

# Laboratorium ETL

PROJEKT ZALICZENIOWY

MAREK JĘDRYCHOWSKI, ADRIAN NAPLOCHA

## Spis treści

Zagadnienia teoretyczne – zastosowanie ROLAP w analityce biznesowej .....	
OLAP - wyjaśnienie pojęcia .....	
Czym jest kostka OLAP? .....	
Drill-down (rozwijanie) .....	
Roll up (zwijanie) .....	
Slice and dice (wycinanie) .....	
Pivot(obracanie) .....	
Krótko o typach OLAP .....	
Porównanie rozwiązań MOLAP, ROLAP, HOLAP .....	
MOLAP .....	
ROLAP .....	
HOLAP .....	
Cel realizacji projektu .....	
Opis tabel .....	
Tabela Address .....	
Tabela Contact .....	
Tabela Product .....	
Tabela ProductCategory .....	
Tabela ProductSubcategory .....	
Tabela SalesOrderDetail .....	
Tabela SalesOrderHeader .....	
Tabela ShipMethod .....	
Schemat gwiazdy .....	
Procesy ETL .....	
Tworzenie tabel wymiarów oraz faktów .....	
Tabele wymiarów i faktów .....	
Customer_DIM .....	
Discount_DIM .....	
Products_DIM .....	
Shipment_DIM .....	
Time_DIM .....	
Orders_FACT .....	
Wykorzystanie schematu gwiazdy .....	
Wartości zamówień dla poszczególnych produktów .....	
Liczba zamówień dostarczonych przez poszczególne firmy transportowe .....	
Wartość sprzedaży odnotowanej na poszczególnych typach promocji .....	

Sprzedaż, liczba sprzedanych produktów w przeliczeniu na poszczególne lata i  
miesiące .....  
Jakość danych .....  
Import pliku .....  
Podsumowanie .....

## Zagadnienia teoretyczne – zastosowanie ROLAP w analityce biznesowej

### OLAP - wyjaśnienie pojęcia

OLAP- Online Analytical Processing to zbiór rozwiązań pozwalających na szybką, wielowymiarową analizę wielkich ilości danych pochodzących z hurtowni danych, składnic danych oraz innych ujednoliconych, scentralizowanych baz danych.

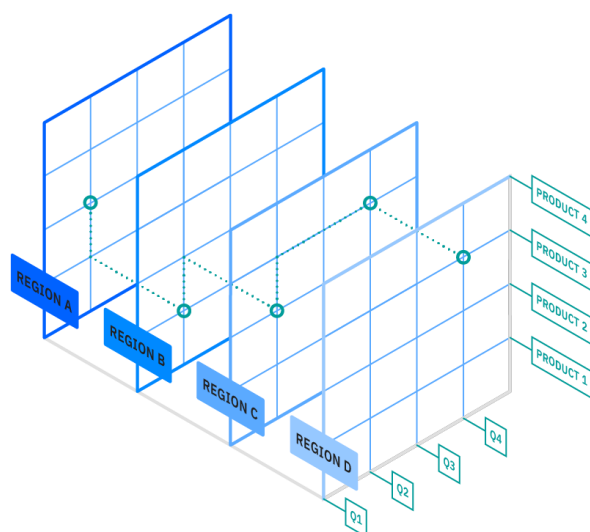
Większość danych biznesowych jest wielowymiarowa – posiada znaczącą liczbę kategorii, wedle których dane są uporządkowane, w celu ich prezentowania, śledzenia oraz analizy. Dla przykładu dane sprzedażowe mogą posiadać kilka wymiarów odpowiadających za lokalizację (region, kraj, województwo, powiat, miasto), czas (rok, miesiąc, tydzień, dzień), produkt (ubranie, płęć, marka, rozmiar, typ).

### Czym jest kostka OLAP?

Rdzeniem większości systemów OLAP jest „Kostka OLAP”. Kostka jest opartą na tablicach wielowymiarowych bazą danych, która umożliwia przetwarzanie i analizę danych zapisanych w postaci wielowymiarowej zdecydowanie szybciej i bardziej efektywnie, niż w tradycyjnej relacyjnej bazie danych.

Tabele w relacyjnych bazach danych posiadają postać zbliżoną do arkusza przechowyującego pojedyncze rekordy w wierszach i kolumnach. Każdy “fakt” taki jak kraj, region zamieszkania umieszczony jest pośród danych znajduje się na przecięciu wiersza i kolumny. Bazy relacyjne pozwalają na budowę raportów i kwerend dla danych, jednak wydajność rozwiązania spada wraz ze wzrostem wielkości bazy. W celu rozwiązania problemów tego typu powstało OLAP. Kostka rozbudowuje pojedynczą tabelę poprzez dodanie kolejnych warstw, z których każda dodaje kolejne wymiary, zwyczajowo kolejny poziom w hierarchii tabeli. Dla przykładu górna warstwa kostki może organizować sprzedaż wedle regionu, kolejne wedle kraju, województwa etc.

W teorii kostka może posiadać nieograniczoną liczbę warstw (mniejsze kostki mogą czasem zawierać się w poszczególnych warstwach większej) dla przykładu każda warstwa sklepu mogłaby porządkować sprzedaż wedle pracownika i produktu. W praktyce jednak tworzy się jedynie kostki, które pozwalają na zoptymalizowaną analizę i wysoką wydajność operacji.



Rysunek 1 Wizualizacja kostki OLAP

Kostki OLAP pozwalają na cztery podstawowe typy wielowymiarowej analizy danych:

#### Drill-down (rozwijanie)

Operacja rozwijania konwertuje mniej szczegółowe dane do postaci bardziej szczegółowej z wykorzystaniem jednej z dwóch metod. Przejścia w głąb hierarchii lub dodania nowego wymiaru do tabeli. Dla przykładu przeglądając dane sprzedażowe organizacji dla danego kwartału możliwe jest przejście do podglądu danych na każdy pojedynczy miesiąc, schodząc tym samym w dół hierarchii po wymiarze czasu.

#### Roll up (zwijanie)

Operacja zwijania pozwala na agregację danych w kostce, poprzez przejście w górę hierarchii lub przez redukcję liczby wymiarów. Jako przykład można podać wgląd w dane dotyczące całego państwa, zamiast poszczególnych miast.

#### Slice and dice (wycinanie)

Operacja pozwala na utworzenie pod-kostki powstałej poprzez wybranie pojedynczego wymiaru spośród tych dostępnych w kostce głównej. Przykładowo wybranie wszystkich danych dostępnych na temat pojedynczego kwartału.

#### Pivot(obracanie)

Obracanie zmienia podgląd kostki, w celu przedstawienia nowej reprezentacji danych. Pozwala to na przegląd poszczególnych wymiarów kostki. Funkcja ta zbliżona jest do pivot dostępnego w excelu. Funkcja excela posiada jednak szybszy czas reakcji i większą wydajność kwerend.

## Krótko o typach OLAP

### Porównanie rozwiązań MOLAP, ROLAP, HOLAP



Rysunek 2 Porównanie rozwiązań OLAP

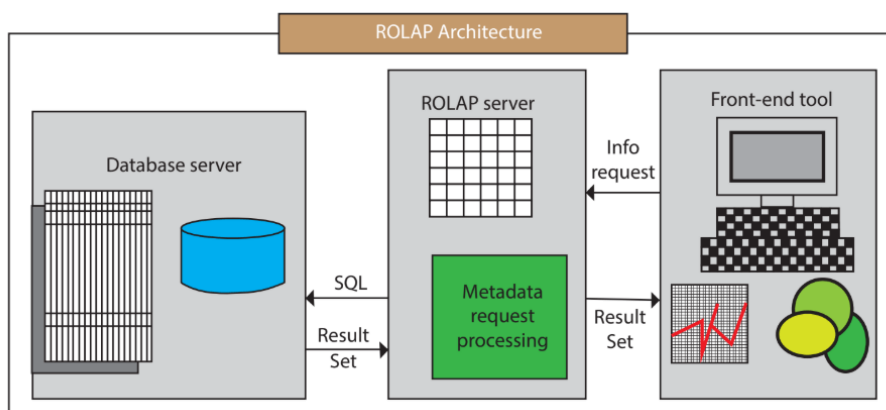
#### MOLAP

Znany jest pod nazwą Multidimensional OLAP (MOLAP). Dla wielu zastosowań MOLAP jest najszybszym i najbardziej praktycznym typem wielowymiarowej analizy danych.

Istnieją jednak również inne rozwiązania OLAP, które dla niektórych zastosowań mogą być preferowane.

#### ROLAP

ROLAP (Relational OLAP) jest to narzędzie wielowymiarowej analizy danych, działające bezpośrednio na danych umieszczonych w tabelach relacyjnych, bez wcześniejszego przekazania ich do kostki, wykorzystujące język SQL. Wysoki poziom skomplikowania tworzonych zapytań może jednak prowadzić do spadków wydajności. Kolejną wadą rozwiązania jest statyczny wgląd w dane, co oznacza, że nie może on zostać zmodyfikowany w celu prezentacji innego podglądu danych. Rozwiązania ROLAP najlepiej spisują się więc w sytuacjach, kiedy możliwość pracy bezpośrednio na wielkich zbiorach danych jest ważniejsza od wydajności czy elastyczności.

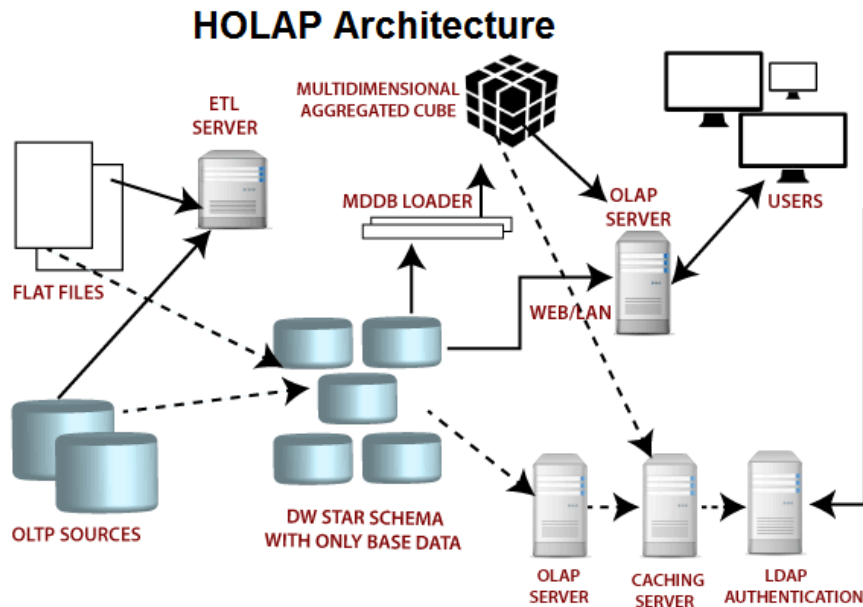


Rysunek 3 Schemat architektury ROLAP

#### HOLAP

HOLAP (Hybrid OLAP) powstał, aby umożliwić optymalny podział pracy, pomiędzy relacyjnymi, a wielowymiarowymi bazami danych z wykorzystaniem pojedynczej struktury

OLAP. Tabele relacyjne zawierają większe ilości danych, natomiast kostki OLAP wykorzystywane są do agregacji. Narzędzie HOLAP może przejść z kostki OLAP do tabel relacyjnych, co umożliwia szybkie przetwarzanie danych i elastyczny dostęp. System hybrydowy pozwala na większą skalowalność, ale jednocześnie nie jest wolny od nieuniknionego spowolnienia podczas przetwarzania tabel relacyjnych. Wysoki stopień komplikacji architektury wymusza dokonywanie częstszych aktualizacji i kładzenie większego nacisku na konserwację. Przekłada się to na większy koszt rozwiązań HOLAP.



Rysunek 4 Schemat architektury HOLAP

## Cel realizacji projektu

Niniejszy projekt powstał na potrzeby zaprezentowania wykorzystania narzędzia SAS Data Integration Studio. W projekcie dokonano analizy danych sprzedażowych fikcyjnego przedsiębiorstwa Adventure Works. Utworzone przykładowe procesy mogłyby zostać wykorzystane w celu przekazania decydującym wiedzy, odpowiedniej dla celów zarządzania przedsiębiorstwem. Wszystkie dane wykorzystane w projekcie zawarte są w bazie Adventure Works. Przedsiębiorstwo to zajmujące się produkcją oraz sprzedażą rowerów i części rowerowych. Sprzedaż dokonywana jest w sklepach stacjonarnych, jak również z wykorzystaniem serwisów internetowych.

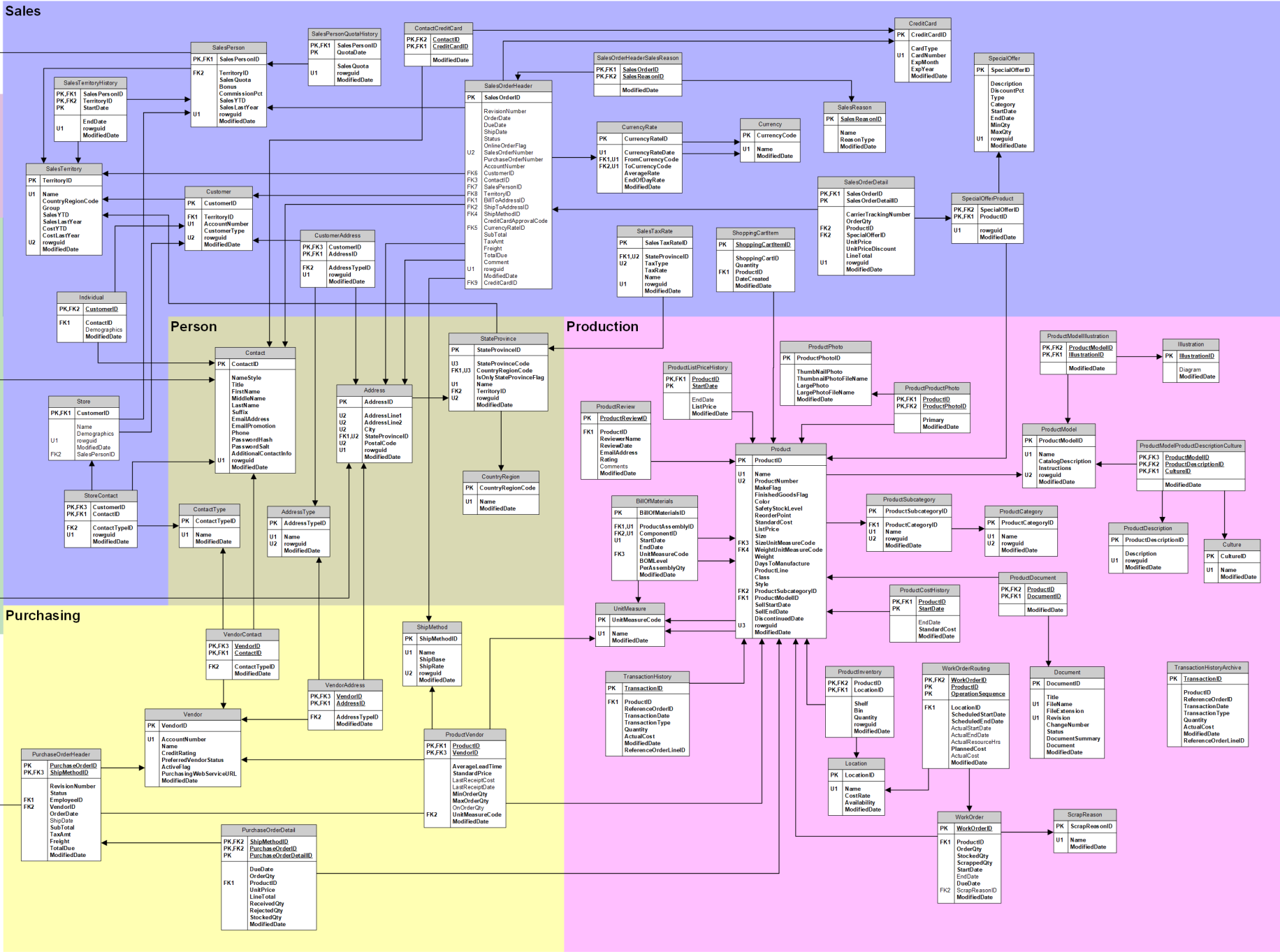
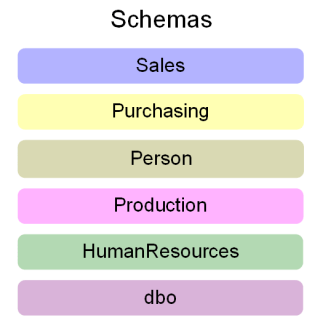
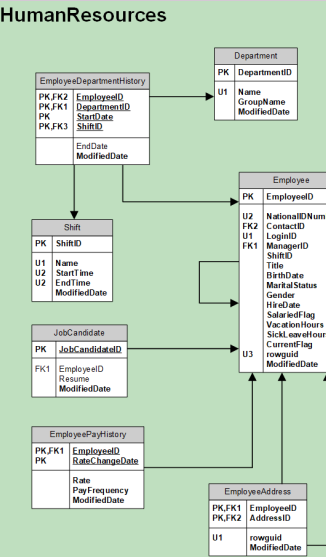
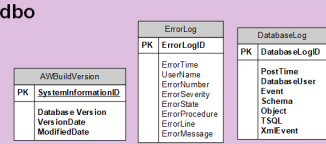
Tabele bazy danych podzielone są na 5 głównych grup w zależności od tematyki danych w niej zawartych:

- Sales
- Production
- Person
- Human Resources
- Purchasing



AdventureWorks OLTP Schema  
November 2005

Best Print Results if  
11X17 paper  
Landscape  
Fit to 1 sheet



Rysunek 5 Schemat bazy Adventure Works

## Opis tabel

W tej sekcji opisana zostanie zawartość tabel wykorzystanych do realizacji projektu

### Tabela Address

Tabela Address składa się z 8 kolumn i zawiera dane adresowe dla każdego klienta, pracownika oraz partnera firmy Adventure Works.

Tabela 1 Address

Column	Data type	Nullability	Description
AddressID	int	Not null	Primary key for address rows.
AddressLine1	nvarchar(60)	Not null	First street-address line.
AddressLine2	nvarchar(60)	Null	Second street address line.
City	nvarchar(30)	Not null	Name of the city.
StateProvinceID	int	Not null	Unique identification number for the state or province. Foreign key to StateProvince.StateProvinceID.
PostalCode	nvarchar(15)	Not null	Postal code for the street address.
rowguid	uniqueidentifier ROWGUIDCOL	Not null	ROWGUIDCOL number uniquely identifying the row. Used to support a merge replication sample.
ModifiedDate	datetime	Not null	Date and time the row was last updated.

### Tabela Contact

Tabela CONTACT składa się z 15 kolumn i zawiera listę nazwisk oraz innych powiązanych danych dla każdego klienta, pracownika oraz partnera.

Tabela 2 Contact

Column	Data type	Nullability	Description
ContactID	int	Not null	Primary key for Contact rows.
NameStyle	bit	Not null	Indicates the column in which the family name is stored.  0 = Family name is stored in LastName and the first name is stored in FirstName.  1 = Family name is stored in FirstName and the first name is stored in LastName.
Title	nvarchar	Null	A courtesy title. For example, Mr. or Ms.
FirstName	nvarchar(50)	Not null	First name of the person.
MiddleName	nvarchar(50)	Null	Middle name or middle initial of the person.
LastName	nvarchar(50)	Not null	Last name of the person.
Suffix	nvarchar(10)	Null	Surname suffix. For example, Sr. or Jr.
EmailAddress	nvarchar(50)	Null	E-mail address for the person.
EmailPromotion	int	Null	0 = Contact does not wish to receive e-mail promotions.

			1 = Contact does wish to receive e-mail promotions.
Phone	nvarchar(25)	Null	Phone number associated with the person.
PasswordHash	varchar(40)	Not null	Password for the email account.
PasswordSalt	varchar(10)	Not null	Random value concatenated with the password string before the password is hashed.
AdditionalContactInfo	xml	Null	Additional contact information about the person, such as additional phone numbers and addresses or special instructions and notes. Data is stored in XML format.
Rowguid	uniqueidentifier ROWGUIDCOL	Not null	ROWGUIDCOL number uniquely identifying the row. Used to support a merge replication sample.
ModifiedDate	datetime	Not null	Date and time the row was last updated.

### Tabela Product

Tabela PRODUCT składa się z 25 kolumn i zawiera informacje o produktach sprzedażowych oraz produktach niehandlowych wykorzystywanych dla celów produkcji w firmie Adventure Works.

Tabela 3 Product

Column	Data type	Nullability	Description
ProductID	int	Not null	Primary key for Product rows.
Name	nvarchar(50)	Not null	Name of the product.
ProductNumber	nvarchar(25)	Not null	Unique product identification number.
MakeFlag	bit	Not null	0 = Product is purchased. 1 = Product is manufactured in-house.
FinishedGoodsFlag	bit	Not null	0 = Product is not a salable item. 1 = Product is salable.
Color	nvarchar(15)	Null	Product color.
SafetyStockLevel	smallint	Not null	Minimum inventory quantity.
ReorderPoint	smallint	Not null	Inventory level that triggers a purchase order or work order.
StandardCost	money	Not null	Standard cost of the product.
ListPrice	money	Not null	Selling price.
Size	nvarchar(5)	Null	Product size.
SizeUnitMeasureCode	nchar(3)	Null	Unit of measure for Size column.
WeightUnitMeasureCode	nchar(3)	Null	Unit of measure for Weight column.
Weight	decimal (8,2)	Null	Product weight.
DaysToManufacture	int	Not null	Number of days required to manufacture the product.

ProductLine	nchar(2)	Null	R = Road M = Mountain T = Touring S = Standard
Class	nchar(2)	Null	H = High M = Medium L = Low
Style	nchar(2)	Null	W = Women's M = Men's U = Universal
ProductSubcategoryID	smallint	Null	Product is a member of this product subcategory. Foreign key to ProductSubCategory.ProductSubCategoryID.
ProductModelID	int	Null	Product is a member of this product model. Foreign key to ProductModel.ProductModelID.
SellStartDate	datetime	Not null	Date the product was available for sale.
SellEndDate	datetime	Null	Date the product was no longer available for sale.
DiscontinuedDate	datetime	Null	Date the product was discontinued.
rowguid	uniqueidentifier ROWGUIDCOL	Not null	ROWGUIDCOL number uniquely identifying the row. Used to support a merge replication sample
ModifiedDate	datetime	Not null	Date and time the row was last updated.

### Tabela ProductCategory

Tabela PRODUCTCATEGORY składa się z 4 kolumn opisujących wszystkie wysokopoziomowe kategorie produktów w bazie danych Adventure Works.

*Tabela 4 ProductCategory*

Column	Data type	Nullability	Description
ProductCategoryID	int	Not null	Primary key for ProductCategory rows.
Name	nvarchar(50)	Not null	Category description.
rowguid	uniqueidentifier ROWGUIDCOL	Not null	ROWGUIDCOL number uniquely identifying the row. Used to support a merge replication sample.
ModifiedDate	datetime	Not null	Date and time the row was last updated.

### Tabela ProductSubcategory

Tabela PRODUCTSUBCATEGORY składa się z 5 kolumn opisujących wszystkie podkategorie produktów w bazie danych Adventure Works.

*Tabela 5 ProductSubcategory*

Column	Data type	Nullability	Description
ProductSubcategoryID	smallint	Not null	Primary key for ProductSubcategory rows.
ProductCategoryID	int	Not null	Product category identification number. Foreign key to ProductCategory.ProductCategoryID.
Name	nvarchar(50)	Not null	Subcategory description.
rowguid	uniqueidentifier ROWGUIDCOL	Not null	ROWGUIDCOL number uniquely identifying the row. Used to support a merge replication sample.

ModifiedDate	datetime	Not null	Date and time the row was last updated.
--------------	----------	----------	---

#### Tabela SalesOrderDetail

Tabela SALESORDERDETAIL składa się z 11 kolumn. Zawiera powiązanie zamówienia z produktami zamawianymi. Każde zamówienie może składać się z jednego lub więcej produktów.

*Tabela 6 salesOrderDetail*

Column	Data type	Nullability	Description
SalesOrderID	int	Not null	Primary key. Foreign key to SalesOrderHeader.SalesOrderID.
SalesOrderDetailID	int	Not null	Primary key. A sequential number used to ensure data uniqueness.
CarrierTrackingNumber	nvarchar(25)	Null	Shipment tracking number supplied by the shipper.
OrderQty	smallint	Not null	Quantity ordered per product.
ProductID	int	Not null	Product sold to customer. Foreign key to Product.ProductID.
SpecialOfferID	int	Not null	Promotional code. Foreign key to SpecialOffer.SpecialOfferID.
UnitPrice	money	Not null	Selling price of a single product.
UnitPriceDiscount	money	Not null	Discount amount.
LineTotal	money	Not null	Per product subtotal. Computed as OrderQty * UnitPrice
rowguid	uniqueidentifier ROWGUIDCOL	Not null	ROWGUIDCOL number uniquely identifying the row. Used to support a merge replication sample.
ModifiedDate	datetime	Not null	Date and time the row was last updated.



## Tabela SalesOrderHeader

Tabela SALESORDERHEADER składa się z 27 kolumn. Dane w niej zawarte opisują nagłówki zamówień. Dokładna specyfikacja zamówień znajduje się w tabeli SALESORDERDETAIL.

Tabela 7 SalesOrderHeader

Column	Data type	Nullability	Description
SalesOrderID	int	Not null	Primary key
RevisionNumber	tinyint	Not null	Incremental number to track changes to the sales order over time.
OrderDate	datetime	Not null	Dates the sales order was created.
DueDate	Datetime	Not null	Date the order is due to the customer.
ShipDate	Datetime	Null	Date the order was shipped to the customer.
Status	tinyint	Not null	Order current status. 1 = In process 2 = Approved 3 = Back ordered 4 = Rejected 5 = Shipped 6 = Canceled
OnlineOrderFlag	bit	Not null	0 = Order placed by sales person. 1 = Order placed online by customer.
SalesOrderNumber	nvarchar(25)	Not null	Unique sales order identification number.
PurchaseOrderNumber	nvarchar(25)	Null	Customer purchase order number reference.
AccountNumber	nvarchar(15)	Null	Financial accounting number reference.
CustomerID	int	Not null	Customer identification number. Foreign key to Customer.CustomerID.
ContactID	int	Not null	Customer contact identification number. Foreign key to Contact.ContactID

SalesPersonID	int	Null	Sales person who created the sales order. Foreign key to SalesPerson.SalePersonID.
TerritoryID	int	Null	Territory in which the sale was made. Foreign key to SalesTerritory.SalesTerritoryID.
BillToAddressID	int	Not null	Customer billing address. Foreign key to Address.AddressID.
ShipToAddressID	Int	Not null	Customer shipping address. Foreign key to Address.AddressID.
ShipMethodID	Int	Not null	Shipping method. Foreign key to ShipMethod.ShipMethodID.
CreditCardID	Int	Null	Credit card identification number. Foreign key to CreditCard.CreditCardID.
CreditCardApprovalCode	varchar(15)	Null	Approval code provided by the credit card company.
CurrencyRateID	int	Null	Currency exchange rate used. Foreign key to CurrencyRate.CurrencyRateID.
SubTotal	money	Not null	Sales subtotal. Computed as SUM(SalesOrderDetail.LineTotal) for the SalesOrderID.
TaxAmt	Money	Not null	Tax amount.
Freight	Money	Not null	Shipping cost.
TotalDue	Money	Not null	Total due from customer. Computed as SubTotal + TaxAmt + Freight
Comment	nvarchar(128)	Null	Comments made by the sales representative.
rowguid	uniqueidentifier ROWGUIDCOL	Not null	ROWGUIDCOL number uniquely identifying the row. Used to support a merge replication sample.
ModifiedDate	datetime	Not null	Date and time the row was last updated.

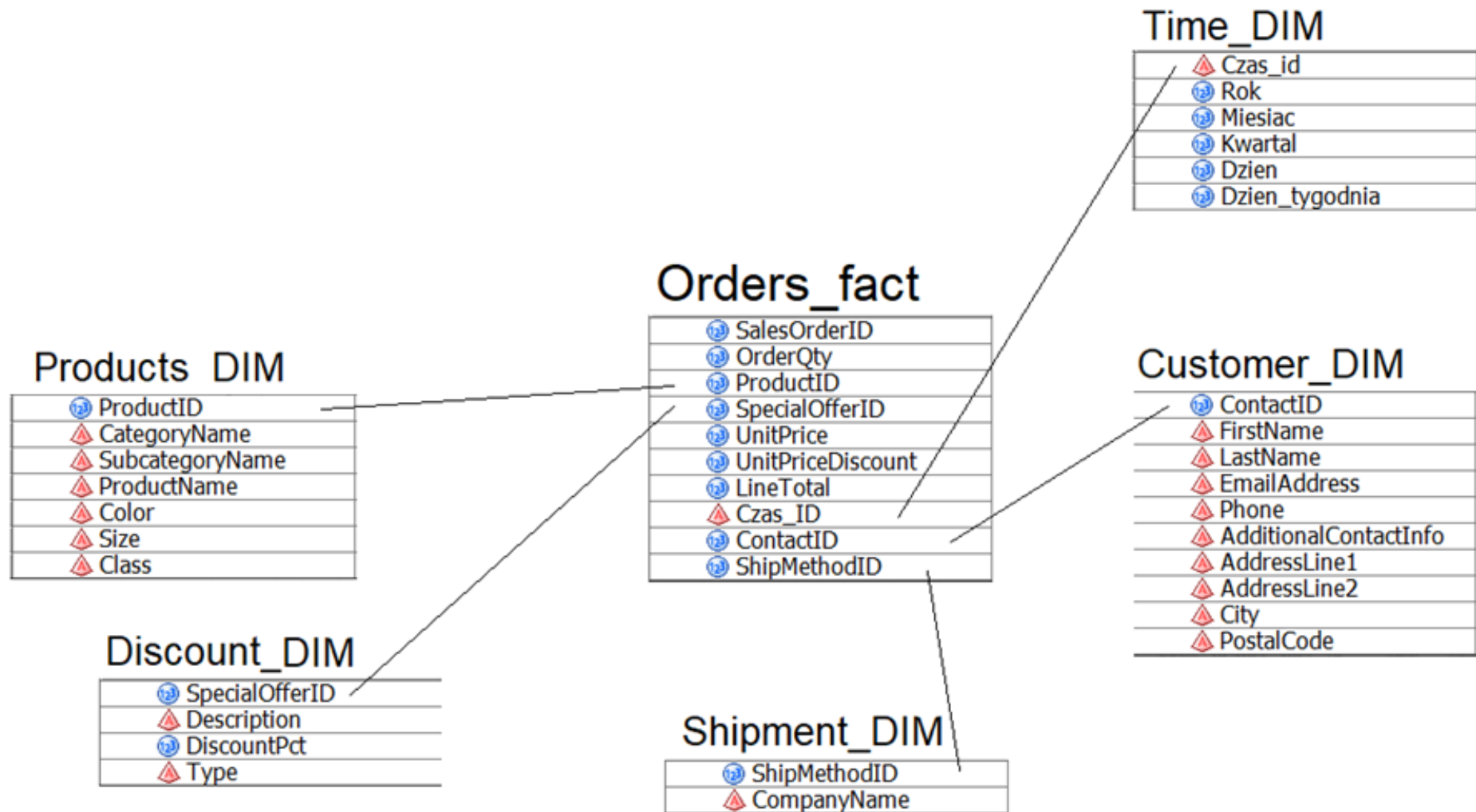
## Tabela ShipMethod

Tabela SHIPMETHOD składa się z 6 kolumn. Dane opisują firmy współpracujące z organizacją Adventure Works Cycles w zakresie dostaw oraz przewozów zamówień.

Tabela 8 ShipMethod

Column	Data type	Nullability	Description
ShipMethodID	int	Not null	Primary key.
Name	nvarchar(50)	Not null	Shipping company name.
ShipBase	money	Not null	Minimum shipping charge..
ShipRate	money	Not null	Shipping charge per pound
rowguid	uniqueidentifier ROWGUIDCOL	Not null	ROWGUIDCOL number uniquely identifying the row. Used to support a merge replication sample.
ModifiedDate	datetime	Not null	Date and time the row was last updated.

## Schemat gwiazdy



Rysunek 6 Schemat gwiazdy

## Procesy ETL

### Tworzenie tabel wymiarów oraz faktów

#### Tabele wymiarów i faktów

Model gwiazdy składa się z 6 tabel wymiarów:

- Customer\_DIM
- Time\_DIM
- Discount\_DIM
- Products\_DIM
- Shipment\_DIM
- Orders\_fact

Tabele modelu powstały z wykorzystaniem następujących procesów:

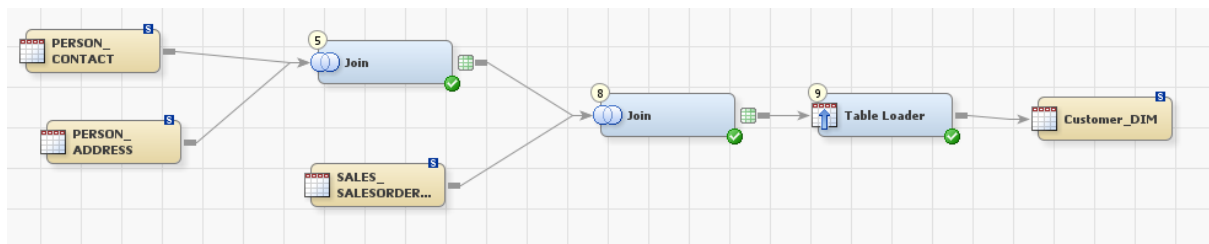
loader\_customer  
loader\_discount  
loader\_fact\_table  
loader\_products  
loader\_shipment  
loader\_time

Rysunek 7 Procesy tworzenia tabel dla schematu gwiazdy

#### Customer\_DIM

Pierwsza opisana tabelą wymiaru zawiera informacje o kliencie składającym zamówienie. Powstała ona dzięki połączeniu trzech tabel źródłowych.

- Person\_CONTACT
- Person\_ADDRESS
- Sales\_SalesOrderHeader



Rysunek 8 Przebieg procesu loader\_customer

Informacje na temat osób pochodzą z tabeli CONTACT, a dane adresowe z tabeli ADDRESS. Połączenie następuje z wykorzystaniem pól ContactID tabeli Contact, AddressID tabeli ADDRESS, oraz ContactID tabeli SalesOrderHeader. Strukturę tabeli opisują rysunki widoczne poniżej.

Customer_DIM Properties									
General Columns Indexes Keys Parameters Physical Storage Options Notes Extended Attributes Authorization									
#	Name	Description	Type	Len...	Informat	Format	Is Nullable	Summary ...	Sort Ord
1	ContactID	ContactID	Numeric	8	11.	11.	No	(None)	(None)
2	FirstName	FirstName	Character	1024	\$1024.	\$1024.	Yes	(None)	(None)
3	LastName	LastName	Character	1024	\$1024.	\$1024.	Yes	(None)	(None)
4	EmailAddress	EmailAddress	Character	1024	\$1024.	\$1024.	Yes	(None)	(None)
5	Phone	Phone	Character	1024	\$1024.	\$1024.	Yes	(None)	(None)
6	AdditionalContactInfo	AdditionalCo...	Character	1024	\$1024.	\$1024.	Yes	(None)	(None)
7	AddressLine1	AddressLine1	Character	1024	\$1024.	\$1024.	Yes	(None)	(None)
8	AddressLine2	AddressLine2	Character	1024	\$1024.	\$1024.	Yes	(None)	(None)
9	City	City	Character	1024	\$1024.	\$1024.	Yes	(None)	(None)
10	PostalCode	PostalCode	Character	1024	\$1024.	\$1024.	Yes	(None)	(None)

Rysunek 9 Pola tabeli Customer\_DIM

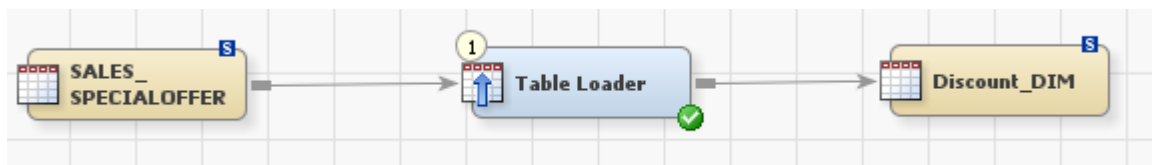
Utworzona tabela docelowa zawiera 10 kolumn, w tym ContactID pełniącą rolę klucza.

#	ContactID	FirstName	LastName	EmailAddress	Phone	AdditionalContactInfo	AddressLine1	AddressLine2	City	PostalCode
1	1	Gustavo	Achong	gustavo0@advent...	398-555-01...	?<AdditionalContactInfo x...	Mall Of Memphis		Memphis	38103
2	2	Catherine	Abel	catherine0@adve...	747-555-01...	?<AdditionalContactInfo x...	57251 Serene Blvd...		Van Nuys	91411
3	3	Kim	Abercrombie	kim2@adventure...	334-555-01...	?<AdditionalContactInfo x...	Tanger Factory		Branch	55056
4	4	Humberto	Acevedo	humberto0@adve...	599-555-01...	?<AdditionalContactInfo x...	Johnny Appleseed		Mansfield	44903
5	5	Pilar	Ackerman	pilar1@adventure...	1 (11) 500	?<AdditionalContactInfo x...	4250 Concord Roa...		Rhodes	2138
6	6	Frances	Adams	frances0@advent...	991-555-01...	?<AdditionalContactInfo x...	6900 Sisk Road		Modesto	95354
7	7	Margaret	Smith	margaret0@adven...	959-555-01...	?<AdditionalContactInfo x...	Lewiston Mall		Lewiston	83501
8	8	Carla	Adams	carla0@adventure...	107-555-01...	?<AdditionalContactInfo x...	Leesburg Premium		Leesburg	20176
9	9	Jay	Adams	jay1@adventure...	158-555-01...	?<AdditionalContactInfo x...	Blue Ridge Mall		Kansas Ci...	64106
10	10	Ronald	Adina	ronald0@adventur...	453-555-01...	?<AdditionalContactInfo x...	Hilton Head Factor...		Bluffton	29910
11	11	Samuel	Agcaoli	samuel0@adventu...	554-555-01...		No. 25800-130 Kin...		Toronto	M4B 1V5
12	12	James	Agular	james2@adventur...	1 (11) 500		Knaackstr 7		Paderbor...	33041
13	13	Robert	Ahlering	robert1@adventur...	678-555-01...		6500 East Grant R...		Tucson	85701
14	14	François	Ferrier	francois1@advent...	571-555-01...		Eastridge Mall		Casper	82601
15	15	Kim	Akers	kim3@adventure...	440-555-01...		Granite State Mark...		Hooksett	03106
16	16	Amy	Alberts	amy1@adventure...	727-555-01...		252851 Rowan Pla...		Richmond...	V6B 3P7
17	17	Anna	Albright	anna0@adventure...	197-555-01...		Flagler Park Plaza		Miami	33127
18	18	Milton	Albury	milton0@adventur...	492-555-01...		Wrentham Village		Wrentha...	02093
19	19	Paul	Alcorn	paul2@adventure...	331-555-01...		White Mountain M...		Rock Spri...	82901
20	20	Gregory	Alderson	gregory0@advent...	968-555-01...		25981 College Stre...		Montreal	H1Y 2H5
21	21	Phillip	Alexander	jphillp0@adventur...	845-555-01...		Belz Factory Outle...		Pigeon F...	37863
22	22	Michelle	Alexander	michelle0@advent...	115-555-01...		22589 West Craig		North Las...	89030
23	23	Sean	Jacobson	sean2@adventure...	555-555-01...		2551 East Warner		Gilbert	85233
24	24	Phyllis	Allen	phyllis0@adventur...	695-555-01...		Gulfport Factory S...		Gulfport	39501
25	25	Marvin	Allen	marvin0@adventu...	447-555-01...		First Colony Mall		Sugar La...	77478
26	26	Michael	Allen	michael10@adven...	607-555-01...		Lake Region Facto...		Tilton	03276

Rysunek 10 Dane przykładowe z tabeli Customer\_DIM

## Discount\_DIM

Następna tabela zawiera informacje o wartościach zniżek przyznawanych na poszczególne produkty. Proces Loader\_Discount przedstawiony poniżej przedstawia przeładowanie tabeli SALES\_SPECIALOFFER przy użyciu węzła TABLE\_LOADER. Węzeł Table\_Loader pozwala nam na wybór stosownych kolumn potrzebnych do utworzenia nowej tabeli.



Rysunek 11 Przebieg procesu loader\_discount

W tabeli Discount\_DIM zawarte jest mniej kolumn, niż w bazie Adventure Works. Schemat poniżej obrazuje strukturę nowo utworzonej Discount\_DIM. Rolę klucza głównego pełni kolumna SpecialOfferID.

Discount_DIM Properties									
General Columns Indexes Keys Parameters Physical Storage Options Notes Extended Attributes Authorization									
#	Name	Description	Type	Len...	Informat	Format	Is Nullable	Summary ...	Sort Ord
1	SpecialOfferID	SpecialOfferID	Numeric	8	11.	11.	No	(None)	(None)
2	Description	Description	Character	1024	\$1024.	\$1024.	Yes	(None)	(None)
3	DiscountPct	DiscountPct	Numeric	8	DOLLA...	DOLLA...	Yes	(None)	(None)
4	Type	Type	Character	1024	\$1024.	\$1024.	Yes	(None)	(None)

Rysunek 12 Pola tabeli Discount\_DIM

View Data: Discount_DIM (16 rows)				
#	SpecialOfferID	Description	DiscountPct	Type
1	1	No Discount	\$0.00	No Discount
2	2	Volume Discount 11 to 14	\$0.02	Volume Discount
3	3	Volume Discount 15 to 24	\$0.05	Volume Discount
4	4	Volume Discount 25 to 40	\$0.10	Volume Discount
5	5	Volume Discount 41 to 60	\$0.15	Volume Discount
6	6	Volume Discount over 60	\$0.20	Volume Discount
7	7	Mountain-100 Clearance Sale	\$0.35	Discontinued Product
8	8	Sport Helmet Discount-2002	\$0.10	Seasonal Discount
9	9	Road-650 Overstock	\$0.30	Excess Inventory
10	10	Mountain Tire Sale	\$0.50	Excess Inventory
11	11	Sport Helmet Discount-2003	\$0.15	Seasonal Discount
12	12	LL Road Frame Sale	\$0.35	Excess Inventory
13	13	Touring-3000 Promotion	\$0.15	New Product
14	14	Touring-1000 Promotion	\$0.20	New Product
15	15	Half-Price Pedal Sale	\$0.50	Seasonal Discount
16	16	Mountain-500 Silver Clearance Sale	\$0.40	Discontinued Product

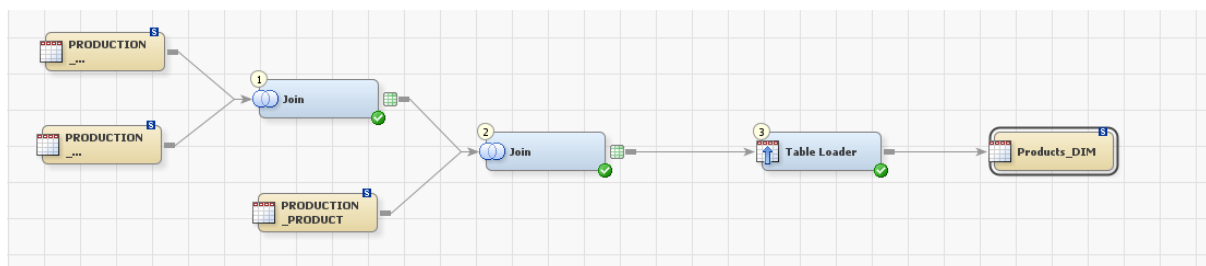
Rysunek 13 Pełne dane z tabeli Discount\_DIM

## Products\_DIM

Utworzona tabela zawiera informacje na temat produktów sprzedawanych przez firmę Adventure Works. Powstaje ona poprzez połączenie trzech tabel:

- Production\_PRODUCT
- Production\_PRODUCTCATEGORY
- Production\_PRODUCTSUBCATEGORY

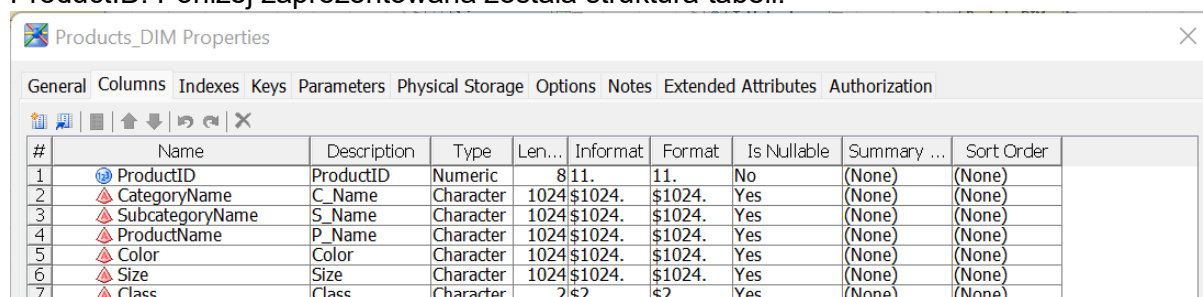
Przeładowanie danych do tabeli docelowej zaprezentowano poniżej.



Rysunek 14 Przebieg procesu loader\_products

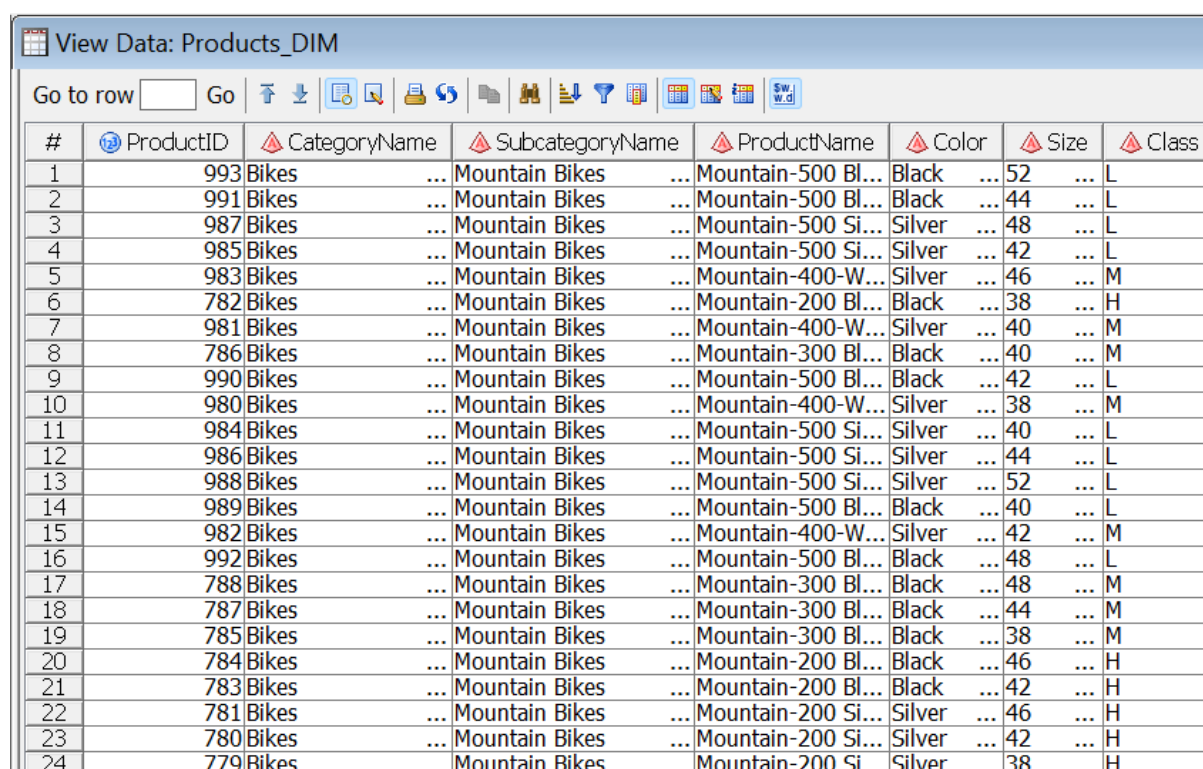


Tabela Products\_DIM składa się z 8 kolumn. Jako klucz główny tabeli wykorzystano ProductID. Poniżej zaprezentowana została struktura tabeli.



#	Name	Description	Type	Len...	Informat	Format	Is Nullable	Summary ...	Sort Order
1	ProductID	ProductID	Numeric	8	11.	11.	No	(None)	(None)
2	CategoryName	C_Name	Character	1024	\$1024.	\$1024.	Yes	(None)	(None)
3	SubcategoryName	S_Name	Character	1024	\$1024.	\$1024.	Yes	(None)	(None)
4	ProductName	P_Name	Character	1024	\$1024.	\$1024.	Yes	(None)	(None)
5	Color	Color	Character	1024	\$1024.	\$1024.	Yes	(None)	(None)
6	Size	Size	Character	1024	\$1024.	\$1024.	Yes	(None)	(None)
7	Class	Class	Character	2	\$2.	\$2.	Yes	(None)	(None)

Rysunek 15 Pola tabeli Products\_DIM

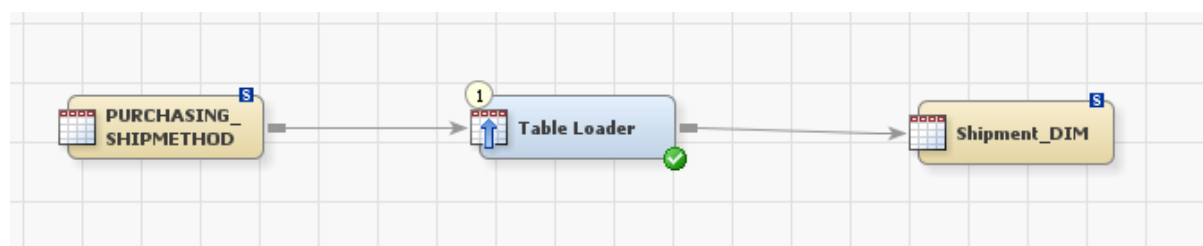


#	ProductID	CategoryName	SubcategoryName	ProductName	Color	Size	Class
1	993	Bikes	Mountain Bikes	Mountain-500 Bl...	Black	52	L
2	991	Bikes	Mountain Bikes	Mountain-500 Bl...	Black	44	L
3	987	Bikes	Mountain Bikes	Mountain-500 Si...	Silver	48	L
4	985	Bikes	Mountain Bikes	Mountain-500 Si...	Silver	42	L
5	983	Bikes	Mountain Bikes	Mountain-400-W...	Silver	46	M
6	782	Bikes	Mountain Bikes	Mountain-200 Bl...	Black	38	H
7	981	Bikes	Mountain Bikes	Mountain-400-W...	Silver	40	M
8	786	Bikes	Mountain Bikes	Mountain-300 Bl...	Black	40	M
9	990	Bikes	Mountain Bikes	Mountain-500 Bl...	Black	42	L
10	980	Bikes	Mountain Bikes	Mountain-400-W...	Silver	38	M
11	984	Bikes	Mountain Bikes	Mountain-500 Si...	Silver	40	L
12	986	Bikes	Mountain Bikes	Mountain-500 Si...	Silver	44	L
13	988	Bikes	Mountain Bikes	Mountain-500 Si...	Silver	52	L
14	989	Bikes	Mountain Bikes	Mountain-500 Bl...	Black	40	L
15	982	Bikes	Mountain Bikes	Mountain-400-W...	Silver	42	M
16	992	Bikes	Mountain Bikes	Mountain-500 Bl...	Black	48	L
17	788	Bikes	Mountain Bikes	Mountain-300 Bl...	Black	48	M
18	787	Bikes	Mountain Bikes	Mountain-300 Bl...	Black	44	M
19	785	Bikes	Mountain Bikes	Mountain-300 Bl...	Black	38	M
20	784	Bikes	Mountain Bikes	Mountain-200 Bl...	Black	46	H
21	783	Bikes	Mountain Bikes	Mountain-200 Bl...	Black	42	H
22	781	Bikes	Mountain Bikes	Mountain-200 Si...	Silver	46	H
23	780	Bikes	Mountain Bikes	Mountain-200 Si...	Silver	42	H
24	779	Bikes	Mountain Bikes	Mountain-200 Si...	Silver	38	H

Rysunek 16 Dane przykładowe z tabeli Products\_DIM

## Shipment\_DIM

Kolejna tabela – Shipment\_DIM zawiera informacje na temat firm tranzytowych wykorzystanych przez przedsiębiorstwo na potrzeby realizacji zamówień. Tabela powstała, dzięki przeładowaniu tabeli Purchasing\_ShipMethod. Rolę klucza głównego pełni kolumna ShipMethodID.



Rysunek 17 Przebieg procesu loader\_shipment

Tabela docelowa posiada 2 kolumny opisujące ID firmy wysyłkowej oraz jej nazwę.



View Data: Shipment_DIM (5 rows)			
Go to row <input type="text"/> Go			
#	ShipMethodID	CompanyName	
1	5	CARGO TRANSPORT 5	...
2	4	OVERNIGHT J-FAST	...
3	3	OVERSEAS - DELUXE	...
4	1	XRQ - TRUCK GROUND	...
5	2	ZY - EXPRESS	...

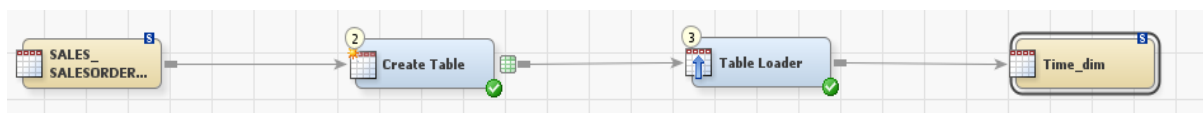
Rysunek 18 Pełne dane z tabeli Shipment\_DIM

Shipment_DIM Properties									
General Columns Indexes Keys Parameters Physical Storage Options Notes Extended Attributes Authorization									
#	Name	Description	Type	Len...	Informat	Format	Is Nullable	Summary ...	Sort Order
1	ShipMethodID	ShipMethodID	Numeric	8	11.	11.	No	(None)	(None)
2	CompanyName	Name	Character	1024	\$1024.	\$1024.	Yes	(None)	(None)

Rysunek 19 Pola tabeli Shipment\_DIM

## Time\_DIM

Kolejną utworzoną tabelą jest Time\_DIM. Powstaje ona poprzez przeładowanie dat zamówień (Order Date) z tabeli SalesOrderHeader. Proces loader\_time został zaprezentowany poniżej.



Rysunek 20 Przebieg procesu loader\_time

W tabeli zapisano 6 kolumn:

- Czas\_ID – zawierającą klucz główny tabeli
- Rok
- Miesiąc
- Kwartał
- Dzień
- Dzień tygodnia.

Time_dim Properties									
General Columns Indexes Keys Parameters Physical Storage Options Notes Extended Attributes Authorization									
#	Name	Description	Type	Len...	Informat	Format	Is Nullable	Summary ...	Sort Order
1	Czas_id		Character	8	\$8.	\$8.	No	(None)	(None)
2	Rok		Numeric	8	(None)	(None)	Yes	(None)	(None)
3	Miesiac		Numeric	8	(None)	(None)	Yes	(None)	(None)
4	Kwartał		Numeric	8	(None)	(None)	Yes	(None)	(None)
5	Dzien		Numeric	8	(None)	(None)	Yes	(None)	(None)
6	Dzien_tygodnia		Numeric	8	(None)	(None)	Yes	(None)	(None)

Rysunek 21 Pola tabeli Time\_DIM

Na potrzeby wygenerowania zawartości kolumn wykorzystano funkcje, takie jak Year(), QTR(), Month(), Day(), Weekday() oraz inne poznane na labolatoriach.

Create Table Properties				
General Source Result Filter and Sort Group Options Table Options Code Precode and Postcode Parameters Notes Extended Attributes				
<input type="checkbox"/> Remove duplicate rows (DISTINCT) <input type="checkbox"/> All fields (SELECT *)				
<div> <div>Source table: SALES_SALESORDERHE...</div> <div>Target table: Create Table (W16QBHP)</div> </div>				
#	Column	Column Des...	Type	Expr
1	Sales...	SalesOrderID	Numeric	
2	Revisi...	RevisionNu...	Numeric	
3	Order...	OrderDate	Numeric	
4	DueD...	DueDate	Numeric	
5	ShipD...	ShipDate	Numeric	
6	Status	Status	Numeric	
7	Onlin...	OnlineOrder...	Numeric	
8	Sales...	SalesOrderN...	Character	
9	Purch...	PurchaseOrd...	Character	
10	Accou...	AccountNum...	Character	
11	Custo...	CustomerID	Numeric	
12	Conta...	ContactID	Numeric	
13	Sales...	SalesPersonID	Numeric	
14	Territ...	TerritoryID	Numeric	
15	BillTo...	BillToAddr...	Numeric	
16	ShipT...	ShipToAddr...	Numeric	
17	ShipM...	ShipMethodID	Numeric	
18	Credit...	CreditCardID	Numeric	
19	Credit...	CreditCard...	Character	
20	Curre...	CurrencyRat...	Numeric	
21	SubT...	SubTotal	Numeric	
22	TaxAmt	TaxAmt	Numeric	
23	Freight	Freight	Numeric	
24	Total...	TotalDue	Numeric	
25	Com...	Comment	Character	
26	rowguid	rowguid	Character	
27	Modifi...	ModifiedDate	Numeric	

Rysunek 22 Tworzenie pól dla tabeli Time\_DIM

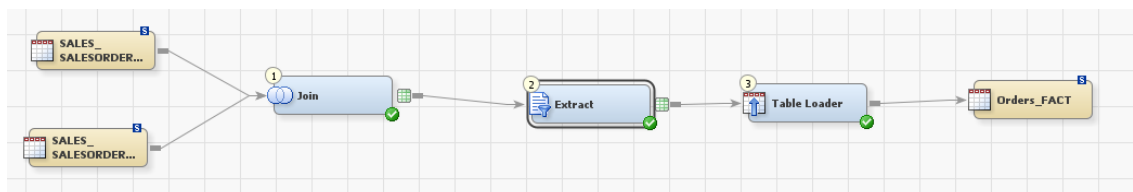
Pierwsze wiersze nowo utworzonej tabeli zaprezentowano poniżej

View Data: Time_dim						
Go to row <input type="text"/> Go						
#	Czas_id	Rok	Miesiac	Kwartal	Dzien	Dzien_tygodnia
1	01010020	2001	10	4	20	7
2	0108030	2001	8	3	30	5
3	01010012	2001	10	4	12	6
4	01011012	2001	11	4	12	2
5	0101202	2001	12	4	2	1
6	0108031	2001	8	3	31	6
7	0108012	2001	8	3	12	1
8	0107031	2001	7	3	31	3
9	010801	2001	8	3	1	4
10	0101201	2001	12	4	1	7
11	01012021	2001	12	4	21	6
12	01012031	2001	12	4	31	2
13	0101202	2001	12	4	2	1
14	0108021	2001	8	3	21	3
15	01012028	2001	12	4	28	6
16	010906	2001	9	3	6	5
17	0101204	2001	12	4	4	3
18	0101105	2001	11	4	5	2
19	01010022	2001	10	4	22	2
20	01012010	2001	12	4	10	2

Rysunek 23 Dane przykładowe z tabeli Time\_DIM

## Orders\_FACT

Tabela faktów modelu gwiazdy jest połączeniem informacji o zamówieniach. Zawarte w niej klucze do tabel wymiarów pozwalają określić czas sprzedaży, podmiot i przedmiot transakcji, jej warunki oraz przewoźnika. Wszystkie dane pochodzą z tabel zawierających dane o zamówieniach, czyli SALESORDERDETAIL oraz SALESORDERHEADER.




Rysunek 24 Przebieg procesu loader\_fact\_table

Tabela FAKT\_ZAMOWIENIA składa się z 10 kolumn. Klucz główny tej tabeli składa się z 6 unikalnych wartości:

- SalesOrderID – kolumna pozwalająca na wyszukanie konkretnych zamówień
- ProductID – ID produktu
- SpecialOfferID - ID wyprzedaży
- Czas\_ID – ID czasu zamówienia
- Contact\_ID – ID pozwalające wyszukać dane kontaktowe klienta
- ShipMethod\_ID – ID przewoźnika

Orders_FACT Properties									
General Columns Indexes Keys Parameters Physical Storage Options Notes Extended Attributes Authorization									
#	Name	Description	Type	Len...	Informat	Format	Is Nullable	Summary ...	Sort On
1	SalesOrderID	SalesOrderID	Numeric	8 11.	11.		No	(None)	(None)
2	OrderQty	OrderQty	Numeric	8 6.	6.		No	(None)	(None)
3	ProductID	ProductID	Numeric	8 11.	11.		No	(None)	(None)
4	SpecialOfferID	SpecialOfferID	Numeric	8 11.	11.		No	(None)	(None)
5	UnitPrice	UnitPrice	Numeric	8 DOLLA...	DOLLA...		Yes	(None)	(None)
6	UnitPriceDiscount	UnitPriceDisc...	Numeric	8 DOLLA...	DOLLA...		Yes	(None)	(None)
7	LineTotal	LineTotal	Numeric	8 30.6	30.6		Yes	(None)	(None)
8	Czas_ID		Character	8 \$8.	\$8.		No	(None)	(None)
9	ContactID	ContactID	Numeric	8 11.	11.		No	(None)	(None)
10	ShipMethodID	ShipMethodID	Numeric	8 11.	11.		No	(None)	(None)

Rysunek 25 Pola tabeli faktów

View Data: Orders_FACT										
Go to row <input type="text"/> Go 										
#	SalesOrderID	OrderQty	ProductID	SpecialOfferID	UnitPrice	UnitPriceDiscount	LineTotal	Czas_ID	ContactID	ShipMethodID
1	75123	1	879	1	\$159.00	\$0.00	159.0000000407031		4189	
2	75123	1	878	1	\$21.98	\$0.00	21.9800000407031		4189	
3	75123	1	712	1	\$8.99	\$0.00	8.9900000407031		4189	
4	75122	1	878	1	\$21.98	\$0.00	21.9800000407031		4781	
5	75122	1	712	1	\$8.99	\$0.00	8.9900000407031		4781	
6	75121	1	921	1	\$4.99	\$0.00	4.9900000407031		16667	
7	75121	1	707	1	\$34.99	\$0.00	34.9900000407031		16667	
8	75121	1	930	1	\$35.00	\$0.00	35.0000000407031		16667	
9	75120	1	884	1	\$53.99	\$0.00	53.9900000407031		18483	
10	75120	1	878	1	\$21.98	\$0.00	21.9800000407031		18483	
11	75120	1	712	1	\$8.99	\$0.00	8.9900000407031		18483	
12	75119	1	873	1	\$2.29	\$0.00	2.2900000407031		7796	
13	75119	1	930	1	\$35.00	\$0.00	35.0000000407031		7796	
14	75119	1	921	1	\$4.99	\$0.00	4.9900000407031		7796	
15	75118	1	873	1	\$2.29	\$0.00	2.2900000407031		17551	
16	75118	1	876	1	\$120.00	\$0.00	120.0000000407031		17551	
17	75118	1	923	1	\$4.99	\$0.00	4.9900000407031		17551	
18	75118	1	877	1	\$7.95	\$0.00	7.9500000407031		17551	
19	75117	1	923	1	\$4.99	\$0.00	4.9900000407031		15638	
20	75117	1	860	1	\$24.49	\$0.00	24.4900000407031		15638	
21	75116	1	923	1	\$4.99	\$0.00	4.9900000407031		17504	
22	75115	1	880	1	\$54.99	\$0.00	54.9900000407031		3783	
23	75115	1	922	1	\$3.99	\$0.00	3.9900000407031		3783	
24	75115	1	931	1	\$21.49	\$0.00	21.4900000407031		3783	
25	75114	1	931	1	\$21.49	\$0.00	21.4900000407031		15056	
26	75113	1	714	1	\$49.99	\$0.00	49.9900000407031		7008	
27	75113	1	928	1	\$24.99	\$0.00	24.9900000407031		7008	
28	75113	1	711	1	\$34.99	\$0.00	34.9900000407031		7008	

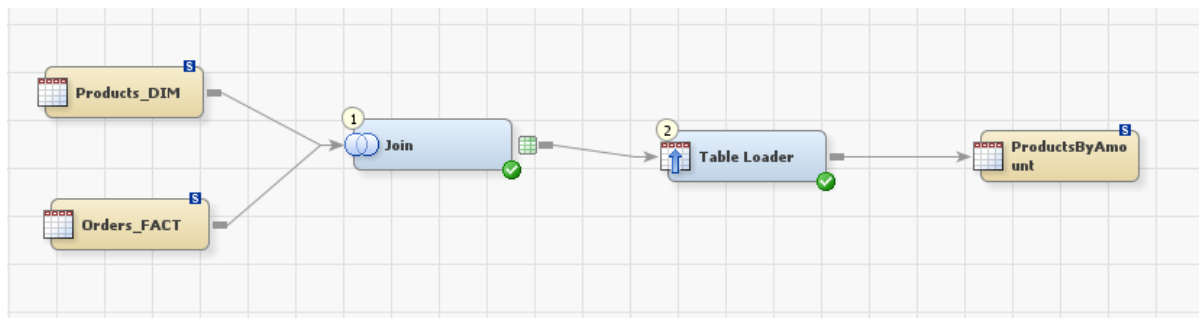
Rysunek 26 Dane przykładowe z tabeli faktów

## Wykorzystanie schematu gwiazdy

Wartości zamówień dla poszczególnych produktów

Pierwszy spośród procesów operujących na wcześniej utworzonym schemacie gwiazdy ma za zadanie ukazać, które produkty kupowano w największych ilościach. Tabela utworzona w tym procesie może również posłużyć do wskazania wartości sprzedaży dla poszczególnych towarów.

Poniżej zaprezentowano schemat procesu.



Rysunek 27 Proces tworzenia tabeli ProductsByAmount

W tabeli docelowej zapisane jest 7 kolumn dostarczających podstawowej wiedzy na temat produktów. Kolumny te to :

- CategoryName – nazwa kategorii
- Wartość – wartość sumaryczna zamówień dla danego produktu
- SubcategoryName – nazwa subkategorii
- ProductName – nazwa produktu
- ProductID – ID produktu
- Class – klasa produktu

#	Name	Description	Type	Len...	Informat	Format	Is Nullable	Summary ...	Sort Order
1	CategoryName	C_Name	Character	1024	\$1024.	\$1024.	Yes	(None)	(None)
2	Wartosc		Numeric	8	DOLLA...	DOLLA...	Yes	(None)	(None)
3	SubcategoryName	S_Name	Character	1024	\$1024.	\$1024.	Yes	(None)	(None)
4	ProductName	P_Name	Character	1024	\$1024.	\$1024.	Yes	(None)	(None)
5	ProductID	ProductID	Numeric	8	11.	11.	No	(None)	(None)
6	Class	Class	Character	2	\$2.	\$2.	Yes	(None)	(None)

Rysunek 28 Pola tabeli ProductsByAmount

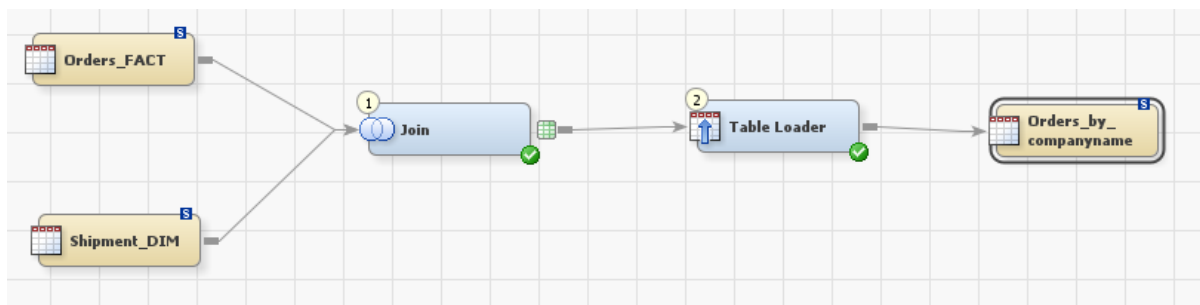
View Data: ProductsByAmount						
Go to row <input type="text"/> Go						
#	CategoryName	Wartosc	SubcategoryName	ProductName	ProductID	Class
1	Bikes	\$4,406,14...	Mountain Bikes	Mountain-200 Black, 38	782	H
2	Bikes	\$4,014,06...	Mountain Bikes	Mountain-200 Black, 42	783	H
3	Bikes	\$3,696,48...	Mountain Bikes	Mountain-200 Silver, 38	779	H
4	Bikes	\$3,441,29...	Mountain Bikes	Mountain-200 Silver, 42	780	H
5	Bikes	\$3,436,08...	Mountain Bikes	Mountain-200 Silver, 46	781	H
6	Bikes	\$3,311,09...	Mountain Bikes	Mountain-200 Black, 46	784	H
7	Bikes	\$2,518,29...	Road Bikes	Road-250 Black, 44	793	H
8	Bikes	\$2,348,24...	Road Bikes	Road-250 Black, 48	794	H
9	Bikes	\$2,012,44...	Road Bikes	Road-250 Black, 52	795	H
10	Bikes	\$1,847,81...	Road Bikes	Road-150 Red, 56	753	H
11	Bikes	\$1,788,32...	Road Bikes	Road-350-W Yellow, 48	976	M
12	Bikes	\$1,769,09...	Road Bikes	Road-150 Red, 62	749	H
13	Bikes	\$1,728,80...	Touring Bikes	Touring-1000 Blue, 60	969	H
14	Bikes	\$1,663,51...	Road Bikes	Road-350-W Yellow, 40	973	M
15	Bikes	\$1,592,38...	Touring Bikes	Touring-1000 Blue, 46	966	H
16	Bikes	\$1,587,00...	Road Bikes	Road-250 Red, 58	792	H
17	Bikes	\$1,578,59...	Touring Bikes	Touring-1000 Yellow, 60	957	H
18	Bikes	\$1,540,80...	Road Bikes	Road-150 Red, 48	751	H
19	Bikes	\$1,506,37...	Road Bikes	Road-250 Black, 58	796	H
20	Bikes	\$1,476,97...	Touring Bikes	Touring-1000 Yellow, 46	954	H
21	Bikes	\$1,449,80...	Road Bikes	Road-250 Red, 44	789	H
22	Bikes	\$1,415,56...	Road Bikes	Road-150 Red, 52	752	H
23	Bikes	\$1,383,86...	Mountain Bikes	Mountain-100 Black, 44	777	H
24	Bikes	\$1,362,05...	Mountain Bikes	Mountain-100 Black, 38	775	H
25	Bikes	\$1,349,07...	Road Bikes	Road-250 Red, 48	790	H
26	Bikes	\$1,340,41...	Road Bikes	Road-150 Red, 44	750	H
27	Bikes	\$1,311,18...	Mountain Bikes	Mountain-100 Silver, 38	771	H
28	Bikes	\$1,272,88...	Mountain Bikes	Mountain-100 Black, 42	776	H
29	Bikes	\$1,252,26...	Mountain Bikes	Mountain-100 Black, 48	778	H
30	Bikes	\$1,235,24...	Road Bikes	Road-550-W Yellow, 48	801	M
31	Bikes	\$1,232,71...	Mountain Bikes	Mountain-100 Silver, 44	773	H
32	Bikes	\$1,228,34...	Road Bikes	Road-550-W Yellow, 38	797	M
33	Bikes	\$1,202,79...	Mountain Bikes	Mountain-100 Silver, 42	772	H
34	Bikes	\$1,121,24...	Road Bikes	Road-350-W Yellow, 42	974	M
35	Bikes	\$1,071,44...	Road Bikes	Road-550-W Yellow, 40	798	M

Rysunek 29 Dane przykładowe w tabeli ProductsByAmount

W tabeli ProductsByAmount produkty uporządkowano w kolejności rosnącej według wartości sprzedaży dla danego towaru. Jak widać najlepiej sprzedającym się towarem był Rower Mountain-200 Black, 38.

Liczba zamówień dostarczonych przez poszczególne firmy transportowe

Kolejny proces ma na celu ukazać, który przewoźnik przewiózł największą liczbę zamówień oraz ile zamówień zostało obsłużonych przez poszczególnych przewoźników.



Rysunek 30 Proces tworzenia tabeli Orders\_by\_companyname

Tabela docelowa składa się z dwóch kolumn:

- CompanyName
- Orders

Struktura utworzonej w wyniku procesu tabeli oraz jej zawartość zostały przedstawione poniżej.

Orders_by_companyname Properties									
General Columns Indexes Keys Parameters Physical Storage Options Notes Extended Attributes Authorization									
#	Name	Description	Type	Len...	Informat	Format	Is Nullable	Summary ...	Sort Order
1	CompanyName	Name	Character	1024	\$1024.	\$1024.	Yes	(None)	(None)
2	Orders		Numeric	8	(None)	(None)	Yes	(None)	(None)

Rysunek 31 Pola tabeli Orders\_by\_companyname

View Data: Orders_by_companyname (2 rows)		
#	CompanyName	Orders
1	CARGO TRANSPORT 5 ...	60919
2	XRQ - TRUCK GROUND ...	60398

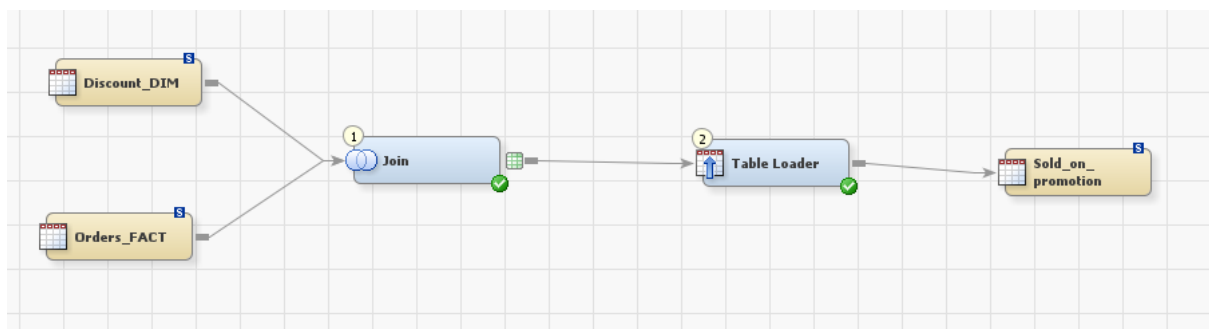
Rysunek 32 Podgląd pełnych danych dla tabeli Orders\_by\_companyname

Co widoczne jedynie 2 przewoźników zajmowało się transportem produktów z badanego zakresu.

Wartość sprzedaży odnotowanej na poszczególnych typach promocji.

Proces ma na celu przedstawić, jakie akcje promocyjne prowadzone były w przedsiębiorstwie Adventure Works oraz jaki był ich sukces. Na podstawie uzyskanych wyników, dział sprzedaży fikcyjnego przedsiębiorstwa mógłby opracować raport na temat łącznej wartości sprzedaży dla poszczególnych promocji wraz z informacjami na temat liczby sztuk sprzedanych towarów.

Poniżej przedstawiono schemat przepływu.



Rysunek 33 Tworzenie tabeli Sold\_on\_promotion

Tabela docelowa składa się z trzech kolumn, są to:

- Description
- OrderQty
- Total

Szczegółowa struktura tabeli wraz z przykładowymi danymi została zaprezentowana poniżej.



Sold_on_promotion Properties									
General Columns Indexes Keys Parameters Physical Storage Options Notes Extended Attributes Authorization									
#	Name	Description	Type	Len...	Informat	Format	Is Nullable	Summary ...	Sort Ord
1	Description	Description	Character	1024	\$1024.	\$1024.	Yes	(None)	(None)
2	OrderQty		Numeric	8	(None)	(None)	Yes	(None)	(None)
3	Total		Numeric	8	(None)	(None)	Yes	(None)	(None)

Rysunek 34 Pola tabeli Sold\_on\_promotion

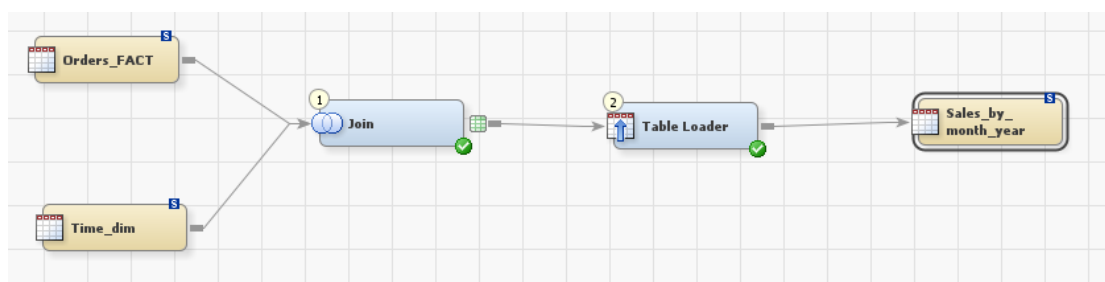
View Data: Sold_on_promotion (12 rows)				
#	Description	OrderQty	Total	
1	No Discount	238944	102372622.26	
2	Volume Discount 11 to 14	18181	4955135.1743	
3	Volume Discount 15 to 24	10713	1091720.4846	
4	Volume Discount 25 to 40	2321	137710.7159	
5	Touring-3000 Promotion	1581	536076.6075	
6	Touring-1000 Promotion	775	757505.046	
7	Sport Helmet Discount-2003	680	10604.94	
8	Sport Helmet Discount-2002	492	8227.2732	
9	Mountain-100 Clearance Sale	456	385883.01	
10	Mountain-500 Silver Clearance Sale	382	43012.436	
11	Road-650 Overstock	304	71317.488	
12	Volume Discount 41 to 60	85	2030.7675	

Rysunek 35 Pełne dane w tabeli Sold\_on\_promotion

Największą liczbę oraz wartość sumaryczną sprzedanych towarów sprzedano po cenach standardowych, bez zniżek.

Sprzedaż, liczba sprzedanych produktów w przeliczeniu na poszczególne lata i miesiące  
Ostatni proces przygotowany na podstawie utworzonego wcześniej schematu gwiazdy przedstawia ilość oraz wartości sprzedaży produktów w przeliczeniu na poszczególne lata i miesiące.

Schemat przepływu przedstawiono poniżej.



Rysunek 36 Tworzenie tabeli Sales\_by\_month\_year

Tabela końcowa składa się z czterech kolumn:

- Rok
- Miesiąc
- Qty (liczba sprzedanych produktów)
- TotalWorth (wartość sprzedanych produktów)

Szczegółowa struktura tabeli wraz z przykładowymi danymi została zaprezentowana poniżej.

Sales_by_month_year Properties									
General Columns Indexes Keys Parameters Physical Storage Options Notes Extended Attributes Authorization									
#	Name	Description	Type	Len...	Informat	Format	Is Nullable	Summary ...	Sort Order
1	Rok		Numeric	8	(None)	(None)	Yes	(None)	(None)
2	Miesiac		Numeric	8	(None)	(None)	Yes	(None)	(None)
3	Qty		Numeric	8	10.	10.	Yes	(None)	(None)
4	TotalWorth		Numeric	8	DOLLA...	DOLLA...	Yes	(None)	(None)

Rysunek 37 Pola tabeli Sales\_by\_month\_year

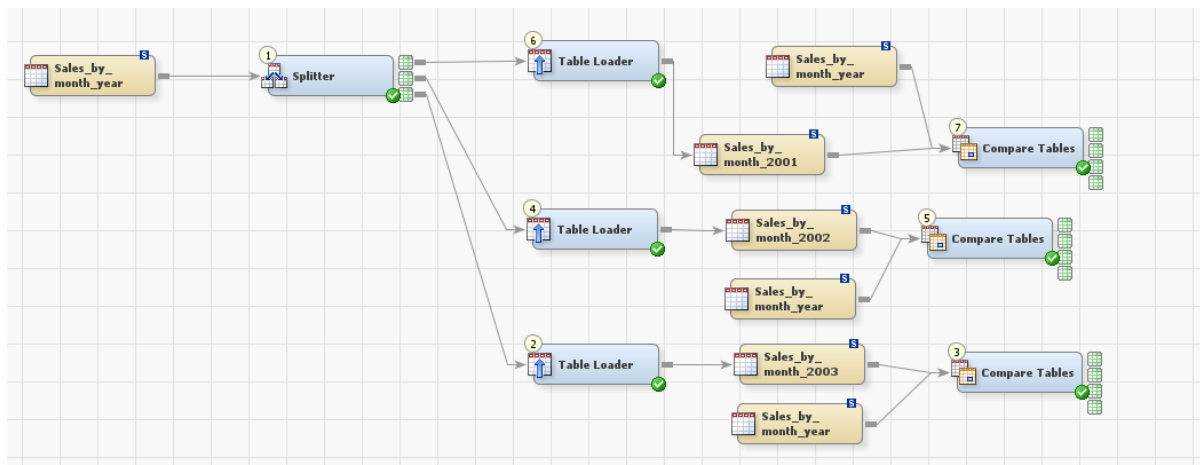
View Data: Sales_by_month_year (37 rows)				
Go to row <input type="text"/> Go				
#	Rok	Miesiac	Qty	TotalWorth
1	2001	7	36256	\$24,207,670.17
2	2001	8	165543	\$127,526,478...
3	2001	9	97918	\$78,463,375.40
4	2001	10	55929	\$41,172,631.69
5	2001	11	286039	\$228,525,893...
6	2001	12	155110	\$123,696,312...
7	2002	1	39796	\$36,605,594.46
8	2002	2	182963	\$167,117,577...
9	2002	3	116793	\$107,764,362...
10	2002	4	54806	\$43,014,160.55
11	2002	5	290333	\$228,707,329...
12	2002	6	160756	\$88,467,873.20
13	2002	7	602991	\$198,433,673...
14	2002	8	1638496	\$542,691,494...
15	2002	9	1003808	\$330,061,253...
16	2002	10	436230	\$150,498,505...
17	2002	11	1124660	\$428,502,390...
18	2002	12	797948	\$282,717,347...
19	2003	1	242680	\$100,841,155...
20	2003	2	736858	\$346,535,950...
21	2003	3	447680	\$186,024,368...
22	2003	4	463054	\$165,812,342...
23	2003	5	1173318	\$433,296,397...
24	2003	6	721590	\$241,832,659...
25	2003	7	1095487	\$313,148,470...
26	2003	8	3739343	\$1,049,942,8...
27	2003	9	3365243	\$940,541,330...
28	2003	10	1369216	\$408,481,006...
29	2003	11	2702102	\$878,576,836...
30	2003	12	2954520	\$981,432,173...
31	2004	1	938847	\$317,831,805...
32	2004	2	1891801	\$766,779,581...
33	2004	3	1943235	\$777,563,657...
34	2004	4	1495883	\$466,867,943...
35	2004	5	3181318	\$1,052,225,5...
36	2004	6	2768132	\$933,235,916...
37	2004	7	72033	\$1,656,499.96

Rysunek 38 Pełne dane w tabeli Sales\_by\_month\_year















## Jakość danych

Proces badania jakości danych został przeprowadzony na podstawie tabeli Sales\_by\_month\_year utworzonej w poprzednim procesie. Przy pomocy węzła splitter podzielono dane na podstawie kolumny rok. Pozwoliło to na uzyskanie tabel zawierających wartość sprzedaży z podziałem na lata. Tabele końcowe zawierające dane transakcyjne na lata 2001, 2002, 2003 zostały następnie porównane z tabelą początkową. Przebieg procesu zobrazony jest poniżej.















Rysunek 39 Wykorzystanie węzła compare tables













Dla węzłów compare table ustawione zostało grupowanie na podstawie kolumny rok. Dla tabeli zawierającej dane na rok 2001 zmieniono kolejność portów wchodzących do węzła względem pozostałych porównań. Dla poszczególnych węzłów compare tables zaprezentowano wszystkie tabele, które dla danego porównania zostały utworzone. Warto zauważyć, że zmiana kolejności łączenia tabel do węzła spowodowała zmianę uzyskanych tabel końcowych. Gdy tabela bazowa połączona była z portem górnym tworzone były tabele New Records oraz Unchanged records, natomiast gdy tabela bazowa połączona była z węzłem dolnym compare tables, to w wyniku procesu uzyskano tabele unchanged records oraz missing records. Dla każdego z węzłów tabela changed records była pusta, jako że nie dokonano modyfikacji danych bazowych, a jedynie ich podziału.

View Data: New records (31 rows)				
Go to row <input type="text"/> Go            				
#	Rok	Miesiac	Qty	TotalWorth
1	2002	1	39796	\$36,605,594.46
2	2002	2	182963	\$167,117,577...
3	2002	3	116793	\$107,764,362...
4	2002	4	54806	\$43,014,160.55
5	2002	5	290333	\$228,707,329...
6	2002	6	160756	\$88,467,873.20
7	2002	7	602991	\$198,433,673...
8	2002	8	1638496	\$542,691,494...
9	2002	9	1003808	\$330,061,253...
10	2002	10	436230	\$150,498,505...
11	2002	11	1124660	\$428,502,390...
12	2002	12	797948	\$282,717,347...
13	2003	1	242680	\$100,841,155...
14	2003	2	736858	\$346,535,950...
15	2003	3	447680	\$186,024,368...
16	2003	4	463054	\$165,812,342...
17	2003	5	1173318	\$433,296,397...
18	2003	6	721590	\$241,832,659...
19	2003	7	1095487	\$313,148,470...
20	2003	8	3739343	\$1,049,942,8...
21	2003	9	3365243	\$940,541,330...
22	2003	10	1369216	\$408,481,006...
23	2003	11	2702102	\$878,576,836...
24	2003	12	2954520	\$981,432,173...
25	2004	1	938847	\$317,831,805...
26	2004	2	1891801	\$766,779,581...
27	2004	3	1943235	\$777,563,657...
28	2004	4	1495883	\$466,867,943...
29	2004	5	3181318	\$1,052,225,5...
30	2004	6	2768132	\$933,235,916...
31	2004	7	72033	\$1,656,499.96













Rysunek 40 Tabela New Records dla porównania danych pierwotnych z danymi dla roku 2001

View Data: Unchanged records (6 rows)				
Go to row <input type="text"/> Go            				
#	Rok	Miesiac	Qty	TotalWorth
1	2001	7	36256	\$24,207,670.17
2	2001	8	165543	\$127,526,478...
3	2001	9	97918	\$78,463,375.40
4	2001	10	55929	\$41,172,631.69
5	2001	11	286039	\$228,525,893...
6	2001	12	155110	\$123,696,312...

Rysunek 41 Tabela Unchanged Records dla porównania danych pierwotnych z danymi dla roku 2001

View Data: Unchanged records (12 rows)				
Go to row <input type="text"/> Go            				
#	Rok	Miesiac	Qty	TotalWorth
1	2002	1	39796	\$36,605,594.46
2	2002	2	182963	\$167,117,577...
3	2002	3	116793	\$107,764,362...
4	2002	4	54806	\$43,014,160.55
5	2002	5	290333	\$228,707,329...
6	2002	6	160756	\$88,467,873.20
7	2002	7	602991	\$198,433,673...
8	2002	8	1638496	\$542,691,494...
9	2002	9	1003808	\$330,061,253...
10	2002	10	436230	\$150,498,505...
11	2002	11	1124660	\$428,502,390...
12	2002	12	797948	\$282,717,347...

Rysunek 42 Tabela Unchanged Records dla porównania danych dla roku 2002 z danymi pierwotnymi

View Data: Missing records (25 rows)			
Go to row <input type="text"/> Go            			
#	Miesiac	Qty	TotalWorth
1	4	463054	\$165,812,342...
2	1	242680	\$100,841,155...
3	1	938847	\$317,831,805...
4	3	1943235	\$777,563,657...
5	5	3181318	\$1,052,225,5...
6	11	2702102	\$878,576,836...
7	6	721590	\$241,832,659...
8	7	36256	\$24,207,670.17
9	5	1173318	\$433,296,397...
10	7	72033	\$1,656,499.96
11	12	155110	\$123,696,312...
12	2	736858	\$346,535,950...
13	10	1369216	\$408,481,006...
14	8	165543	\$127,526,478...
15	8	3739343	\$1,049,942,8...
16	11	286039	\$228,525,893...
17	3	447680	\$186,024,368...
18	9	97918	\$78,463,375.40
19	12	2954520	\$981,432,173...
20	4	1495883	\$466,867,943...
21	7	1095487	\$313,148,470...
22	9	3365243	\$940,541,330...
23	10	55929	\$41,172,631.69
24	6	2768132	\$933,235,916...
25	2	1891801	\$766,779,581...

Rysunek 43 Tabela Missing Records dla porównania danych dla roku 2002 z danymi pierwotnymi

View Data: Unchanged records (12 rows)				
Go to row <input type="text"/> Go				
#	12 Rok	12 Miesiac	12 Qty	12 TotalWorth
1	2003	1	242680	\$100,841,155...
2	2003	2	736858	\$346,535,950...
3	2003	3	447680	\$186,024,368...
4	2003	4	463054	\$165,812,342...
5	2003	5	1173318	\$433,296,397...
6	2003	6	721590	\$241,832,659...
7	2003	7	1095487	\$313,148,470...
8	2003	8	3739343	\$1,049,942,8...
9	2003	9	3365243	\$940,541,330...
10	2003	10	1369216	\$408,481,006...
11	2003	11	2702102	\$878,576,836...
12	2003	12	2954520	\$981,432,173...

Rysunek 44 Tabela Unchanged Records dla porównania danych dla roku 2003 z danymi pierwotnymi

View Data: Missing records (25 rows)			
Go to row <input type="text"/> Go			
#	12 Miesiac	12 Qty	12 TotalWorth
1	1	938847	\$317,831,805...
2	10	436230	\$150,498,505...
3	3	1943235	\$777,563,657...
4	5	290333	\$228,707,329...
5	9	1003808	\$330,061,253...
6	5	3181318	\$1,052,225,5...
7	7	602991	\$198,433,673...
8	11	1124660	\$428,502,390...
9	8	1638496	\$542,691,494...
10	4	54806	\$43,014,160.55
11	7	36256	\$24,207,670.17
12	7	72033	\$1,656,499.96
13	12	155110	\$123,696,312...
14	8	165543	\$127,526,478...
15	3	116793	\$107,764,362...
16	11	286039	\$228,525,893...
17	2	182963	\$167,117,577...
18	9	97918	\$78,463,375.40
19	12	797948	\$282,717,347...
20	4	1495883	\$466,867,943...
21	1	39796	\$36,605,594.46
22	10	55929	\$41,172,631.69
23	6	160756	\$88,467,873.20
24	6	2768132	\$933,235,916...
25	2	1891801	\$766,779,581...

Rysunek 45 Tabela Missing Records dla porównania danych dla roku 2003 z danymi pierwotnymi

## Import pliku

W ostatnim procesie dokonany zostanie import danych zewnętrznych do tabeli programu Data Integration Studio. W tym celu wykorzystany zostanie węzeł user written statement pokazany poniżej.



Rysunek 46 User written statement

Kod importu pliku zaprezentowany został poniżej.

```
%let _OUTPUT1_label = %nrquote();
%let _OUTPUT1_col_count = 0;
%let _OUTPUT1_filetype = WorkTable;

filename resp temp;
proc http
url='https://raw.githubusercontent.com/SyntheticAquamarine/SAS_Data/main/restaurant_details.json'
method = "get"
out = resp;
run;
libname rss json filename = resp

LIBNAME stud2021s015etlgl BASE "E:\Studenci\stud2021s015etlgl"
data stud_015.pizza
run;
%rcSet(&syserr);
%rcSet(&sqlrc);
%perfstop(metrVal6=%sysfunc(max(&etls_recnt,-1)));
%let etls_recnt=-1;
```

Rysunek 47 Kod importu pliku

```
"title": "Royal Fast Food",
"dining_rating": "3.8",
"dining_total_review": "132",
"delivery_rating": "3.4",
"delivery_total_review": "8,292",
"category": ["Chinese", "North Indian", "Fast Food", "Mughlai", "Sichuan"],
"address": "New Textile Market, Surat",
"time": "8am \u2013 9pm (Today)",
"location": [
  "https://www.google.com/maps/dir/?api=1&destination=21.1891766227,72.8182196617"
],
"users_review": {
  "Lizarani": "",
  "ILYAS SHAIKH": "",
  "SANDIP THAKUR": "perfect",
```

Rysunek 48 Podgląd importowanych danych

## Podsumowanie

Celem niniejszego projektu było zaprezentowanie możliwości narzędzia SAS Data Integration Studio. Przy wykorzystaniu programu utworzono procesy, tabele danych służące do dalszych operacji oraz wykorzystano wbudowane funkcje w celu przetwarzania danych bazowych. Działania te pozwoliły na uzyskanie informacji, które mogłyby być pomocne w procesie podejmowania decyzji w przedsiębiorstwie Adventure Works. SAS Data Integration Studio posiada olbrzymi zasób możliwości, pozwalając na budowę skomplikowanych procesów ETL przy wykorzystaniu danych pochodzących z różnych źródeł. W projekcie wykorzystano operacje mające na celu utworzenie szczegółowych kwerend, umożliwiających wygenerowanie raportów, jak również narzędzia do tworzenia oraz ładowania tabel danych z tabel źródłowych. Do stworzenia raportów dotyczących rozkładu zamówień w kolejnych latach i miesiącach podzielono kolumnę z datą zamówienia, na kolumny w tabeli docelowej reprezentujące: rok, kwartał, miesiąc, dzień oraz dzień tygodnia. Pozwala to na wskazanie, w których okresach sprzedaż była największa. W kolejnym procesie zaprezentowano, które produkty kupowane były najczęściej oraz ile wyniósł sumaryczny zysk z ich sprzedaży. Pozwoli to na uniknięcie nadprodukcji oraz zmaksymalizowanie zysków, poprzez rezygnację z produkcji mało popularnych artykułów i skupienie na rozwoju linii produkcyjnych, dla przedmiotów pożądanых. W kolejnym procesie przedstawiono wykaz przewoźników wraz z ilością towaru przez nich przetransportowaną. Dla procesu następnego utworzono raport dotyczący sprzedaży na poszczególnych typach promocji, jak również liczby artykułów podczas wyprzedaży. Pozwoli to na odpowiednie wybranie typów promocji w celu osiągnięcia zamierzonych zysków ze sprzedaży. Należy wspomnieć, że do stworzenia projektu wykorzystano połączone tabele z bazy Adventure Works, co pozwoliło na stworzenie dodatkowych raportów, z uwzględnieniem potrzeb firmy. Dzięki złączeniom poszczególnych tabel, możliwe jest przedstawienie kompleksowej analizy firmy.