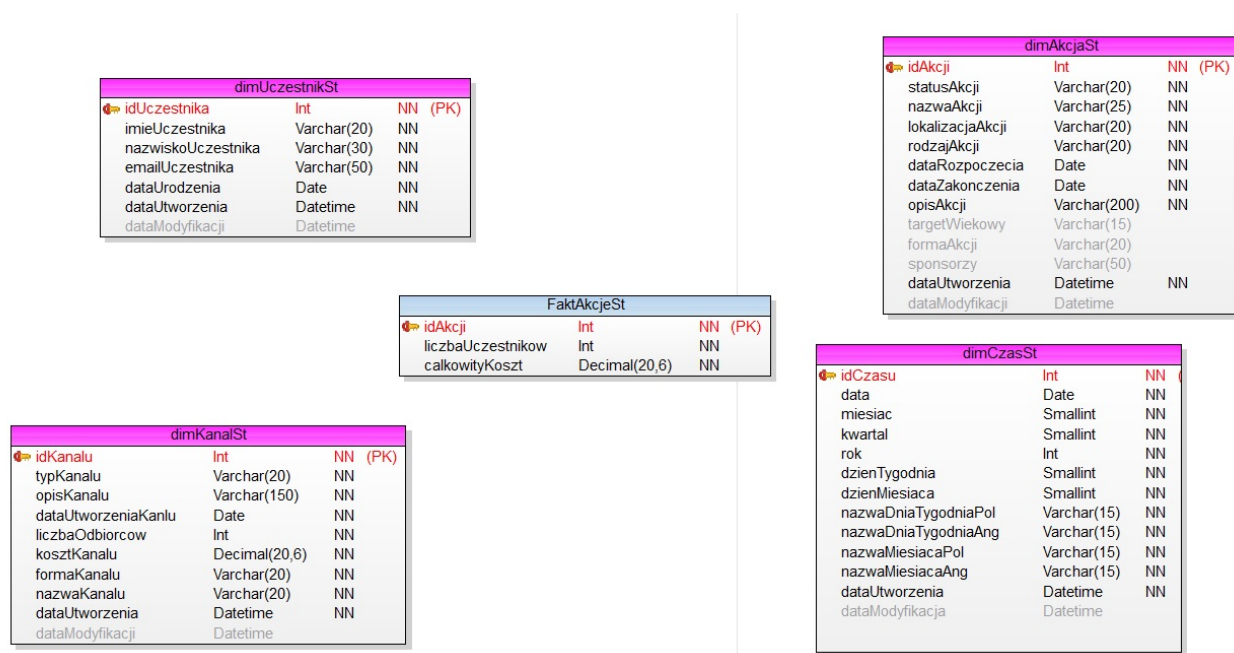
3.1.4. Tabela faktów

Tabela 9: FaktAkcjaMarketingowa

Atrybut	Typ i dziedzina	Czy obowiązkowe?	Opis
idCzasu	Integer	Tak	Klucz obcy unikatowy numer czasu
idAkcji	Integer	Tak	Klucz obcy unikatowy numer akcji
idKanal	Integer	Tak	Klucz obcy unikatowy numer kanału
idUczestnika	Integer	Tak	Klucz obcy unikatowy numer uczestnika
calkowityKoszt	Decimal(20,6)	Tak	Całkowity koszt
liczbaUczestników	Integer	Tak	Całkowita liczba uczestników

3.2. Modelowanie ODS i stałego wymiaru czasu

3.2.1. Schemat w Toad Modeler



Rysunek 13: Schemat ODS

3.3. Opracowanie mechanizmu przyrostowego względem źródeł danych

W celu opracowania mechanizmu przyrostowego względem źródeł danych, do każdego wymiaru zostały dodane pola dataUtworzenia oraz dataModyfikacji z odpowiednimi typami danych.

Pole dataUtworzenia przechowuje informację o dacie i czasie utworzenia danego rekordu w bazie danych. Dzięki temu można śledzić, kiedy dane zostały wprowadzone do systemu, co może być przydatne do analizy danych historycznych.

Z kolei pole dataModyfikacji przechowuje informację o dacie i czasie ostatniej modyfikacji danego rekordu. To umożliwia monitorowanie zmian w danych oraz identyfikację, kiedy i przez kogo dany rekord został zmodyfikowany. Jest to kluczowe dla zachowania spójności i wiarygodności danych w systemie, a także dla śledzenia historii zmian.

Dzięki temu mechanizmowi możliwe będzie skuteczne monitorowanie zmian w danych i identyfikacja ich źródeł oraz historii.

Dodanie atrybutów Data Modyfikacji i Data Utworzenia w bazie src2:

SQLQuery3.sql - 1...C2 (ABDGR06 (190))* SQLQuery2.sql - 1...C2 (ABDGR06 (191))* SQLQuery1.sql - 1...C2 (ABDGR06 (187))

```

SELECT TOP (1000) [idKanal]
, [typKanal]
, [nazwaKanal]
, [dataUtworzeniaKanal]
, [liczbaOdbiorcow]
, [idAkcji]
, [kosztKanal]
, [dataUtworzenia]
, [dataModyfikacji]
FROM [ABD_GR06_SRC2].[dbo].[KanalAkcji]

```

100 %

Results Messages

	idKanal	typKanal	nazwaKanal	dataUtworzeniaKanal	liczbaOdbiorcow	idAkcji	kosztKanal	dataUtworzenia	dataModyfikacji
1	1	Informacyjny	ParkInfotv	2023-03-15	51	3	6000.000000	2024-04-06 13:30:22.467	2024-04-13 17:55:36.917
2	2	Interaktywny	ParkQuest	2023-03-28	100	3	1000.000000	2024-04-06 13:30:22.467	2024-04-06 13:30:22.467
3	3	Reklamowy	AtrakcjeExtra	2024-02-01	100	13	1000.000000	2024-04-06 13:30:22.467	2024-04-06 13:30:22.467
4	4	Społecznościowy	ParkFunCommunity	2023-03-01	150	1	2000.000000	2024-04-06 13:30:22.467	2024-04-12 17:01:39.967

Rysunek 14: Data Utworzenia i data Modyfikacji

Przykład działania mechanizmu przyrostowego względem źródeł danych:

SQLQuery2.sql - 1...DS (ABDGR06 (203)) SQLQuery1.sql - 1...DS (ABDGR06 (179))

```

SELECT TOP (1000) [idKanal]
, [typKanal]
, [opisKanal]
, [dataUtworzeniaKanal]
, [liczbaOdbiorcow]
, [kosztKanal]
, [formaKanal]
, [nazwaKanal]
, [dataUtworzenia]
, [dataModyfikacji]
FROM [ABD_GR06_ODS].[dbo].[dimKanalSt]

```

100 %

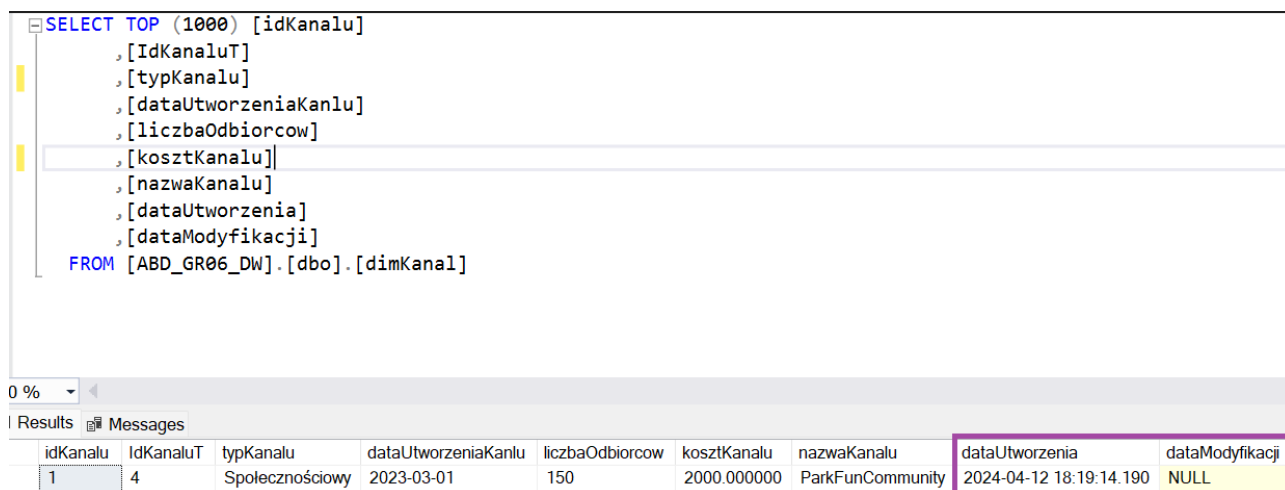
Results Messages

	idKanal	typKanal	opisKanal	dataUtworzeniaKanal	liczbaOdbiorcow	kosztKanal	formaKanal	nazwaKanal	dataUtworzenia	dataModyfikacji
1	4	Społecznościowy	Oficjalny profil społecznościowy parku, gdzie go...	2023-03-01	150	2000.000000	Strona internetowa	ParkFunCommunity	2024-04-06 13:30:22.467	2024-04-12 17:01:39.967

Rysunek 15: Przykład działania na tabeli wymiaru Kanał

W powyższym przykładzie można zauważyć, że do ODS dimKanal został dodany tylko jeden rekord, który uległ modyfikacji. Mechanizm przyrostowy dostarczył pełnej informacji o dacie i godzinie zmiany.

Dodawanie do hurtowni tylko tych rekordów, które uległy zmianie:



The screenshot shows a SQL query window with the following text:

```
SELECT TOP (1000) [idKanal]
, [IdKanalUT]
, [typKanal]
, [dataUtworzeniaKanal]
, [liczbaOdbiorcow]
, [kosztKanal]
, [nazwaKanal]
, [dataUtworzenia]
, [dataModyfikacji]
FROM [ABD_GR06_DW].[dbo].[dimKanal]
```

Below the query, the 'Results' tab is active, displaying a single row of data. The columns are: idKanal, IdKanalUT, typKanal, dataUtworzeniaKanal, liczbaOdbiorcow, kosztKanal, nazwaKanal, dataUtworzenia, and dataModyfikacji. The values for the single row are: 1, 4, Społecznościowy, 2023-03-01, 150, 2000.000000, ParkFunCommunity, 2024-04-12 18:19:14.190, and NULL.

idKanal	IdKanalUT	typKanal	dataUtworzeniaKanal	liczbaOdbiorcow	kosztKanal	nazwaKanal	dataUtworzenia	dataModyfikacji
1	4	Społecznościowy	2023-03-01	150	2000.000000	ParkFunCommunity	2024-04-12 18:19:14.190	NULL

Rysunek 16: Przykład działania na tabeli wymiaru Kanał

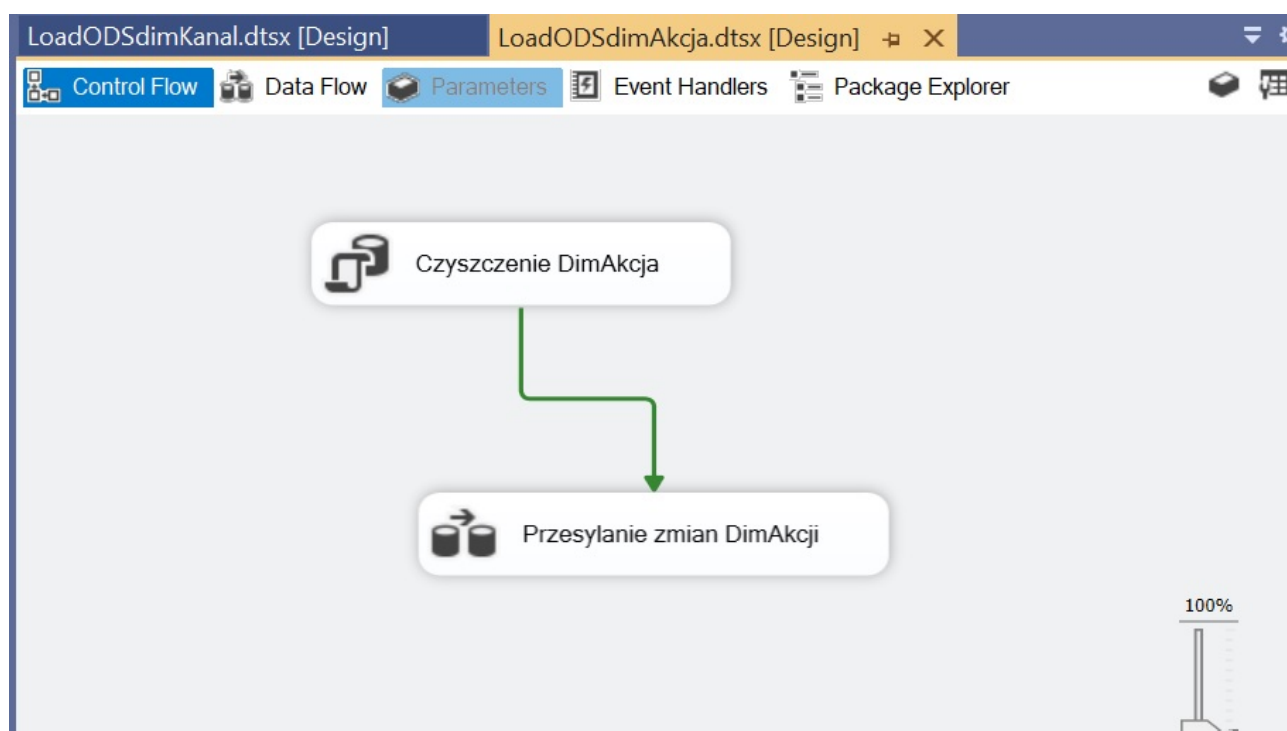
Na powyższym przykładzie można zauważyć że, do hurtowni został dodany tylko jeden rekord. W kolumnie dataUtworzenia podana jest dokładna data dodania rekordu do hurtowni, która jest inna w porównaniu do daty zmiany w źródle, ponieważ do późniejszej analizy interesuje nas, kiedy nowy rekord lub zmiana trafiły do bazy. Wartość dataModyfikacji jest równa NULL ponieważ od datyUtworzenia nie było żadnej modyfikacji.

3.4. Opracowanie ETL do ODS

Opracowanie ETL do ODS stanowi bardzo ważny krok w projektowaniu systemów informatycznych, zwłaszcza tych o złożonym charakterze biznesowym.

ODS pełni istotną rolę jako tymczasowy magazyn danych, który integruje dane z różnych źródeł, transformuje je oraz przechowuje w formie optymalnej do szybkiego dostępu i analizy.

Celem ODS jest zapewnienie jednolitego, zintegrowanego źródła danych, które stanowi solidną podstawę dla procesów decyzyjnych oraz operacyjnych w organizacji. Dzięki niemu odciążą się centralną hurtownię danych i można ją częściej aktualizować.

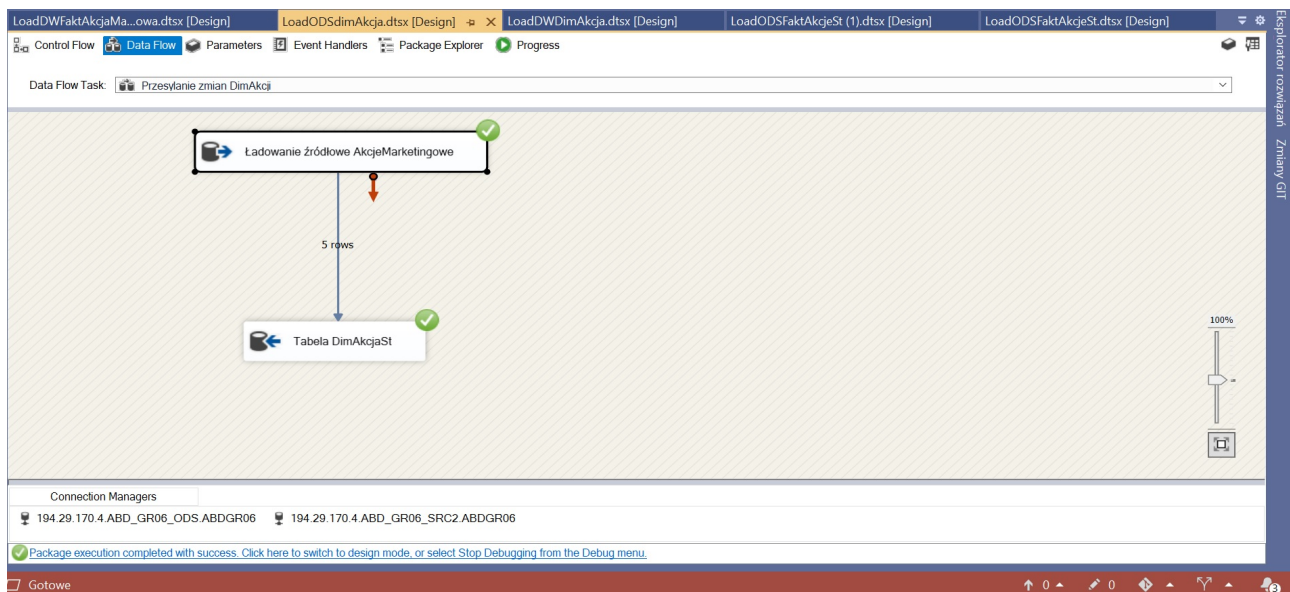


Rysunek 17: Przykład schematu ładowania ODS w Visual Studio

Ładowanie odbywa się w etapach - na początku następuje czyszczenie tabeli za pomocą operacji TRUNCATE a następnie dopiero przesyłanie zmian.

General	
Name	Czyszczenie DimAkcja
Description	Execute SQL Task
Options	
TimeOut	0
CodePage	1250
TypeConversionMode	Allowed
Result Set	
ResultSet	None
SQL Statement	
ConnectionType	OLE DB
Connection	194.29.170.4.ABD_GR06_ODS.ABDGR06
SQLSourceType	Direct input
SQLStatement	TRUNCATE TABLE dimAkcjaSt
IsQueryStoredProcedure	False
BypassPrepare	True

Rysunek 18: Czyszczenie wymiaru przed przesyłaniem zmian



Rysunek 19: Przykład schematu ładowania ODS w Visual Studio

Na powyższym rysunku można zauważyć poprawne załadowanie się tabeli wymiaru dimAkcjaSt.

The screenshot shows a SQL query in the 'SQLQuery1.sql' file. The query is a SELECT statement with 14 columns: idAkcji, statusAkcji, nazwaAkcji, lokalizacjaAkcji, rodzajAkcji, dataRozpoczecia, dataZakonczenia, opisAkcji, targetWiekowy, formaAkcji, sponsorzy, dataUtworzenia, and dataModyfikacji. The results table shows 5 rows of data.

	idAkcji	statusAkcji	nazwaAkcji	lokalizacjaAkcji	rodzajAkcji	dataRozpoczecia	dataZakonczenia	opisAkcji	targetWiekowy	formaAkcji	sponsorzy	dataUtworzenia	dataModyfikacji
1	1	zakonczone	SuperPromo2024	Park rozrywki	Promocja na bilety	2023-04-01	2023-04-15	Wielka promocja na bilety!	18-45	Stagionama	Brak	2024-04-06 13:30:22.380	2024-04-13 17:56:09.143
2	8	zakonczone	Lyzowy	Lodowisko	Lodowisko	2023-12-05	2024-01-05	Lodowisko, jarmark i zimowe atrakcje	5-99	Stagionama	Sprzęt sportowy	2024-04-06 13:30:22.380	2024-04-13 18:37:20.487
3	16	aktywna	ParkowegoRumuńca	Park rozrywki	Turniej tańca	2024-03-08	2024-03-20	Turniej tańca pod gołym niebem	18-40	Stagionama	Branża taneczna	2024-04-06 13:30:22.380	2024-04-13 14:42:09.643
4	33	planowana	PiknikKulturowy	park	piknik kulturowy	2027-08-05	2027-08-20	Spotkanie z różnymi kulturami, jedzenie i występ...	0-99	stagionama	brak	2024-04-06 13:30:22.380	2024-04-13 14:33:43.883
5	41	zakonczone	ZimoweZabawy	park	zimowe zabawy	2022-12-10	2022-12-25	Rodzinne atrakcje zimowe, kulgi i jazda na sank...	0-99	stagionama	brak	2024-04-06 13:30:22.380	2024-04-13 16:37:37.617

Rysunek 20: Przykład załadowanego wymiaru Akcja

Na rysunku 11 można zaobserwować jak wyglądają załadowane dane do bazy ODS, wraz z data utworzenia i datą modyfikacji która odpowiednio się zmienia wraz z jakąkolwiek modyfikacją danych.

SQLQuery4.sql - 1...W (ABDGR06 (167)) SQLQuery3.sql - 1...DS (ABDGR06 (154))

```

SELECT TOP (1000) [idAkcji]
, [liczbaUczestnikow]
, [calkowityKoszt]
, [dataRozpoczenia]
FROM [ABD_GR06_ODS].[dbo].[FaktAkcjeSt]

```

100 %

Results Messages

	idAkcji	liczbaUczestnikow	calkowityKoszt	dataRozpoczenia
1	1	41	25000.000000	2023-04-01
2	8	35	60000.000000	2023-12-05
3	16	40	35000.000000	2024-03-08
4	33	0	0.000000	2027-08-05
5	41	100	100.000000	2022-12-10

Rysunek 21: Przykład załadowanego wymiaru FaktAkcjaSt

Na powyższym rysunku 14 można zauważyć załadowane dane do bazy ODS do Faktu. Jest tu idAkcji, dataRozpoczenia jako indeks odnoszący się do wymiaru czasu oraz dwie miary sumowalne czyli liczba uczestników i całkowity koszt.

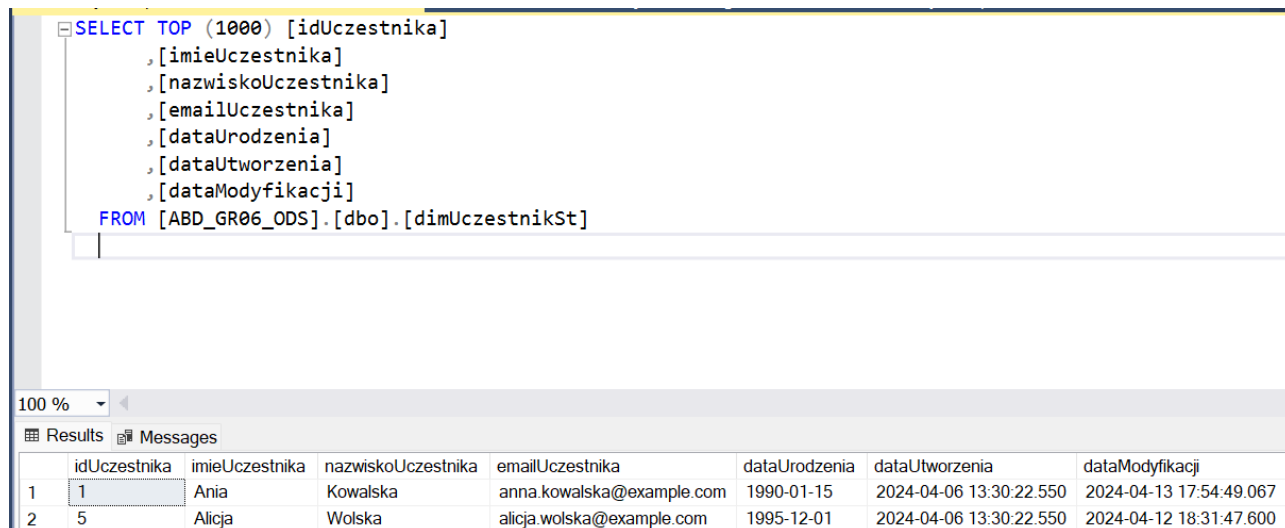
Brakuje tutaj natomiast idKanału oraz idUczestnika, zostały one usunięte ze względu na problemy z ich obsługą (proces ETL nie pozwalał na wielokrotne wykorzystanie tego samego indeksu idAkcji w przypadku uczestników i kanału). Pociąga to za sobą konsekwencje również w przypadku Faktu w centralnej hurtowni danych. Przed realizacją kolejnego etapu zamierzamy uzupełnić tę nieścisłość.

3.5. Opracowanie ETL do DW

Ostatnim etapem projektu było zasilenie docelowej hurtowni danych z warstwy poprzedniej czyli z warstwy pośredniej ODS, przechowującej dodane lub zmodyfikowane rekordy.

3.5.1. Przykład ładowania hurtowni

1. Zmiana rekordu z Anna na Ania w bazie źródłowej SRC2 i załadowanie do ODS:



The screenshot shows a SQL query window with the following query:

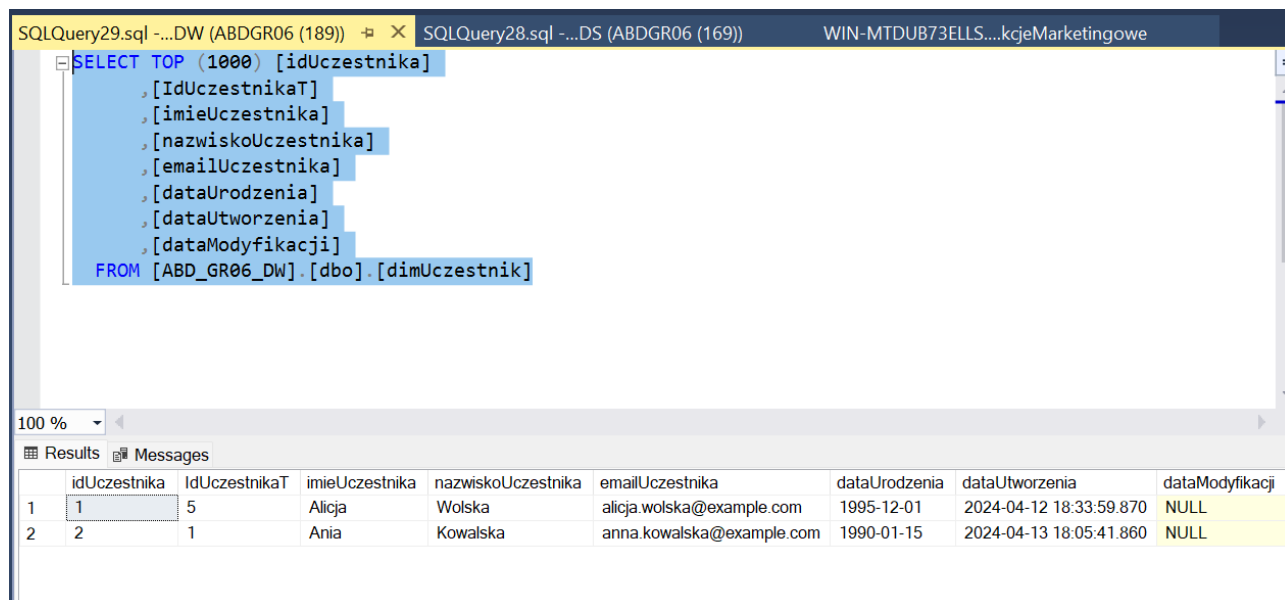
```
SELECT TOP (1000) [idUczestnika]
, [imieUczestnika]
, [nazwiskoUczestnika]
, [emailUczestnika]
, [dataUrodzenia]
, [dataUtworzenia]
, [dataModyfikacji]
FROM [ABD_GR06_ODS].[dbo].[dimUczestnikSt]
```

Below the query, the 'Results' tab shows the following data:

	idUczestnika	imieUczestnika	nazwiskoUczestnika	emailUczestnika	dataUrodzenia	dataUtworzenia	dataModyfikacji
1	1	Ania	Kowalska	anna.kowalska@example.com	1990-01-15	2024-04-06 13:30:22.550	2024-04-13 17:54:49.067
2	5	Alicja	Wolska	alicia.wolska@example.com	1995-12-01	2024-04-06 13:30:22.550	2024-04-12 18:31:47.600

Rysunek 22: Załadowanie zmiany do ODS.

2. Załadowanie zmienionego rekordu do hurtowni danych:



The screenshot shows a SQL query window with the following query:

```
SELECT TOP (1000) [idUczestnika]
, [IdUczestnikaT]
, [imieUczestnika]
, [nazwiskoUczestnika]
, [emailUczestnika]
, [dataUrodzenia]
, [dataUtworzenia]
, [dataModyfikacji]
FROM [ABD_GR06_DW].[dbo].[dimUczestnik]
```

Below the query, the 'Results' tab shows the following data:

	idUczestnika	IdUczestnikaT	imieUczestnika	nazwiskoUczestnika	emailUczestnika	dataUrodzenia	dataUtworzenia	dataModyfikacji
1	1	5	Alicja	Wolska	alicia.wolska@example.com	1995-12-01	2024-04-12 18:33:59.870	NULL
2	2	1	Ania	Kowalska	anna.kowalska@example.com	1990-01-15	2024-04-13 18:05:41.860	NULL

Rysunek 23: Załadowanie hurtowni ze zmianą na Ania.

Widzimy, że w hurtowni zmiana została załadowana o godzinie 18:05.

3. Teraz z powrotem: Ania zostało zamienione na Anna w bazie źródłowej. Na początku dane również zostały załadowane do ODS.

SQLQuery30.sql -...DS (ABDGR06 (194)) SQLQuery29.sql -...DW (ABDGR06 (189)) WIN-MTDUB73ELLS.A...o.UczestnicyAkcji

```

SELECT TOP (1000) [idUczestnika]
, [imieUczestnika]
, [nazwiskoUczestnika]
, [emailUczestnika]
, [dataUrodzenia]
, [dataUtworzenia]
, [dataModyfikacji]
FROM [ABD_GR06_ODS].[dbo].[dimUczestnikSt]

```

100 %

Results Messages

	idUczestnika	imieUczestnika	nazwiskoUczestnika	emailUczestnika	dataUrodzenia	dataUtworzenia	dataModyfikacji
1	1	Anna	Kowalska	anna.kowalska@example.com	1990-01-15	2024-04-06 13:30:22.550	2024-04-13 18:06:52.247
2	5	Alicja	Wolska	alicja.wolska@example.com	1995-12-01	2024-04-06 13:30:22.550	2024-04-12 18:31:47.600

Rysunek 24: Załadowanie ODS.

4. Ponowne ładowanie zmiany do hurtowni danych.

SQLQuery31.sql -...DW (ABDGR06 (174)) SQLQuery30.sql -...DS (ABDGR06 (194)) SQLQuery29.sql -...DW (ABDGR06 (189))

```

SELECT TOP (1000) [idUczestnika]
, [IdUczestnikaT]
, [imieUczestnika]
, [nazwiskoUczestnika]
, [emailUczestnika]
, [dataUrodzenia]
, [dataUtworzenia]
, [dataModyfikacji]
FROM [ABD_GR06_DW].[dbo].[dimUczestnik]

```

100 %

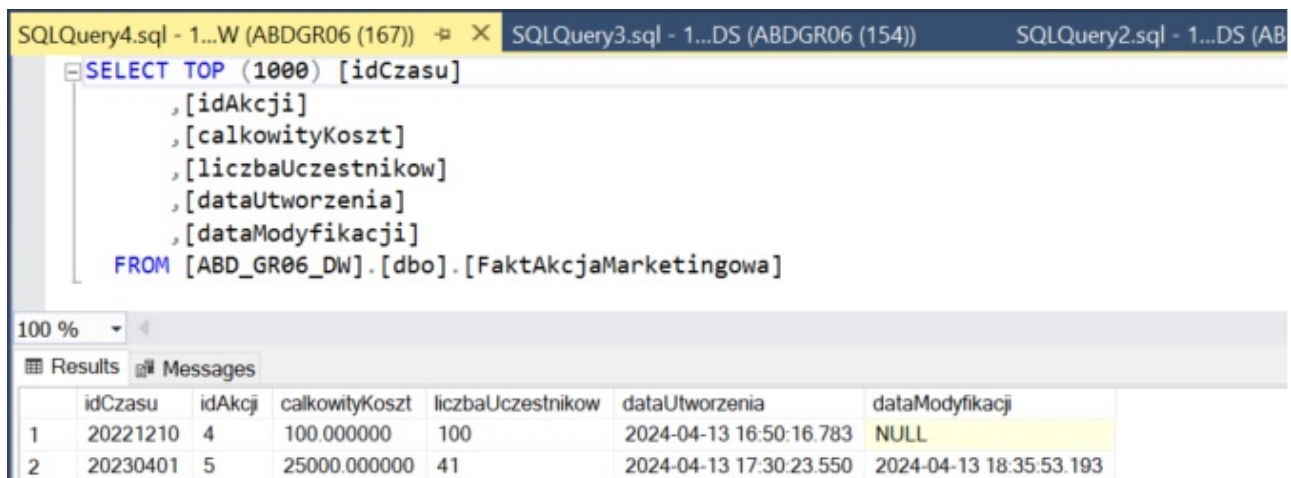
Results Messages

	idUczestnika	IdUczestnikaT	imieUczestnika	nazwiskoUczestnika	emailUczestnika	dataUrodzenia	dataUtworzenia	dataModyfikacji
1	1	5	Alicja	Wolska	alicja.wolska@example.com	1995-12-01	2024-04-12 18:33:59.870	NULL
2	2	1	Anna	Kowalska	anna.kowalska@example.com	1990-01-15	2024-04-13 18:05:41.860	2024-04-13 18:09:15.733

Rysunek 25: Załadowanie ODS

Widzimy, że cały proces przebiegł pomyślnie i finalnie otrzymano rekord w hurtowni z ponowną zmianą na Anna. Data Utworzenia została identyczna, zaś zmieniła się data Modyfikacji na 18:09.

Ładowanie faktu do hurtowni danych



The screenshot shows a SQL Server Enterprise Manager interface. At the top, there are three tabs: 'SQLQuery4.sql - 1...W (ABDGR06 (167))', 'SQLQuery3.sql - 1...DS (ABDGR06 (154))', and 'SQLQuery2.sql - 1...DS (ABDGR06 (154))'. The active tab is 'SQLQuery4.sql'. Below the tabs, the SQL query is displayed:
`SELECT TOP (1000) [idCzasu]`
 `, [idAkcji]`
 `, [calkowityKoszt]`
 `, [liczbaUczestnikow]`
 `, [dataUtworzenia]`
 `, [dataModyfikacji]`
`FROM [ABD_GR06_DW].[dbo].[FaktAkcjaMarketingowa]`
Below the query, there is a 'Results' tab showing the output of the query. The results are displayed in a table with the following columns: 'idCzasu', 'idAkcji', 'calkowityKoszt', 'liczbaUczestnikow', 'dataUtworzenia', and 'dataModyfikacji'. The table contains two rows of data. The first row has values: 1, 20221210, 4, 100.000000, 100, 2024-04-13 16:50:16.783, and NULL. The second row has values: 2, 20230401, 5, 25000.000000, 41, 2024-04-13 17:30:23.550, and 2024-04-13 18:35:53.193.

	idCzasu	idAkcji	calkowityKoszt	liczbaUczestnikow	dataUtworzenia	dataModyfikacji
1	20221210	4	100.000000	100	2024-04-13 16:50:16.783	NULL
2	20230401	5	25000.000000	41	2024-04-13 17:30:23.550	2024-04-13 18:35:53.193

Rysunek 26: Przykład załadowanego wymiaru FaktAkcja w DW

Na rysunku 16 widać przykład załadowanej tabeli faktów oraz przykład gdy dla $idAkcji = 5$ zmienia się liczba uczestników co jest natychmiast odnotowywane poprzez kolumnę `dataModyfikacji`.