

# Projekt integracji systemów baz danych sklepów komputerowych

---

*Patryk Kopyściński D502*

## Spis treści

1. Uzasadnienie biznesowe projektu.....	3
2. Wymagania.....	4
2.1. Funkcjonalne (analityczne).....	4
W01: Analiza ilości zamówień zrealizowanych przez pracowników .....	4
W02: Analiza wielkości sprzedaży według miesiąca .....	4
W03: Analiza wielkości cen podzespołów .....	4
W04: Analiza ilości zamówień klientów .....	4
W05: Analiza wielkości użycia podzespołów według producentów .....	4
2.2. Pozafunkcjonalne .....	4
W01: Baza danych .....	4
W02: ETL.....	4
W03: Analiza danych .....	4
W04: Reporting Services .....	4
3. Projekty baz danych .....	5
3.1.1. OLTP A – BazaA.....	5
3.1.2. OLTP B – BazaB .....	6
3.2. OLAP C - Struktura Hurtowni Danych - BazaC .....	7
SHD1. Wymiary.....	8
SHD2. Fakty.....	12
4. Proces ETL.....	13
4.1. Omówienie .....	13
4.2. Proces czyszczenia .....	15
4.3. Kopiowanie dat z Baz A i B do Bazy C .....	16
4.4. Kopiowanie słowników z Baz A i B do Bazy C .....	17
4.5. Kopiowanie tabel transakcyjnych z Baz A i B do Bazy C .....	21
5. Baza OLAP - Analiza danych w SSAS .....	23
6. Realizacja raportów w Reporting Services .....	25
R01: Analiza ilości zamówień zrealizowanych przez pracowników .....	25
R02: Analiza wielkości sprzedaży według miesiąca .....	25
R03: Analiza wielkości cen podzespołów .....	26
R04: Analiza ilości zamówień klientów .....	26
R05: Analiza wielkości użycia podzespołów według producentów .....	27
7. Job wykonujący nocne przeładowanie procesu ETL.....	27

## 1. Uzasadnienie biznesowe projektu

Moi rodzice prowadzili w Radzyminie sklep komputerowy - Sklep A. Od zawsze interesowałem się informatyką i starałem się im pomagać. Gdy ukończyłem studia informatyczne z bardzo dobrym wynikiem, postanowili mi go przekazać. Firma bardzo dobrze prosperowała i przynosiła spore dochody. Po głębszym zastanowieniu się i zapoznaniu nad sytuacją panującą na rynku lokalnym Radzymina stwierdziłem, że otwarcie kolejnego lokalu w tym samym małym mieście nie będzie przynosiło tak dużych zysków. Rozpocząłem poszukiwania ofert sprzedaży podobnych działalności w Warszawie. Przedsięwzięcie miało wielką szansę na realizację i utrzymanie się na rynku, gdyż w stolicy na dzień dzisiejszy jest ogromne zapotrzebowanie na takie usługi. Znalazłem bardzo korzystną ofertę sklepu komputerowego w dogodnej lokalizacji. Postanowiłem więc pozyskać Sklep B.

Obie firmy posiadają bazy danych OLTP. Zawierają te same dane, lecz w odmiennie nazwanych tabelach i z różnymi typami danych. W celu ujednolicenia systemu, należy stworzyć trzecią bazę. Baza ma za zadanie przechowywać informacje o produktach i klientach sklepu komputerowego. Będzie przechowywać zestawy komputerowe i dane o klientach. Kolejną opcją jest wyświetlanie poszczególnych zamówień klientów. Dzięki temu będzie można obliczyć przychody z funkcjonowania działalności. Będzie wiadomo jakie zamówienia są najczęściej zlecane i jakie podzespoły są najbardziej opłacalne.

## 2. Wymagania

### 2.1. Funkcjonalne (analityczne)

#### **W01: Analiza ilości zamówień zrealizowanych przez pracowników**

Ma przedstawiać wykres ilości zamówień obsłużonych przez pracowników z poszczególnych stanowisk.

#### **W02: Analiza wielkości sprzedaży według miesiąca**

Ma przedstawiać wykres ilości zamówień według miesiąca.

#### **W03: Analiza wielkości cen podzespołów**

Ma przedstawiać wykres zróżnicowania wartości zamówień.

#### **W04: Analiza ilości zamówień klientów**

Ma przedstawiać wykres ilości zamówień poszczególnych klientów.

#### **W05: Analiza wielkości użycia podzespołów według producentów**

Ma przedstawiać wykres ilości zamówień poszczególnych producentów.

### 2.2. Pozafunkcjonalne

#### **W01: Baza danych**

Dostęp do dwóch baz transakcyjnych BazaA oraz BazaB - SQL Server.

Baza OLAP - hurtownia danych, magazynująca rekordy z dwóch baz transakcyjnych.

#### **W02: ETL**

Integracja danych poprzez SQL Server Integration Services.

#### **W03: Analiza danych**

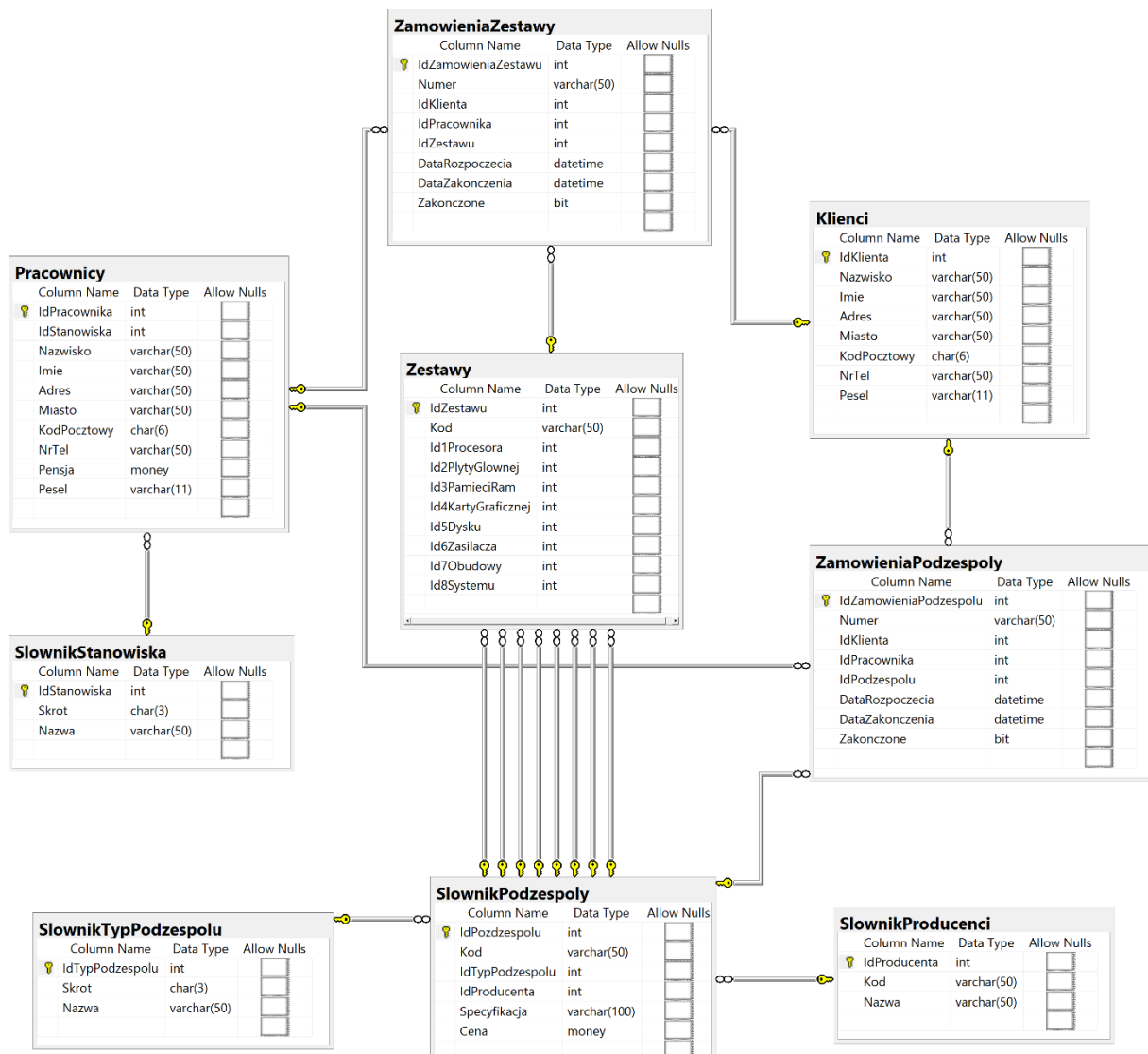
Analityka w oparciu o środowisko Microsoft SQL Server Analysis Services (SSAS).

#### **W04: Reporting Services**

Stworzenie raportów prezentujących opisane w wymaganiach analitycznych realizacje wskaźników KPI.

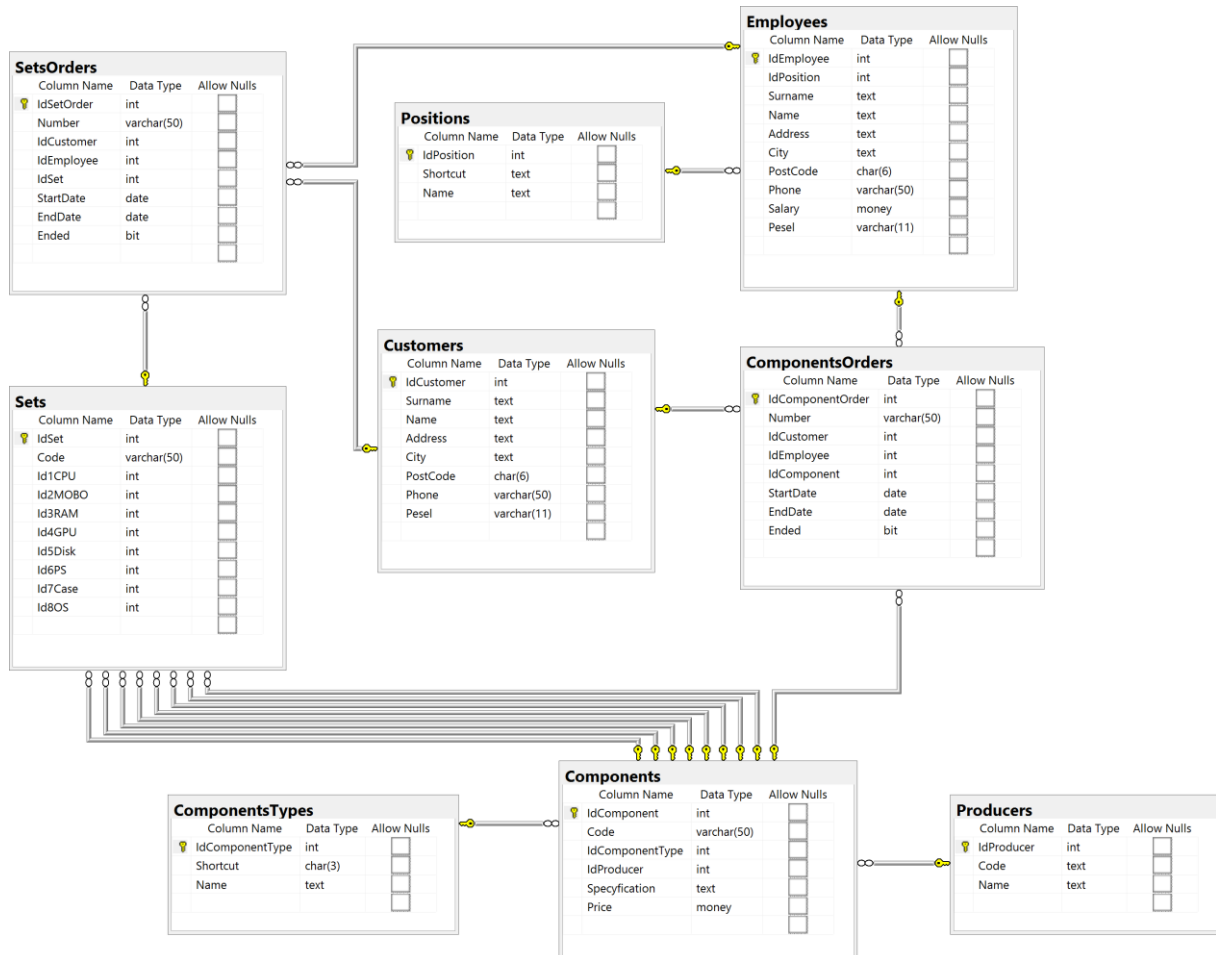
### 3. Projekty baz danych

#### 3.1.1. OLTP A – BazaA



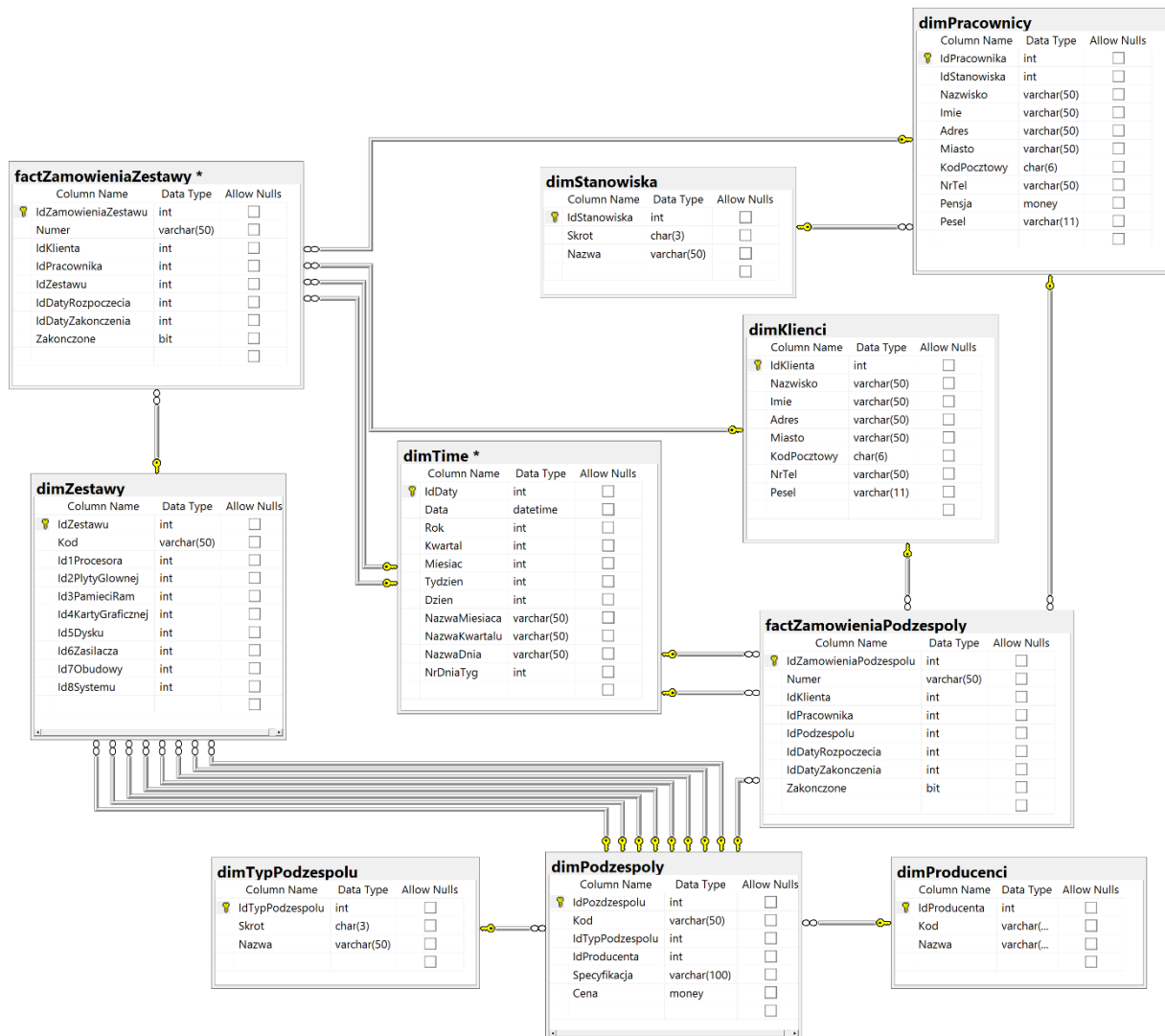
Baza transakcyjna BazaA zawiera tabele budujące strukturę firmy A. Została wykonana z zachowanymi zasadami projektowania. Zawiera słowniki producentów, typów podzespołów i stanowisk pracowników. Posiada informacje o pracownikach, klientach i ich zamówieniach. Na zamówienia składają się pojedyncze podzespoły lub całe zestawy komputerowe. Zostały zachowane odpowiednie typy danych dla poszczególnych rekordów.

### 3.1.2. OLTP B – BazaB



Baza transakcyjna BazaB zawiera tabele budujące strukturę firmy B. W strukturze jest bardzo podobna do bazy pierwszej, ale jest w języku angielskim. Posiada nieodpowiednie typy danych dla poszczególnych rekordów.

### 3.2. OLAP C - Struktura Hurtowni Danych - BazaC



BazaC - OLAP. Będą w niej przechowywane wspólne dane składowane na skutek wszystkich procesów ETL. Jest kopią bazy A, lecz została podzielona na wymiary i fakty. Aby ułatwić raportowanie, została dodana tabela z czasem dimTime.

## SHD1. Wymiary


dimPracownicy			
	Column Name	Data Type	Allow Nulls
	IdPracownika	int	<input type="checkbox"/>
	IdStanowiska	int	<input type="checkbox"/>
	Nazwisko	varchar(50)	<input type="checkbox"/>
	Imie	varchar(50)	<input type="checkbox"/>
	Adres	varchar(50)	<input type="checkbox"/>
	Miasto	varchar(50)	<input type="checkbox"/>
	KodPocztowy	char(6)	<input type="checkbox"/>
	NrTel	varchar(50)	<input type="checkbox"/>
	Pensja	money	<input type="checkbox"/>
	Pesel	varchar(11)	<input type="checkbox"/>
			<input type="checkbox"/>

Tabela w której ewidencjonuje się pracowników zatrudnionych w firmie.


dimStanowiska			
	Column Name	Data Type	Allow Nulls
	IdStanowiska	int	<input type="checkbox"/>
	Skrot	char(3)	<input type="checkbox"/>
	Nazwa	varchar(50)	<input type="checkbox"/>
			<input type="checkbox"/>

Tabela która jest słownikiem stanowisk w firmie.



### dimKlienci


	Column Name	Data Type	Allow Nulls
	IdKlienta	int	<input type="checkbox"/>
	Nazwisko	varchar(50)	<input type="checkbox"/>
	Imie	varchar(50)	<input type="checkbox"/>
	Adres	varchar(50)	<input type="checkbox"/>
	Miasto	varchar(50)	<input type="checkbox"/>
	KodPocztowy	char(6)	<input type="checkbox"/>
	NrTel	varchar(50)	<input type="checkbox"/>
	Pesel	varchar(11)	<input type="checkbox"/>
			<input type="checkbox"/>

Tabela w której ewidencjonuje się klientów.

### dimProducenci


	Column Name	Data Type	Allow Nulls
	IdProducenta	int	<input type="checkbox"/>
	Kod	varchar(...)	<input type="checkbox"/>
	Nazwa	varchar(...)	<input type="checkbox"/>
			<input type="checkbox"/>

Tabela która jest słownikiem producentów podzespołów.

### dimTypPodzespolu


	Column Name	Data Type	Allow Nulls
	IdTypPodzespolu	int	<input type="checkbox"/>
	Skrot	char(3)	<input type="checkbox"/>
	Nazwa	varchar(50)	<input type="checkbox"/>
			<input type="checkbox"/>

Tabela która jest słownikiem typów podzespołów komputerowych.


dimPodzespolu			
	Column Name	Data Type	Allow Nulls
	IdPozdzespolu	int	<input type="checkbox"/>
	Kod	varchar(50)	<input type="checkbox"/>
	IdTypPodzespolu	int	<input type="checkbox"/>
	IdProducenta	int	<input type="checkbox"/>
	Specyfikacja	varchar(100)	<input type="checkbox"/>
	Cena	money	<input type="checkbox"/>
			<input type="checkbox"/>

Tabela w której ewidencjonuje się dostępne podzespoły.


dimZestawy			
	Column Name	Data Type	Allow Nulls
	IdZestawu	int	<input type="checkbox"/>
	Kod	varchar(50)	<input type="checkbox"/>
	Id1Procesora	int	<input type="checkbox"/>
	Id2PlytyGlownej	int	<input type="checkbox"/>
	Id3PamieciRam	int	<input type="checkbox"/>
	Id4KartyGraficznej	int	<input type="checkbox"/>
	Id5Dysku	int	<input type="checkbox"/>
	Id6Zasilacza	int	<input type="checkbox"/>
	Id7Obudowy	int	<input type="checkbox"/>
	Id8Systemu	int	<input type="checkbox"/>
			<input type="checkbox"/>

Tabela w której ewidencjonuje się złożone zestawy.

**dimTime \***


	Column Name	Data Type	Allow Nulls
	IdDaty	int	<input type="checkbox"/>
	Data	datetime	<input type="checkbox"/>
	Rok	int	<input type="checkbox"/>
	Kwartal	int	<input type="checkbox"/>
	Miesiac	int	<input type="checkbox"/>
	Tydzien	int	<input type="checkbox"/>
	Dzien	int	<input type="checkbox"/>
	NazwaMiesiaca	varchar(50)	<input type="checkbox"/>
	NazwaKwartalu	varchar(50)	<input type="checkbox"/>
	NazwaDnia	varchar(50)	<input type="checkbox"/>
	NrDniaTyg	int	<input type="checkbox"/>
			<input type="checkbox"/>

Tabela czasu która jest powiązana z różnymi tabelami faktów.

## SHD2. Fakty


factZamowieniaPodzespoly			
	Column Name	Data Type	Allow Nulls
	IdZamowieniaPodzespolu	int	<input type="checkbox"/>
	Numer	varchar(50)	<input type="checkbox"/>
	IdKlienta	int	<input type="checkbox"/>
	IdPracownika	int	<input type="checkbox"/>
	IdPodzespolu	int	<input type="checkbox"/>
	IdDatyRozpoczecia	int	<input type="checkbox"/>
	IdDatyZakonczenia	int	<input type="checkbox"/>
	Zakonczzone	bit	<input type="checkbox"/>
			<input type="checkbox"/>

Tabela w której ewidencjonuje się zamówienia podzespołów.


factZamowieniaZestawy *			
	Column Name	Data Type	Allow Nulls
	IdZamowieniaZestawu	int	<input type="checkbox"/>
	Numer	varchar(50)	<input type="checkbox"/>
	IdKlienta	int	<input type="checkbox"/>
	IdPracownika	int	<input type="checkbox"/>
	IdZestawu	int	<input type="checkbox"/>
	IdDatyRozpoczecia	int	<input type="checkbox"/>
	IdDatyZakonczenia	int	<input type="checkbox"/>
	Zakonczzone	bit	<input type="checkbox"/>
			<input type="checkbox"/>

Tabela w której ewidencjonuje się zamówienia zestawów podzespołów.

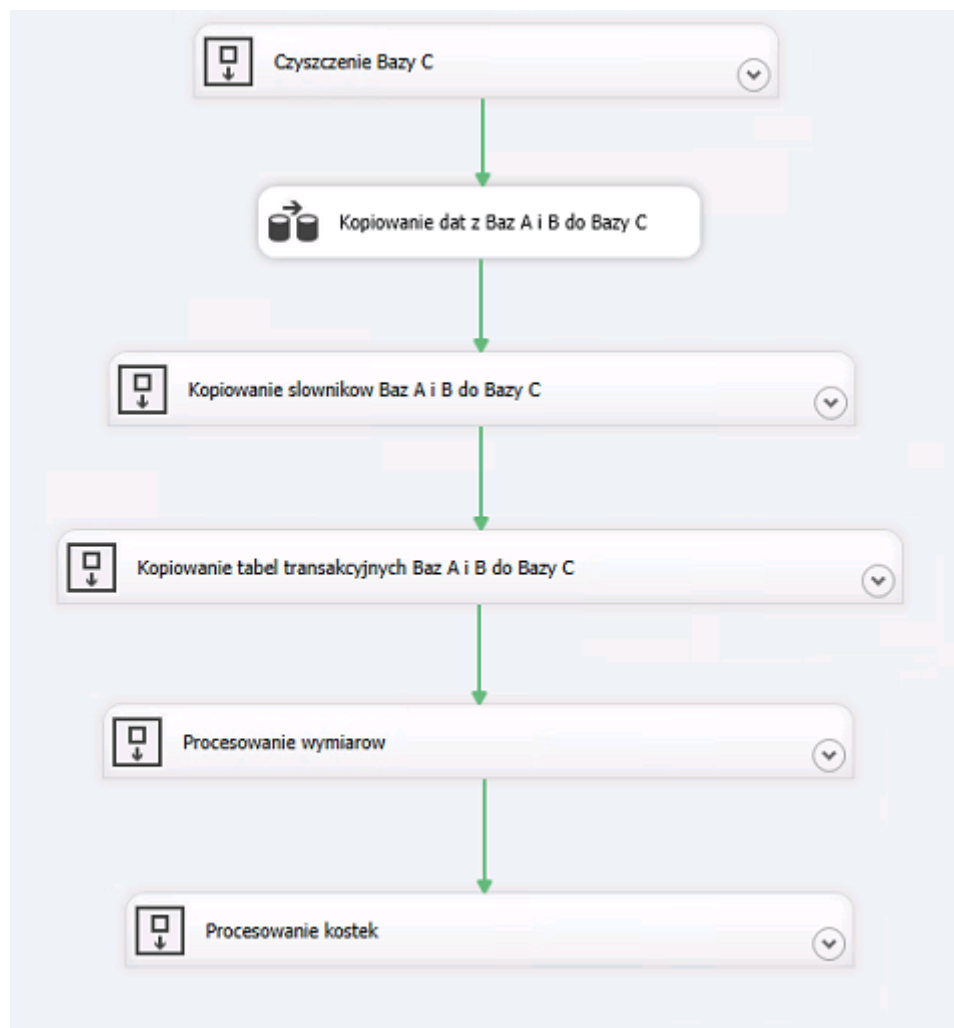
## 4. Proces ETL

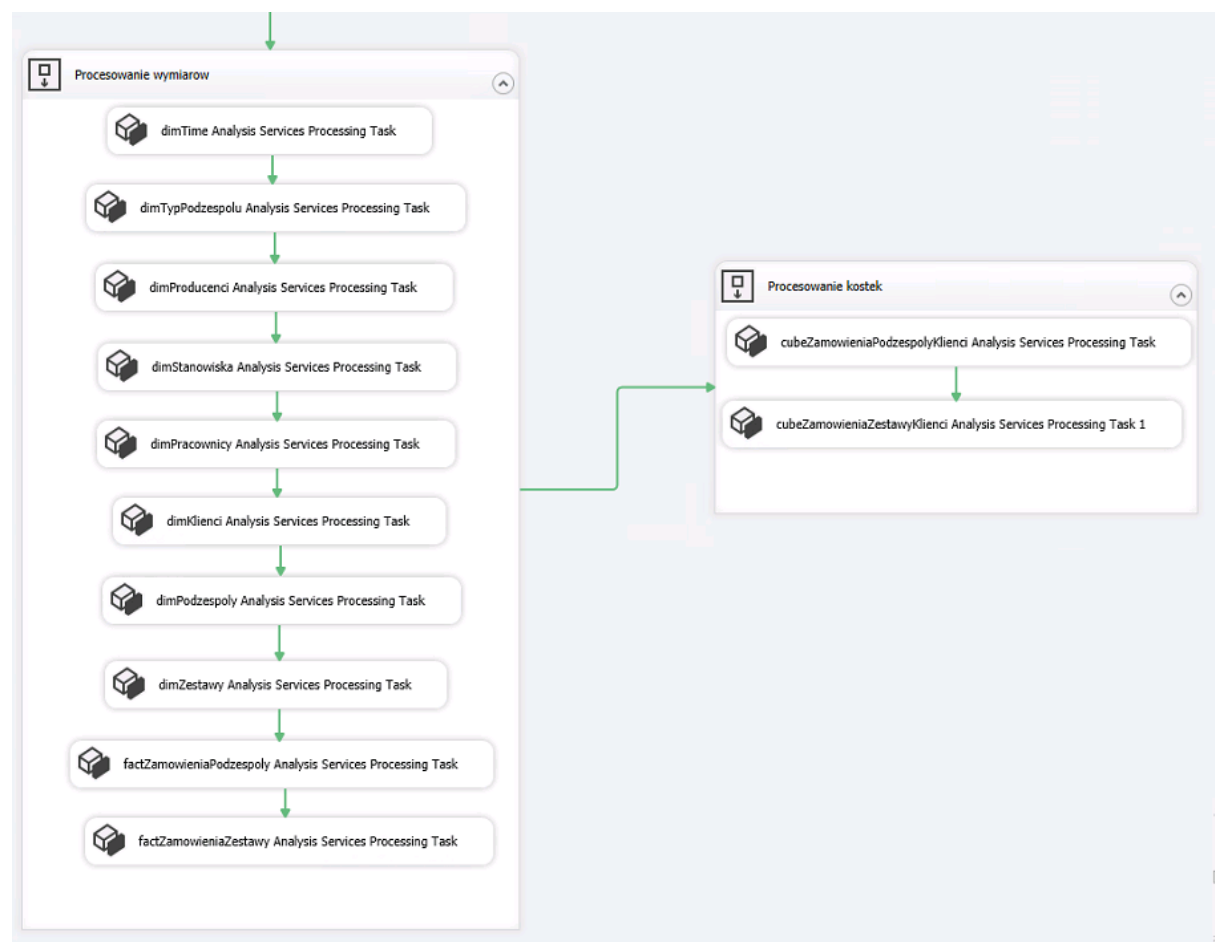
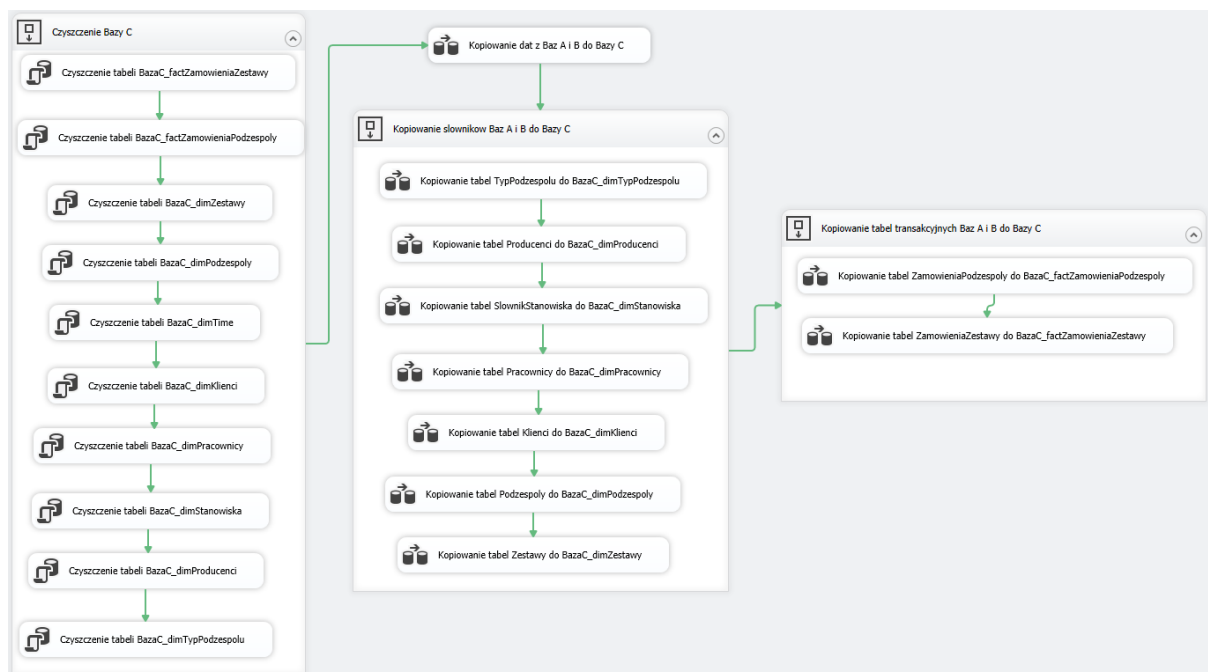
### 4.1. Omówienie

ETL – Extract, Transform and Load. Są to narzędzia wspomagające proces pozyskiwania danych dla hurtowni danych. Zadaniem narzędzi ETL jest:

- pozyskiwanie danych ze źródeł zewnętrznych
- przekształcanie danych
- załadowanie danych do bazy źródłowej, często będącej hurtownią danych

W projekcie hurtowni jest baza OLAP: BazaC. Do procesów ETL użyjemy narzędzi SSIS. Efekt końcowy został przedstawiony poniżej:



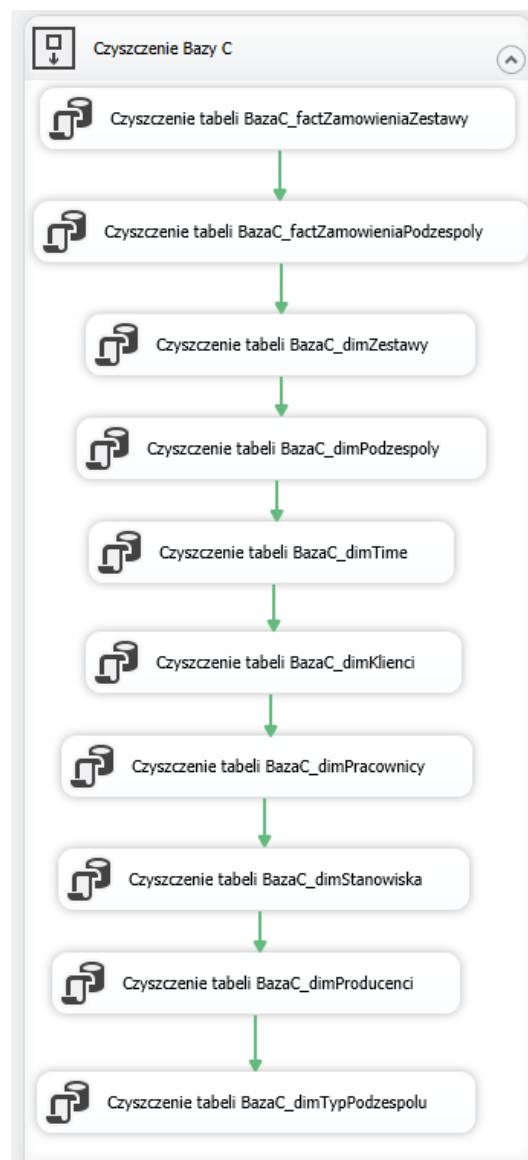


## 4.2. Proces czyszczenia

Przed procesem integracji danych należy oczyścić bazę C.

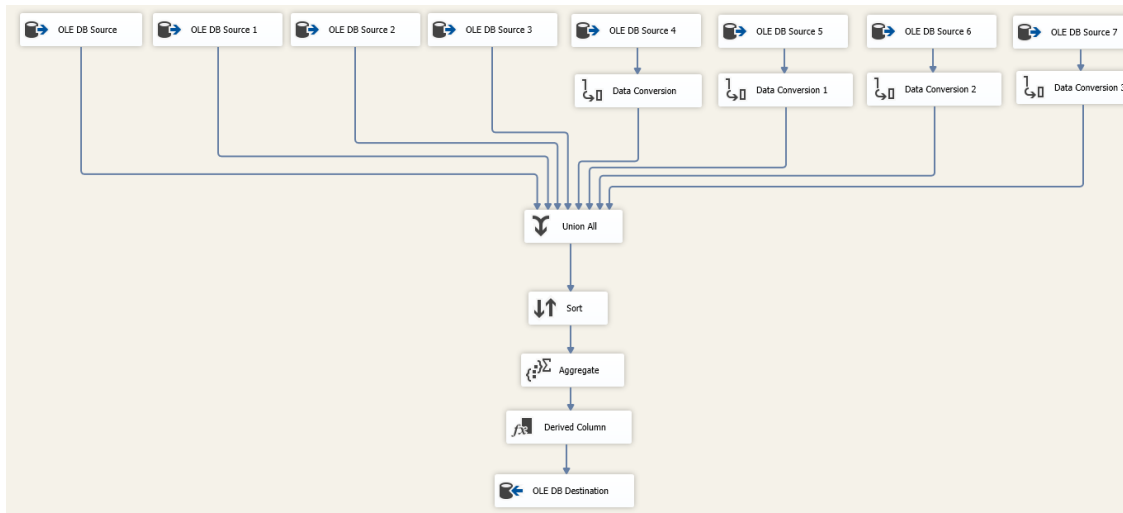
Użyjemy komponentu Execute SQL Task w którym wykonamy następujące zapytania:

```
delete from factZamowieniaZestawy DBCC CHECKIDENT ('dbo.factZamowieniaZestawy', RESEED, 0)
delete from factZamowieniaPodzespolu DBCC CHECKIDENT ('dbo.factZamowieniaPodzespolu', RESEED, 0)
delete from dimZestawy DBCC CHECKIDENT ('dbo.dimZestawy', RESEED, 0)
delete from dimPodzespolu DBCC CHECKIDENT ('dbo.dimPodzespolu', RESEED, 0)
delete from dimTime DBCC CHECKIDENT ('dbo.dimTime', RESEED, 0)
delete from dimKlienci DBCC CHECKIDENT ('dbo.dimKlienci', RESEED, 0)
delete from dimPracownicy DBCC CHECKIDENT ('dbo.dimPracownicy', RESEED, 0)
delete from dimStanowiska DBCC CHECKIDENT ('dbo.dimStanowiska', RESEED, 0)
delete from dimProducenci DBCC CHECKIDENT ('dbo.dimProducenci', RESEED, 0)
delete from dimTypPodzespolu DBCC CHECKIDENT ('dbo.dimTypPodzespolu', RESEED, 0)
```



### 4.3. Kopiowanie dat z Baz A i B do Bazy C

Przed skopiowaniem słowników, zasilimy bazę Datami z Baz A i B. Musimy doprowadzić je do wspólnego formatu za pomocą Data Conversion, a następnie połączyć za pomocą Union All. Następnie sortujemy dane. Po sortowaniu Agregujemy dane. Za pomocą Derived Column wyłuskujemy poszczególne elementy daty.



Derived Column Transformation Editor

Specify the expressions used to create new column values, and indicate whether the values update existing columns or populate new columns.

Variables and Parameters

Columns

Mathematical Functions

String Functions

Date/Time Functions

NULL Functions

Type Casts

Operators

Description:

Derived Column Name	Derived Column	Expression	Data Type	
Rok	<add as new column>	YEAR(Data)	four-byte signed integer [...]	
Kwartal	<add as new column>	DATEPART("qq",Data)	four-byte signed integer [...]	
Miesiac	<add as new column>	MONTH(Data)	four-byte signed integer [...]	
Dzien	<add as new column>	DAY(Data)	four-byte signed integer [...]	
Tydzien	<add as new column>	DATEPART("ww",Data)	four-byte signed integer [...]	
NrDniaTyg	<add as new column>	DATEPART("dw",Data)	four-byte signed integer [...]	

Configure Error Output... OK Cancel Help



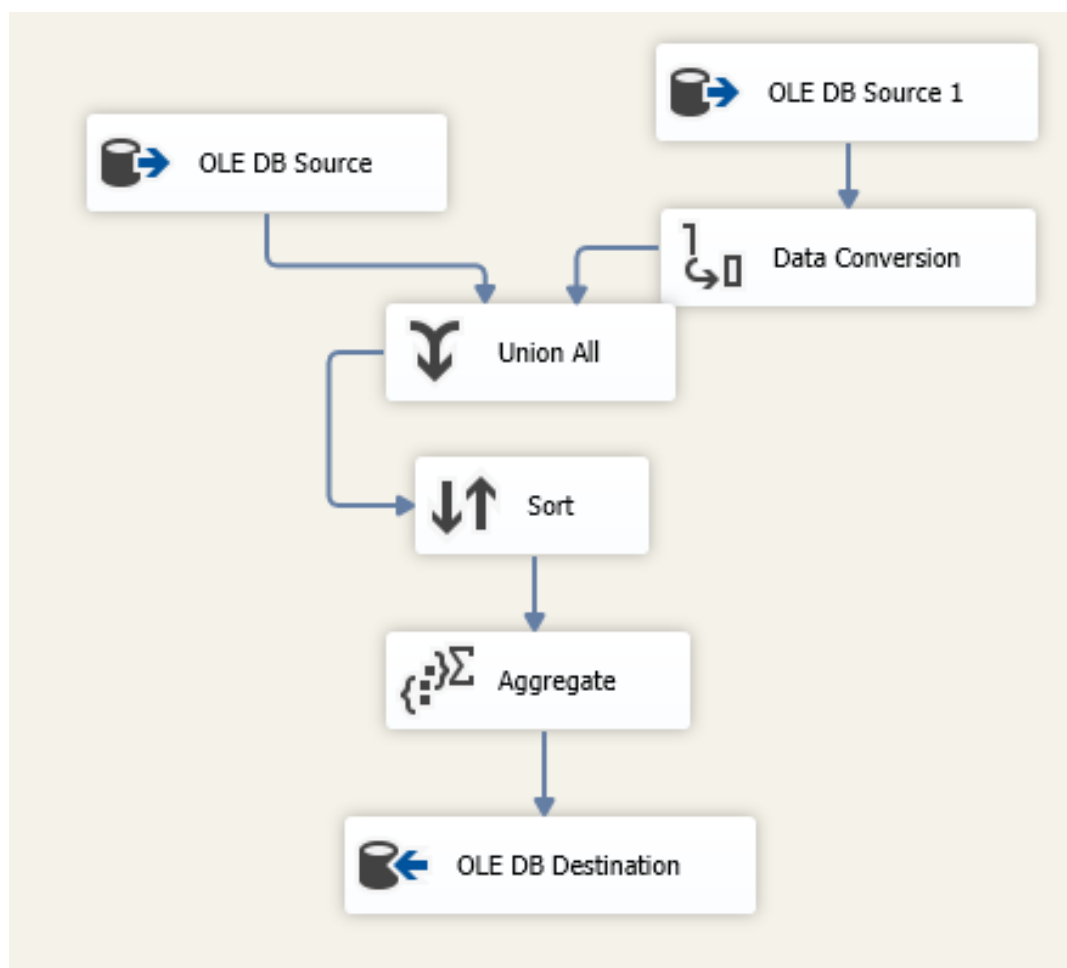
#### 4.4. Kopiowanie słowników z Baz A i B do Bazy C

Teraz na podstawie jednego przykładu, zaprezentuję transfer słowników z Bazy A i B do Bazy C.

Do wykonania tego zostały użyte następujące narzędzia:

- Data Conversion do konwersji typów danych
- Union All do połączenia kilku strumieni danych
- Sort do posortowania danych
- Aggregate do zgrupowania i usunięcia duplikatów

**Data flow: Kopiowanie tabel TypPodzespolu do BazaC\_dimTypPodzespolu**



## OLE DB Source

Configure the properties used by a data flow to obtain data from any OLE DB provider.

**Connection Manager**  
Columns  
Error Output

Specify an OLE DB connection manager, a data source, or a data source view, and select the data access mode. If using the SQL command access mode, specify the SQL command either by typing the query or by using Query Builder.

OLE DB connection manager:  
WIN-E6VLVMUO30E.BazaA New...

Data access mode:  
Table or view

Name of the table or the view:  
[dbo].[SlownikTypPodzespolu]

## OLE DB Source 1

Configure the properties used by a data flow to obtain data from any OLE DB provider.

**Connection Manager**  
Columns  
Error Output

Specify an OLE DB connection manager, a data source, or a data source view, and select the data access mode. If using the SQL command access mode, specify the SQL command either by typing the query or by using Query Builder.

OLE DB connection manager:  
WIN-E6VLVMUO30E.BazaB New...

Data access mode:  
Table or view

Name of the table or the view:  
[dbo].[ComponentsTypes]

## Data Conversion

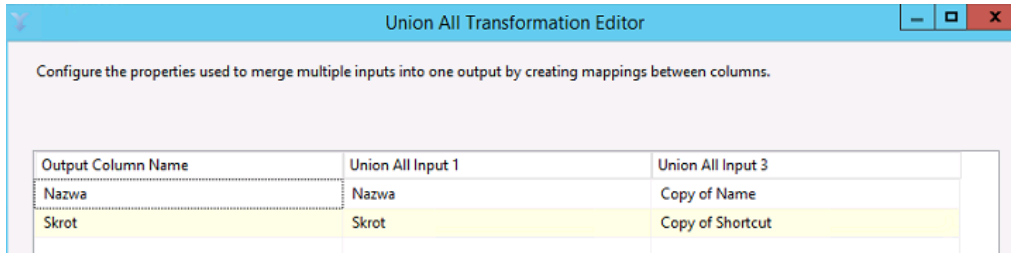
Configure the properties used to convert the data type of an input column to a different data type. Depending on the data type to which the column is converted, set the length, precision, scale, and code page of the column.

**Available Input Columns**

- ☒ Name
- ☒ Shortcut
- ☒ Name

Input Column	Output Alias	Data Type	Length	Precision	Scale	Code Page
Shortcut	Copy of Shortcut	string [DT_STR]	3			1252 (ANSI - La
Name	Copy of Name	string [DT_STR]	50			1252 (ANSI - La

## Union All

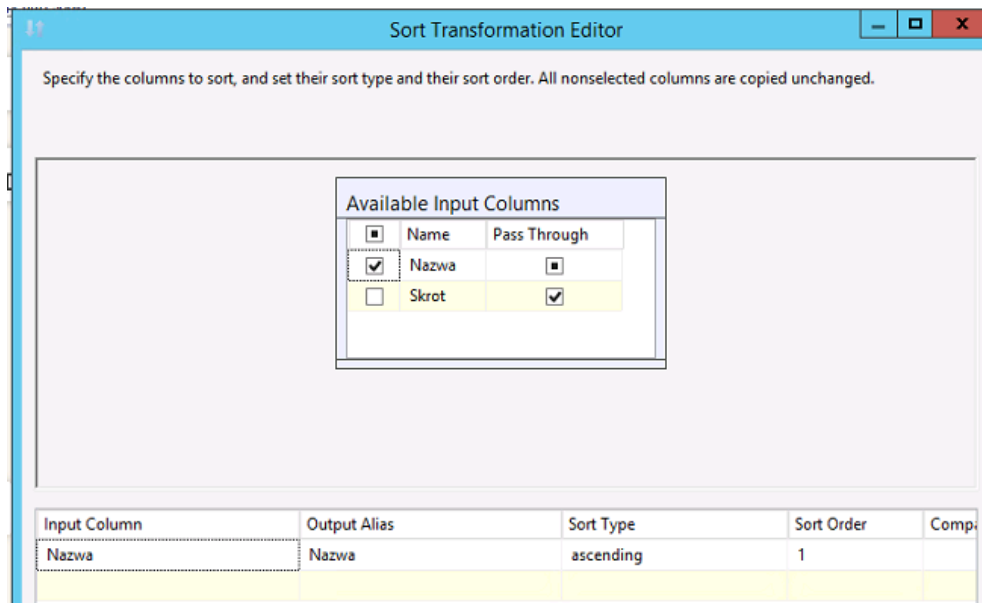


Union All Transformation Editor

Configure the properties used to merge multiple inputs into one output by creating mappings between columns.

Output Column Name	Union All Input 1	Union All Input 3
Nazwa	Nazwa	Copy of Name
Skrot	Skrot	Copy of Shortcut

## Sort



Sort Transformation Editor

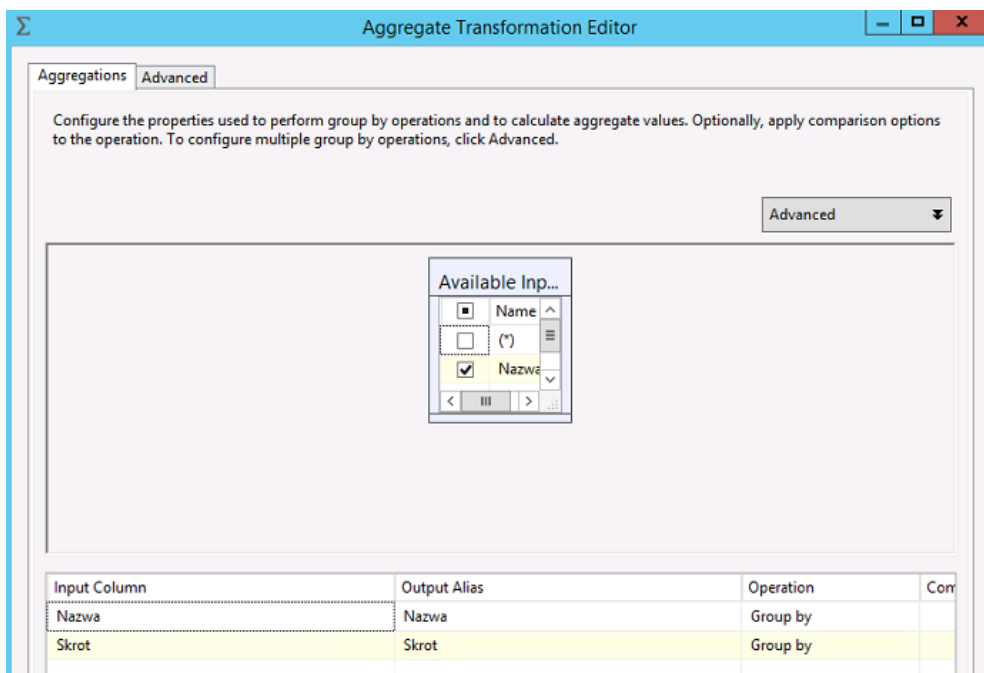
Specify the columns to sort, and set their sort type and their sort order. All nonselected columns are copied unchanged.

Available Input Columns

<input type="checkbox"/>	Name	Pass Through
<input checked="" type="checkbox"/>	Nazwa	<input type="checkbox"/>
<input type="checkbox"/>	Skrot	<input checked="" type="checkbox"/>

Input Column	Output Alias	Sort Type	Sort Order	Comp.
Nazwa	Nazwa	ascending	1	

## Aggregate



Aggregate Transformation Editor

Aggregations Advanced

Configure the properties used to perform group by operations and to calculate aggregate values. Optionally, apply comparison options to the operation. To configure multiple group by operations, click Advanced.

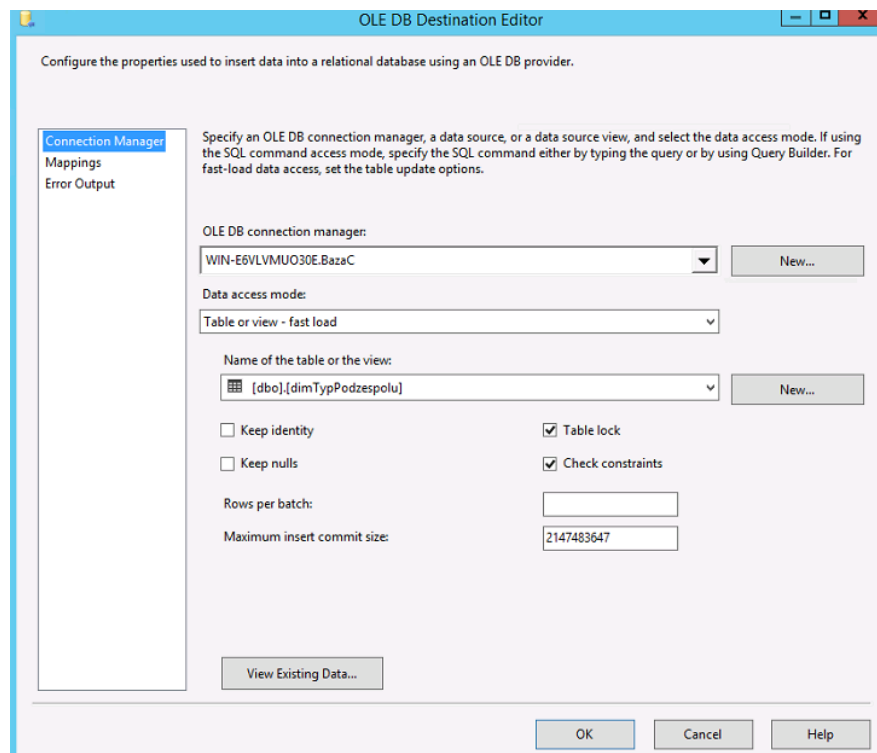
Advanced

Available Inp...

<input type="checkbox"/>	Name
<input type="checkbox"/>	(*)
<input checked="" type="checkbox"/>	Nazwa

Input Column	Output Alias	Operation	Comp.
Nazwa	Nazwa	Group by	
Skrot	Skrot	Group by	

## OLE DB Destination



Configure the properties used to insert data into a relational database using an OLE DB provider.

**Connection Manager**  
Mappings  
Error Output

Specify an OLE DB connection manager, a data source, or a data source view, and select the data access mode. If using the SQL command access mode, specify the SQL command either by typing the query or by using Query Builder. For fast-load data access, set the table update options.

OLE DB connection manager:  
WIN-E6VLVMUO30E.BazaC

Data access mode:  
Table or view - fast load

Name of the table or the view:  
[dbo].[dimTypPodzespolu]

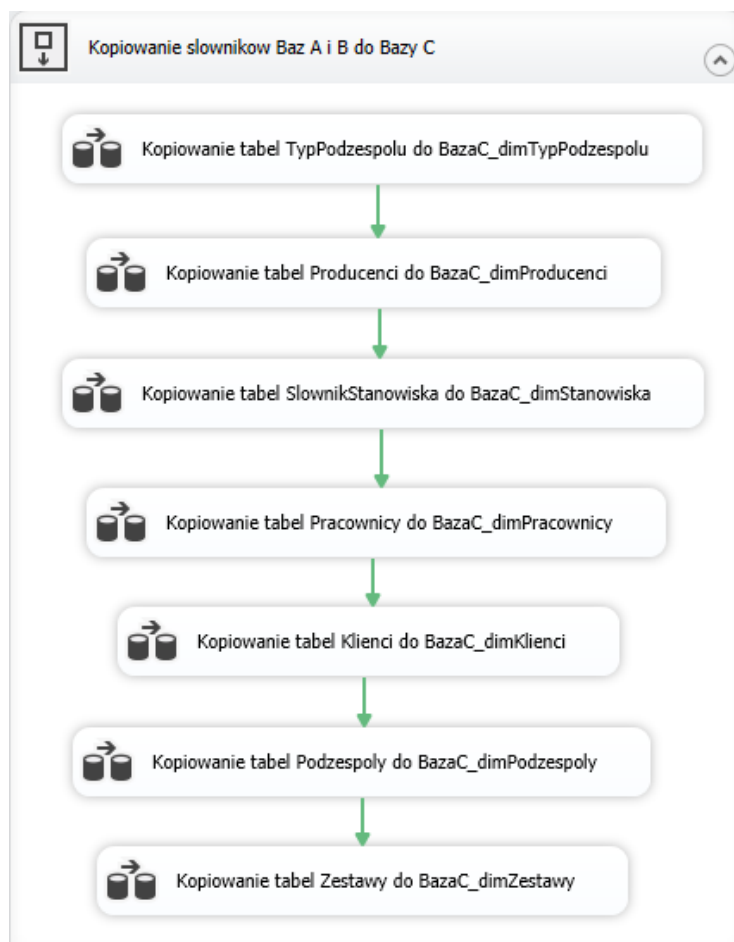
☐ Keep identity ☒ Table lock  
☐ Keep nulls ☒ Check constraints

Rows per batch:   
Maximum insert commit size: 2147483647

View Existing Data...

OK Cancel Help

## Kontener: Kopiowanie słowników Baz A i B do Bazy C

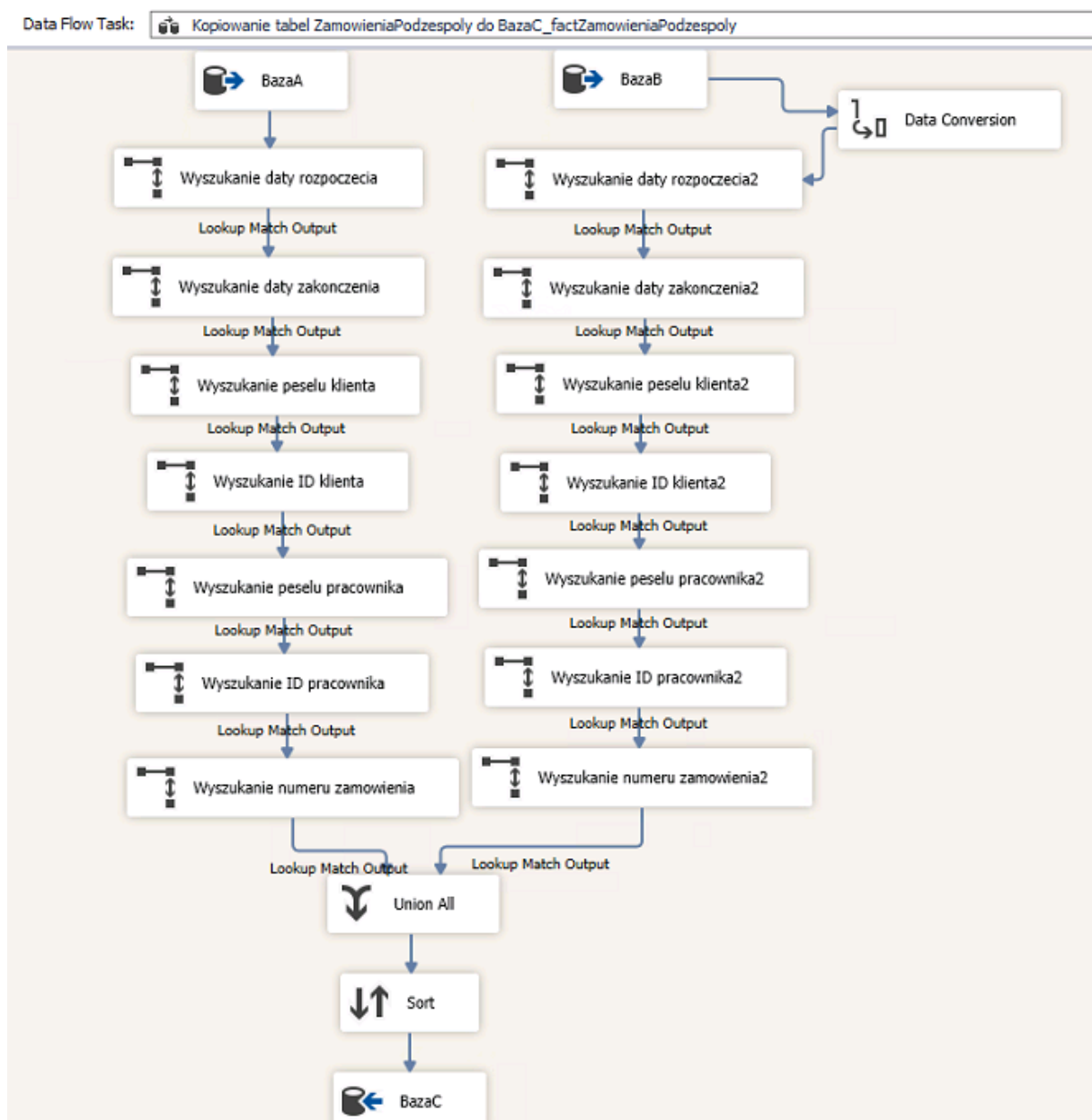


## 4.5. Kopiowanie tabel transakcyjnych z Baz A i B do Bazy C

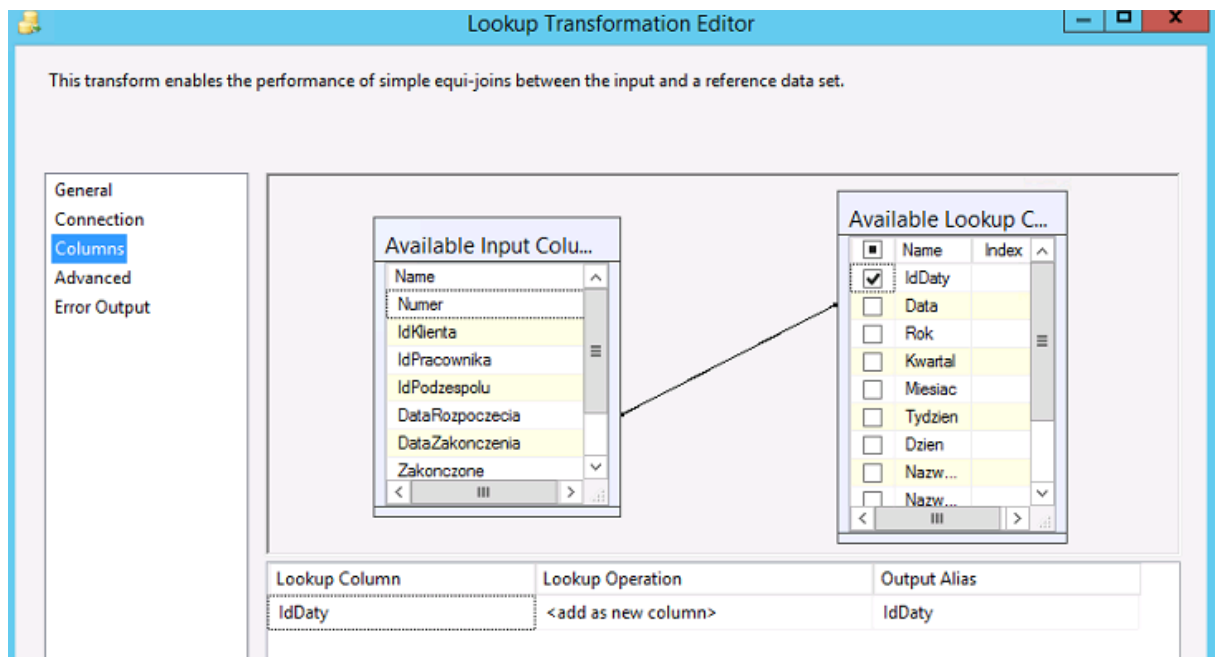
Przenoszenie tabel transakcyjnych na przykładzie jednej z tabel.

Do wykonania tego zostały użyte następujące narzędzia:

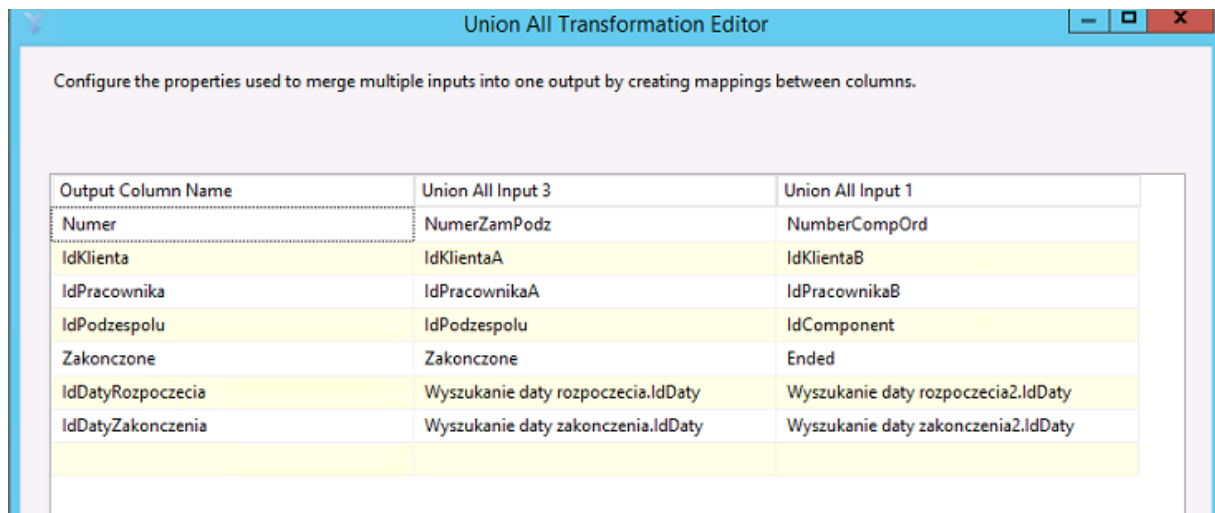
- Data Conversion do konwersji typów danych
- Lookup, które sprawdzają czy rekordy nie znajdują się już w bazie i odpowiednio pobierają dane z bazy OLAP



## Lookup: Wyszukanie daty rozpoczęcia

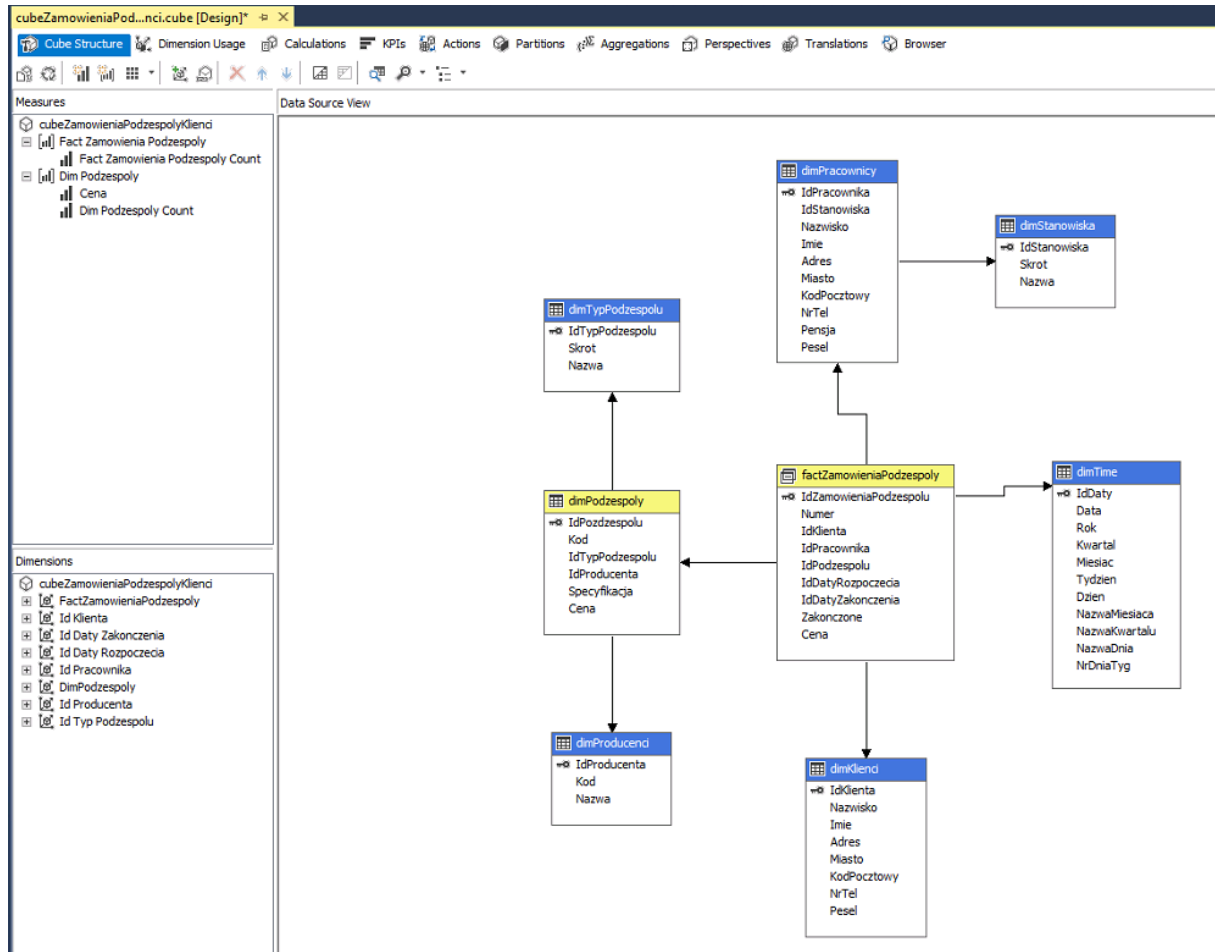


## Union All



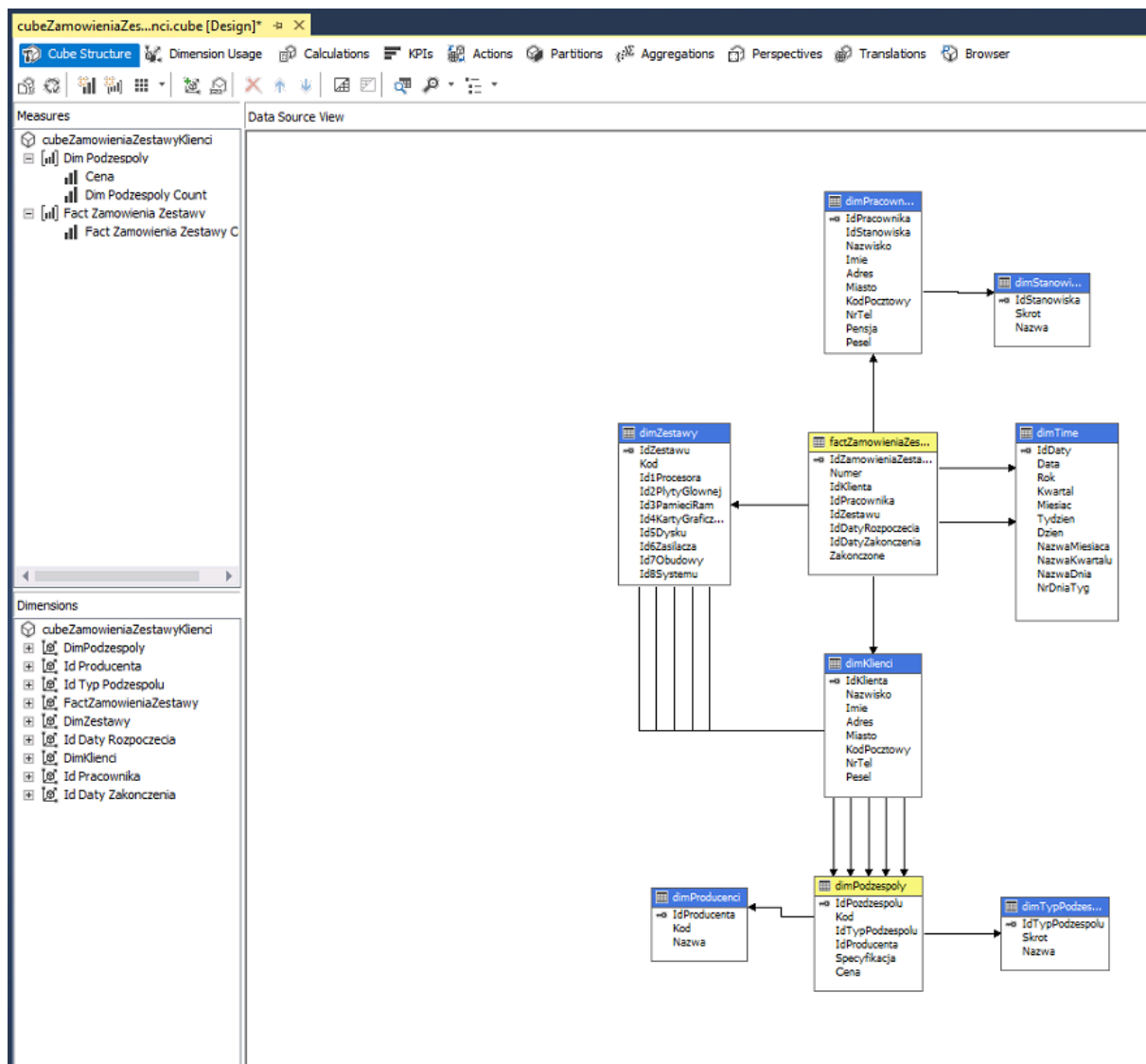
## 5. Baza OLAP - Analiza danych w SSAS

### Przykładowa kostka wraz z wymiarami "cubeZamowieniaPodzespolowyKlienti"



Kostka w której miarą jest ilość zamówień podzespołów, a także cena i ilość podzespołów. Wymiarami będą Pracownicy, Stanowiska, Czas, Klienci, Producenci, Podzespoły i TypPodzespolu.

### Przykładowa kostka wraz z wymiarami "cubeZamowieniaZestawyKlienci"



Kostka w której miarą jest ilość zamówień podzespołów, a także cena i ilość podzespołów.

Wymiarami będą Pracownicy, Stanowiska, Czas, TypPodzespolu, Podzespolny, Producenci, Zestawy.



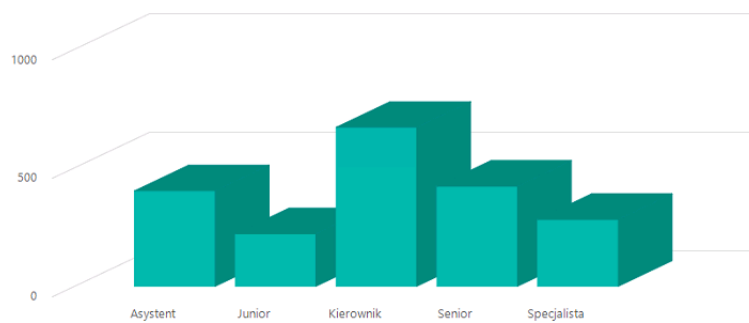
## 6. Realizacja raportów w Reporting Services

### R01: Analiza ilości zamówień zrealizowanych przez pracowników

#### Obsługa zamówień

Stanowisko	Ilość zamówień	Imię	Nazwisko
Asystent	40	Karrah	Lazare
	42	Annadiane	Getch
	45	Raquela	Shulem
	51	Gustav	Trett
	54	Rollo	Rounds
	57	Anastasia	Mardee
	63	Darb	Amberger
		Reynold	Andrewartha
Junior			
Kierownik			
Senior			
Specjalista			

Ilość zamówień obsługiwanych przez pracowników z poszczególnych stanowisk



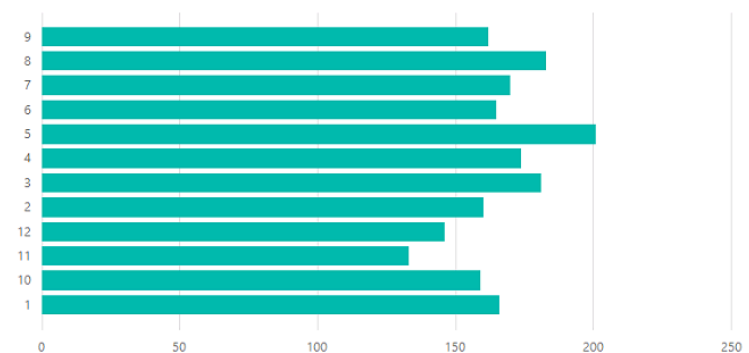
Wykres przedstawia ilość zamówień obsługiwanych przez pracowników z poszczególnych stanowisk.

### R02: Analiza wielkości sprzedaży według miesiąca

#### Ilość zamówień - miesiące

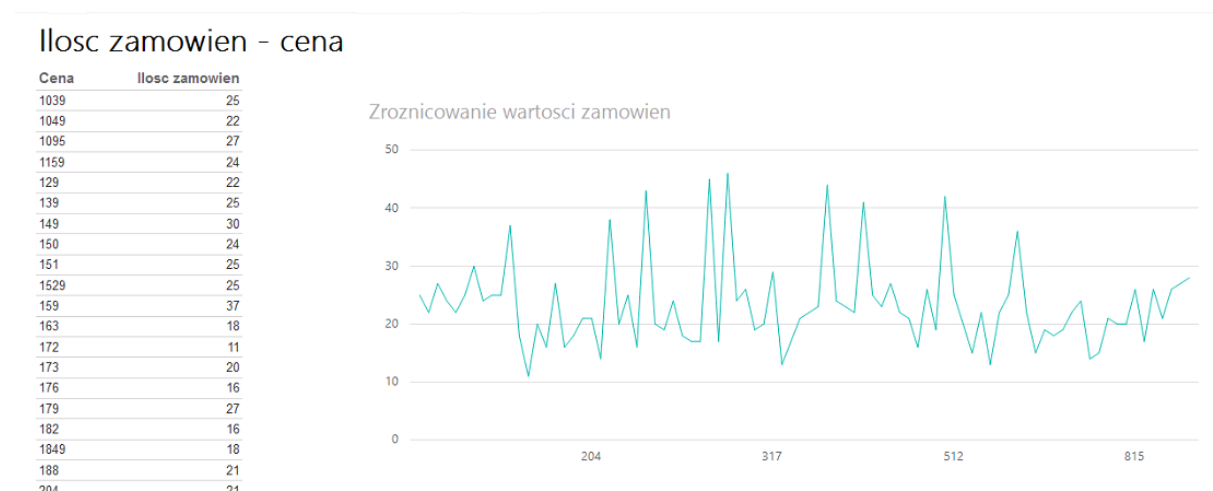
Miesiąc	Ilość zamówień
1	166
10	159
11	133
12	146
2	160
3	181
4	174
5	201
6	165
7	170
8	183
9	162

Ilość zamówień według miesiąca



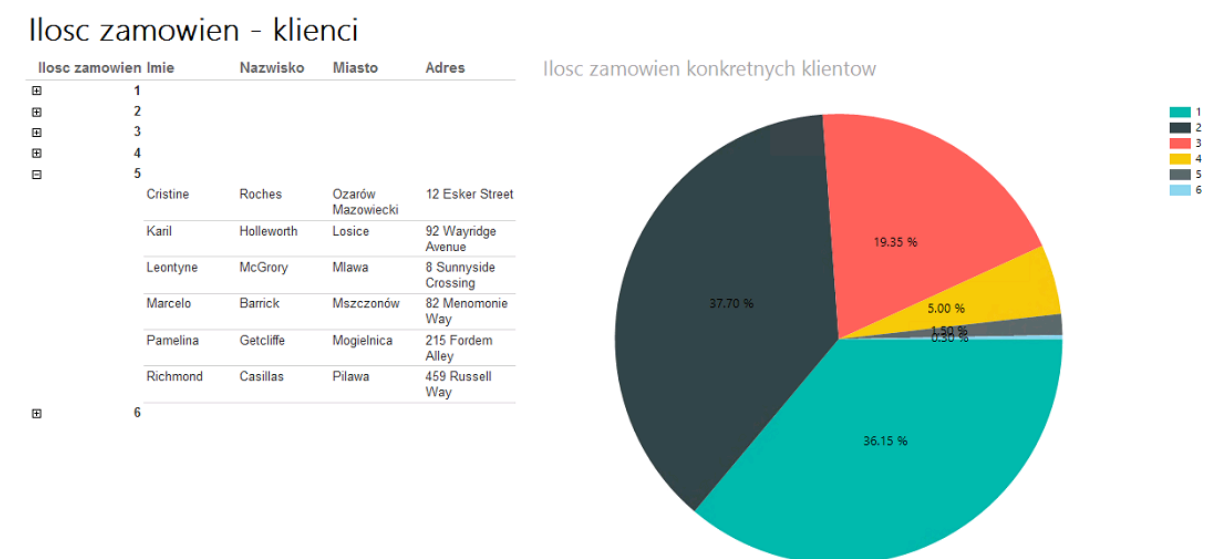
Wykres przedstawia ilość zamówień według miesiąca.

### R03: Analiza wielkości cen podzespołów



Wykres przedstawia zróżnicowanie wartości zamówień.

### R04: Analiza ilości zamówień klientów



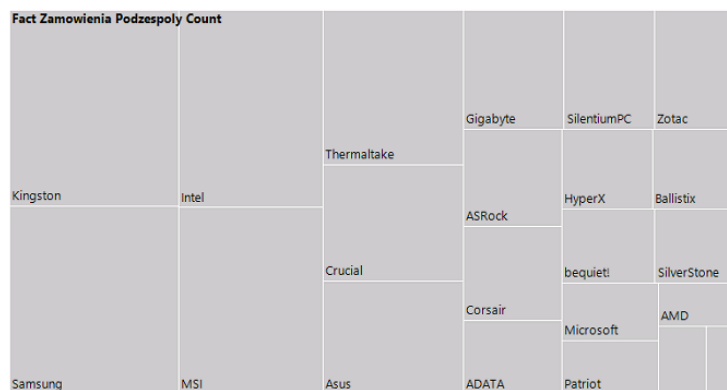
Wykres przedstawia ilość zamówień poszczególnych klientów.

## R05: Analiza wielkości użycia podzespołów według producentów

### Ilość zamówień - producenci

Ilość zamówień	Nazwa
16	Cooler Master
23	Western Digital
25	AMD
37	Patriot
40	Microsoft
44	SilverStone
49	Ballistix
	bequiet!
52	ADATA
	HyperX
66	Corsair
68	ASRock
72	Zotac
77	SilentiumPC
86	Gigabyte
114	Asus
116	Crucial
155	Thermaltake
192	MSI
203	Intel
227	Samsung
237	Kingston

Ilość zamówień poszczególnych producentów



Wykres przedstawia ilość zamówień podzespołów według poszczególnych producentów.

## 7. Job wykonujący nocne przeładowanie procesu ETL

New Job

Select a page

- General
- Steps
- Schedules
- Alerts
- Notifications
- Targets

Script ? Help

Name: HurtowniaDanychJob

Owner: WIN-E6VLVMUO30E\Administrator

Category: [Uncategorized (Local)]

Description: Odswiezanie hurtowni danych

Connection

Server: WIN-E6VLVMUO30E

Connection: WIN-E6VLVMUO30E\Administrator

[View connection properties](#)

☒ Enabled

Progress

Ready

OK Cancel

**New Job Step**

Select a page  
 General  
 Advanced

Script ? Help

Step name:  
 ETL

Type:  
 SQL Server Integration Services Package

Run as:  
 SQL Server Agent Service Account

Set values Verification Command line

Package Configurations Command files Data sources Execution options Logging

Package source: File system

Server:

Log on to the server  
☒ Use Windows Authentication  
☐ Use SQL Server Authentication

User name:

Password:

Package:  
 C:\Package.dtsx

Next Previous

OK Cancel

Connection  
 Server: WIN-E6VLVMUO30E  
 Connection: WIN-E6VLVMUO30E\Administrator  
[View connection properties](#)

Progress  
 Ready

**New Job**

Select a page  
 General  
 Steps  
 Schedules  
 Alerts  
 Notifications  
 Targets

Script ? Help

Schedule list:

ID	Name	Ena...	Description	J...
New	Przeladowanie_w_no...	Yes	Occurs every day at 2:00:00 AM. Schedule will...	

New... Pick... Edit Remove

OK Cancel

Connection  
 Server: WIN-E6VLVMUO30E  
 Connection: WIN-E6VLVMUO30E\Administrator  
[View connection properties](#)

Progress  
 Ready