

Hurtownie Danych - laboratorium Lista 6

Proces ETL

Uwaga: realizacja zadań bazuje na wynikach zadań listy 4 i 5.

Wstęp teoretyczny

Usuwanie schematu:

```
DROP SCHEMA [ IF EXISTS ] schema_name
```

Usuwanie tabeli:

```
DROP TABLE [ IF EXISTS ]  
{ database_name.schema_name.table_name | schema_name.ta-  
ble_name | table_name } [ ,...n ]
```

Źródła:

<https://docs.microsoft.com/en-us/sql/t-sql/statements/drop-schema-transact-sql?view=sql-server-ver15>

<https://docs.microsoft.com/en-us/sql/t-sql/statements/drop-table-transact-sql?view=sql-server-ver15>

Zadania do wykonania

Baza danych: **AdventureWorks**

Zad. 1. Przygotowanie instrukcji usuwania tabel

Przygotować instrukcje SQL usuwające każdą z tabel wymiarów i tabeli faktów utworzonych w trakcie pracy nad listą 4 i 5.

Uwaga: Instrukcja powinna być wykonana tylko pod warunkiem istnienia usuwanej tabeli.

Należy sprawdzić, czy dana tabela istnieje.

Rozwiązania:

```
DROP TABLE IF EXISTS STROZIK.FACT_SALES;  
  
DROP TABLE IF EXISTS STROZIK.DIM_PRODUCT;  
  
DROP TABLE IF EXISTS STROZIK.DIM_CUSTOMER;  
  
DROP TABLE IF EXISTS STROZIK.DIM SALESPERSON;  
  
DROP TABLE IF EXISTS STROZIK.DIM_TIME;  
  
DROP TABLE IF EXISTS STROZIK.months_names;  
  
DROP TABLE IF EXISTS STROZIK.weekday_names;  
  
DROP SCHEMA IF EXISTS STROZIK;
```

Zad. 2. Procedury czyszczenia danych

Przygotować instrukcje SQL, które zamienią wszystkie wartości NULL:

- w kolumnie Color (DIM_PRODUCT) na „Unknown”,
- w kolumnie SubCategoryName (DIM_PRODUCT) na „Unknown”.
- w kolumnie CountryRegionCode (DIM_CUSTOMER, DIM_SALESPERSON) na 000,
- w kolumnie Group (DIM_CUSTOMER, DIM_SALESPERSON) na „Unknown”.

Rozwiązania:

```
UPDATE STROZIK.DIM_PRODUCT  
SET Color = 'Unknown'  
WHERE Color IS NULL;  
  
SELECT * FROM STROZIK.DIM_PRODUCT;
```

	ProduktID	Name	ListPrice	Color	SubCategoryName	CategoryName	Weight	Size	IsPurchased
1	1	Adjustable Race	0.00	Unknown	NULL	NULL	NULL	NULL	1
2	2	Bearing Ball	0.00	Unknown	NULL	NULL	NULL	NULL	1
3	3	BB Ball Bearing	0.00	Unknown	NULL	NULL	NULL	NULL	0
4	4	Headset Ball Bearings	0.00	Unknown	NULL	NULL	NULL	NULL	1
5	316	Blade	0.00	Unknown	NULL	NULL	NULL	NULL	0
6	317	LL Crankarm	0.00	Black	NULL	NULL	NULL	NULL	1
7	318	ML Crankarm	0.00	Black	NULL	NULL	NULL	NULL	1
8	319	HL Crankarm	0.00	Black	NULL	NULL	NULL	NULL	1
9	320	Chainring Bolts	0.00	Silver	NULL	NULL	NULL	NULL	1
10	321	Chainring Nut	0.00	Silver	NULL	NULL	NULL	NULL	1
11	322	Chainring	0.00	Black	NULL	NULL	NULL	NULL	1
12	323	Crown Race	0.00	Unknown	NULL	NULL	NULL	NULL	1
13	324	Chain Stays	0.00	Unknown	NULL	NULL	NULL	NULL	0
14	325	Decal 1	0.00	Unknown	NULL	NULL	NULL	NULL	1
15	326	Decal 2	0.00	Unknown	NULL	NULL	NULL	NULL	1
16	327	Down Tube	0.00	Unknown	NULL	NULL	NULL	NULL	0
17	328	Mountain End Caps	0.00	Unknown	NULL	NULL	NULL	NULL	0
18	329	Road End Caps	0.00	Unknown	NULL	NULL	NULL	NULL	0

```
UPDATE STROZIK.DIM_PRODUCT
SET SubCategoryName = 'Unknown'
WHERE SubCategoryName IS NULL;
```

```
SELECT * FROM STROZIK.DIM_PRODUCT;
```

	ProduktID	Name	ListPrice	Color	SubCategoryName	CategoryName	Weight	Size	IsPurchased
1	1	Adjustable Race	0.00	Unknown	Unknown	NULL	NULL	NULL	1
2	2	Bearing Ball	0.00	Unknown	Unknown	NULL	NULL	NULL	1
3	3	BB Ball Bearing	0.00	Unknown	Unknown	NULL	NULL	NULL	0
4	4	Headset Ball Bearings	0.00	Unknown	Unknown	NULL	NULL	NULL	1
5	316	Blade	0.00	Unknown	Unknown	NULL	NULL	NULL	0
6	317	LL Crankarm	0.00	Black	Unknown	NULL	NULL	NULL	1
7	318	ML Crankarm	0.00	Black	Unknown	NULL	NULL	NULL	1
8	319	HL Crankarm	0.00	Black	Unknown	NULL	NULL	NULL	1
9	320	Chainring Bolts	0.00	Silver	Unknown	NULL	NULL	NULL	1
10	321	Chainring Nut	0.00	Silver	Unknown	NULL	NULL	NULL	1
11	322	Chainring	0.00	Black	Unknown	NULL	NULL	NULL	1

```
UPDATE STROZIK.DIM_CUSTOMER
SET CountryRegionCode = '000'
WHERE CountryRegionCode IS NULL;
```

```
SELECT * FROM STROZIK.DIM_CUSTOMER;
```

	CustomerID	FirstName	LastName	Title	City	TerritoryName	CountryRegionCode	Group
1	11000	Jon	Yang	NULL	Rockhampton	Australia	AU	Pacific
2	11001	Eugene	Huang	NULL	Seaford	Australia	AU	Pacific
3	11002	Ruben	Torres	NULL	Hobart	Australia	AU	Pacific
4	11003	Christy	Zhu	NULL	North Ryde	Australia	AU	Pacific
5	11004	Elizabeth	Johnson	NULL	Wollongong	Australia	AU	Pacific
6	11005	Julio	Ruiz	NULL	East Brisbane	Australia	AU	Pacific
7	11006	Janet	Alvarez	NULL	Matraville	Australia	AU	Pacific
8	11007	Marco	Mehta	NULL	Warrnambool	Australia	AU	Pacific
9	11008	Rob	Verhoff	NULL	Bendigo	Australia	AU	Pacific

Akurat w tym przypadku nie wprowadzono żadnych zmian.

```

UPDATE STROZIK.DIM_SALESPERSON
SET CountryRegionCode = '000'
WHERE CountryRegionCode IS NULL;

SELECT * FROM STROZIK.DIM_SALESPERSON;

```

	SalesPersonID	FirstName	LastName	Title	Gender	CountryRegionCode	Group
1	274	Stephen	Jiang	NULL	M	000	NULL
2	275	Michael	Blythe	NULL	M	US	North America
3	276	Linda	Mitchell	NULL	F	US	North America
4	277	Jillian	Carson	NULL	F	US	North America
5	278	Garrett	Vargas	NULL	M	CA	North America
6	279	Tsvi	Reiter	NULL	M	US	North America
7	280	Pamela	Ansman-Wolfe	NULL	F	US	North America
8	281	Shu	Ito	NULL	M	US	North America
9	282	José	Saraiva	NULL	M	CA	North America
10	283	David	Campbell	NULL	M	US	North America
11	284	Tete	Mensa-Annan	Mr.	M	US	North America
12	285	Syed	Abbas	Mr.	M	000	NULL
13	286	Lynn	Tsoflias	NULL	F	AU	Pacific
14	287	Amy	Alberts	NULL	F	000	NULL
15	288	Rachel	Valdez	NULL	F	DE	Europe
16	289	Jae	Pak	NULL	F	GB	Europe
17	290	Ranjit	Varkey Chudukatil	NULL	M	FR	Europe

```

UPDATE STROZIK.DIM_CUSTOMER
SET [Group] = 'Unknown'
WHERE [Group] IS NULL;

SELECT * FROM STROZIK.DIM_CUSTOMER;

```

	CustomerID	FirstName	LastName	Title	City	TerritoryName	CountryRegionCode	Group
1	11000	Jon	Yang	NULL	Rockhampton	Australia	AU	Pacific
2	11001	Eugene	Huang	NULL	Seaford	Australia	AU	Pacific
3	11002	Ruben	Torres	NULL	Hobart	Australia	AU	Pacific
4	11003	Christy	Zhu	NULL	North Ryde	Australia	AU	Pacific
5	11004	Elizabeth	Johnson	NULL	Wollongong	Australia	AU	Pacific
6	11005	Julio	Ruiz	NULL	East Brisbane	Australia	AU	Pacific
7	11006	Janet	Alvarez	NULL	Matrville	Australia	AU	Pacific

Tutaj podobna sytuacja, również nie wprowadzono zmian.

```
UPDATE STROZIK.DIM_SALESPERSON
SET [Group] = 'Unknown'
WHERE [Group] IS NULL;
```

```
SELECT * FROM STROZIK.DIM_SALESPERSON;
```

	SalesPersonID	FirstName	LastName	Title	Gender	CountryRegionCode	Group
1	274	Stephen	Jiang	NULL	M	000	Unknown
2	275	Michael	Blythe	NULL	M	US	North America
3	276	Linda	Mitchell	NULL	F	US	North America
4	277	Jillian	Carson	NULL	F	US	North America
5	278	Garrett	Vargas	NULL	M	CA	North America
6	279	Tsvi	Reiter	NULL	M	US	North America
7	280	Pamela	Ansman-Wolfe	NULL	F	US	North America
8	281	Shu	Ito	NULL	M	US	North America
9	282	José	Saraiva	NULL	M	CA	North America
10	283	David	Campbell	NULL	M	US	North America
11	284	Tete	Mensa-Annan	Mr.	M	US	North America
12	285	Syed	Abbas	Mr.	M	000	Unknown
13	286	Lynn	Tsoflias	NULL	F	AU	Pacific
14	287	Amy	Alberts	NULL	F	000	Unknown
15	288	Rachel	Valdez	NULL	F	DE	Europe
16	289	Jae	Pak	NULL	F	GB	Europe
17	290	Ranjit	Varkey Chudukatil	NULL	M	FR	Europe

Zad. 3. Tabela dziennika

Opracować i utworzyć tabelę dziennika operacji ETL, która będzie zawierać informacje o zakończonych sukcesem (Status=0) bądź błędem (Status<>0) wykonaniach procesów ETL, np.:

ETL_LOG (LogID, ProcessID, ExecDayTime, ExecStatus)

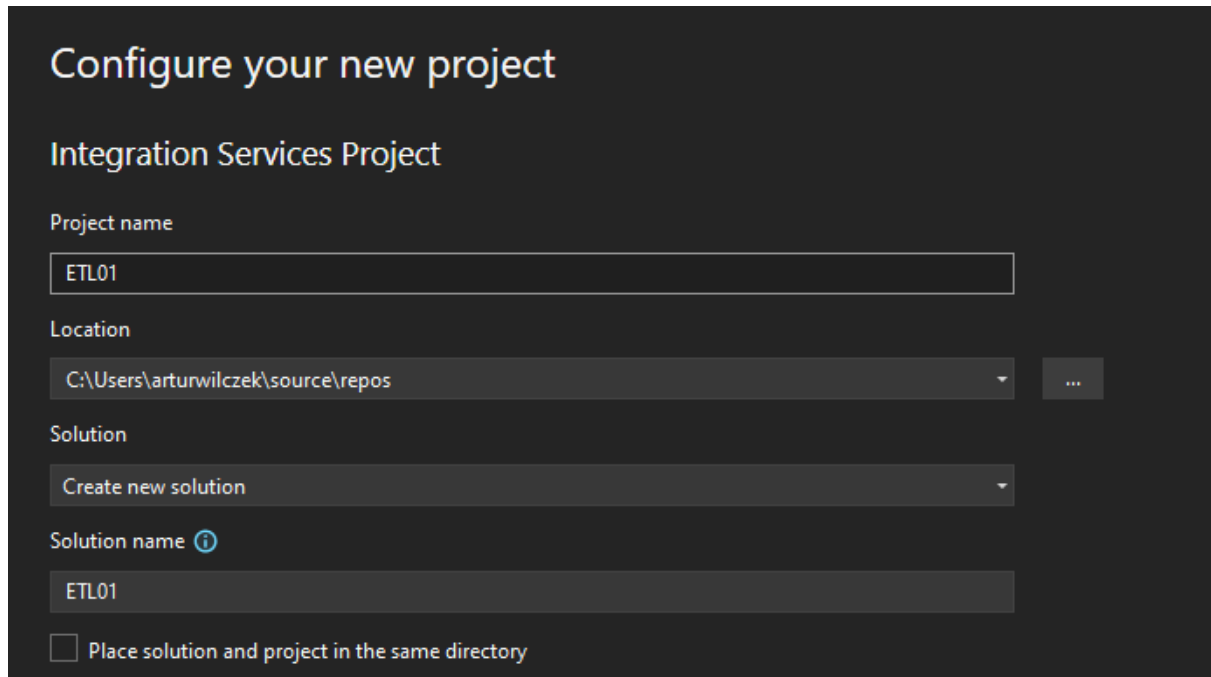
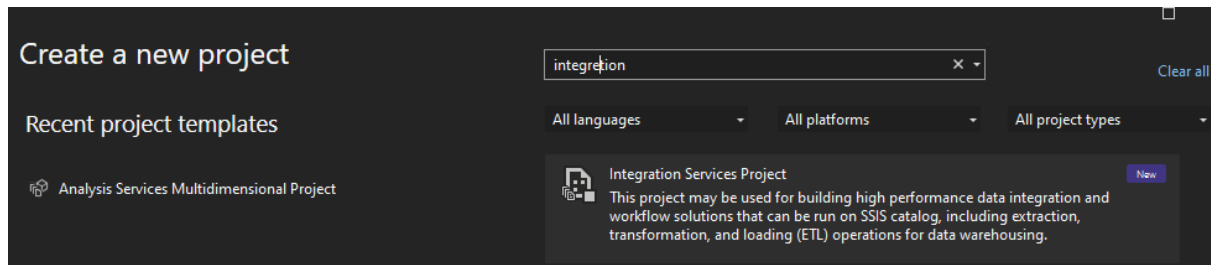
Rozwiązania:

```
CREATE TABLE STROZIK.ETL_LOG
(
    LogID INT IDENTITY(1,1) PRIMARY KEY NOT NULL,
    ProcessID UNIQUEIDENTIFIER NOT NULL,
    ExecDayTime DATETIME NOT NULL,
    ExecStatus INT NOT NULL
);
```

W ramach dynamicznego ustalania wartości dla atrybutów ProcessID, ExecDayTime, ExecStatus, zdecydowano się aby użyć wartości z sysssislog, która będzie tworzona w ramach kolejnych zadań tej listy. Aby to wykonać ProcessID będzie miał ten sam typ co sourceid w tabeli sysssislog czyli UNIQUEIDENTIFIER.

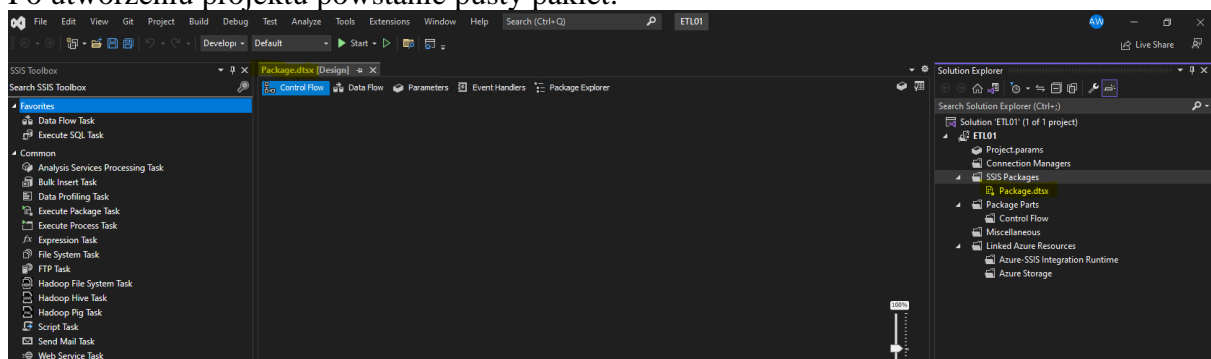
Zad. 4. Proces Extact – Transform - Load

Używając Visual Studio utworzyć projekt typu Integration Services (wybierając z Menu File - > New Project).



Projekt będzie służył do zbudowania procesu ETL wykonującego instrukcje SQL opracowane w ramach poprzedni list zadań dla przygotowania i wypełnienia danymi tabel kostki danych.

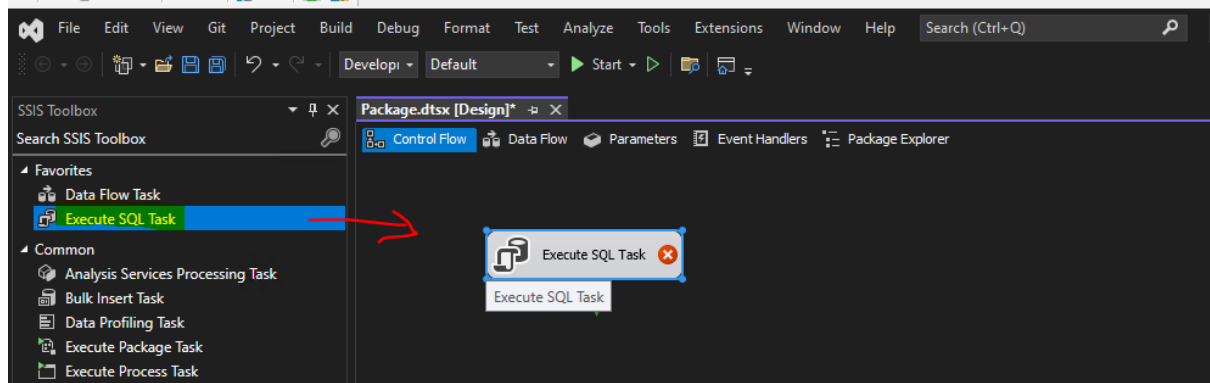
Po utworzeniu projektu powstanie pusty pakiet:



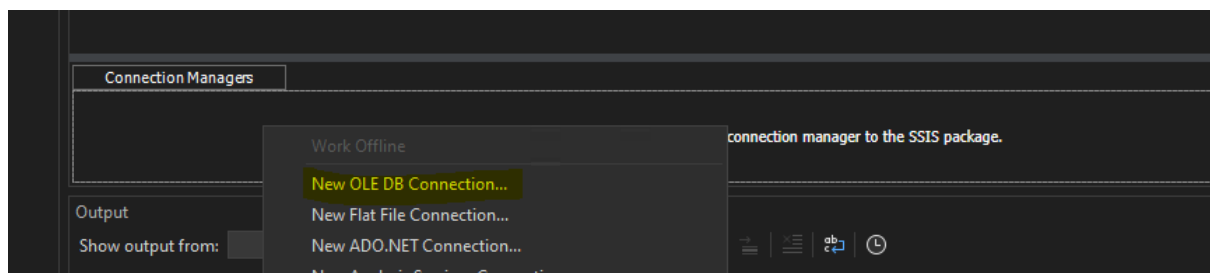
Pakiet będzie zawieraj projekt procesu ETL składającego się z różnych operacji. Pakiet jest jednostką, która może być uruchamiana.

Pakiet może zawierać zadania wykonujące instrukcje SQL.

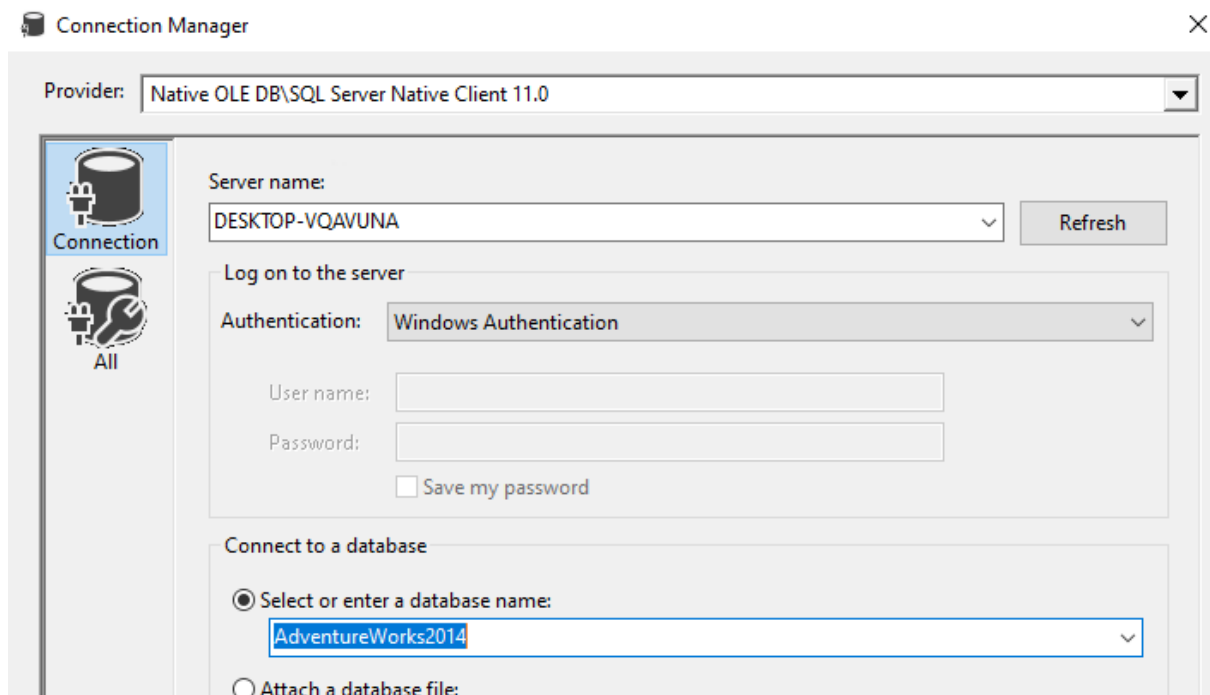
Uwaga: Do umieszczenia instrukcji SQL w treści pakietu użyć zadania Execute SQL Task:



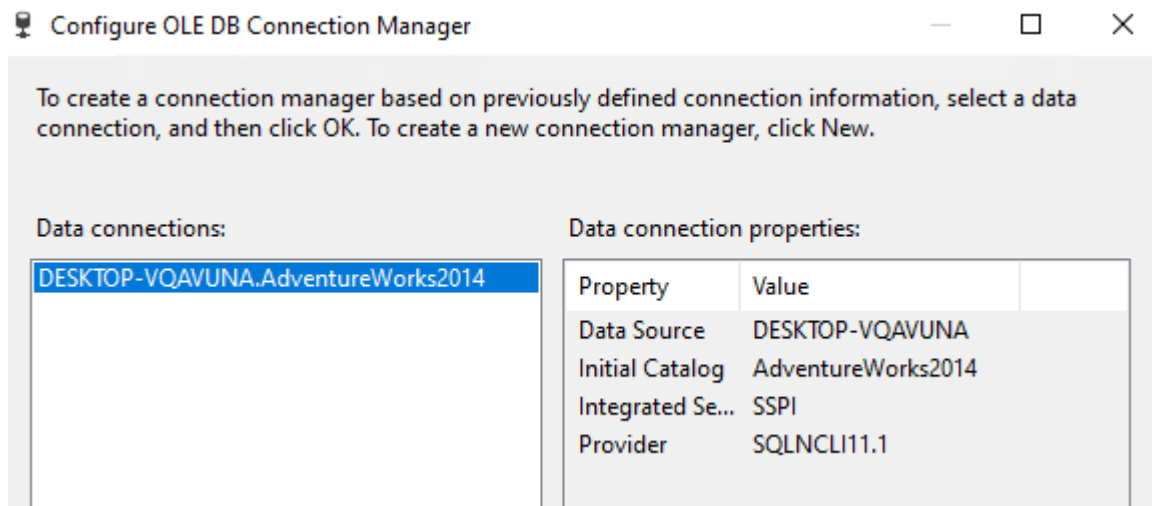
Zadania SQL wymagają podania połączenia do bazy danych, które wcześniej należy utworzyć w obszarze Connection Managers:



Utwórz połączenie do bazy danych **AdventureWorks** (*Polecenie New...*):



Wynik:



Utwórz pakiet wykonujący wcześniej przygotowane instrukcje SQL:

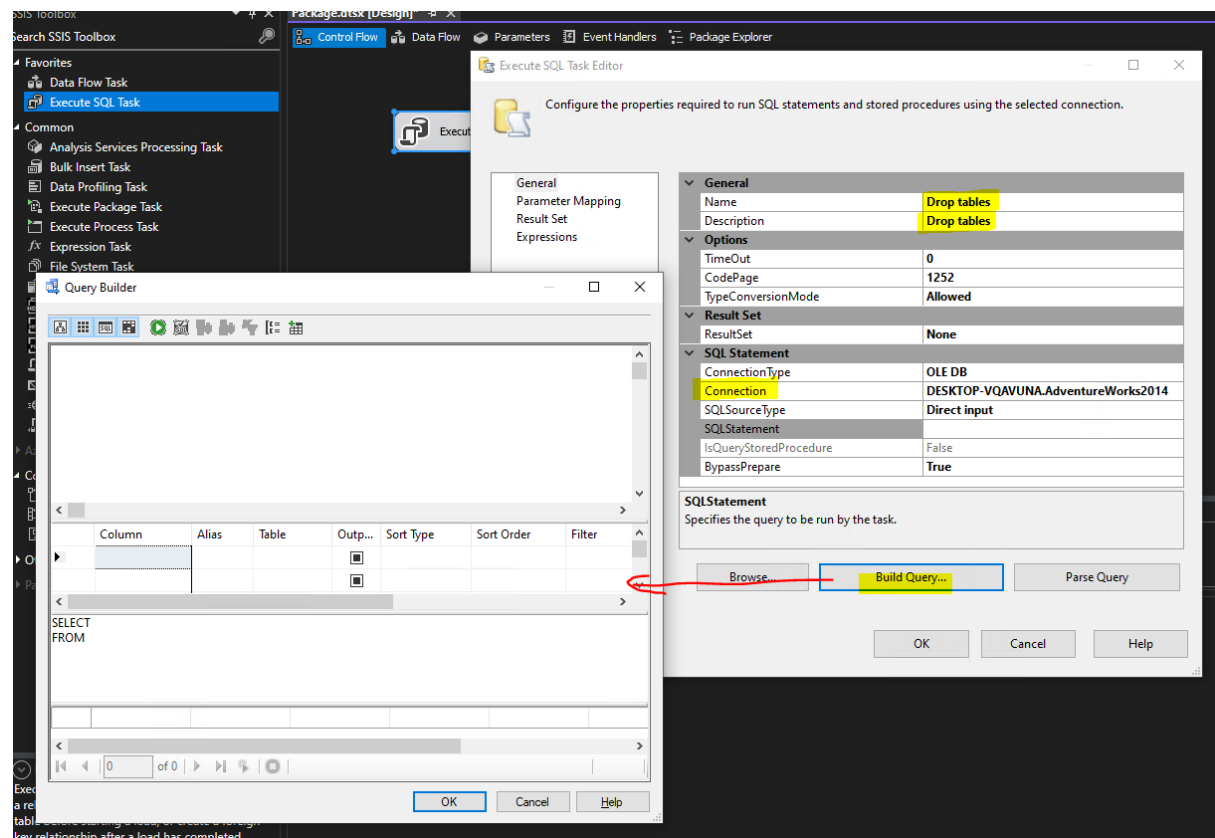
- Usuwa tabele z przedrostkiem DIM i FACT (tylko te, które istnieją),
- Utworzyć tabele z przedrostkiem DIM i FACT,
- Wypełnić tabele danymi (instrukcje INSERT INTO) i wykonać czyszczenia danych,
- Dodać więzy integralności,
- Obsłużyć błędy poprzez zapis statusu do tabeli dziennika
- (Opcjonalne) obsłużyć błędy i wyjątki w zakładce Event Handlers.
- (Opcjonalne) zapisać logi działania procesu w bazie danych

Uruchomić pakiet i sprawdzić i udokumentować poprawność jego działania.

Uwagi:

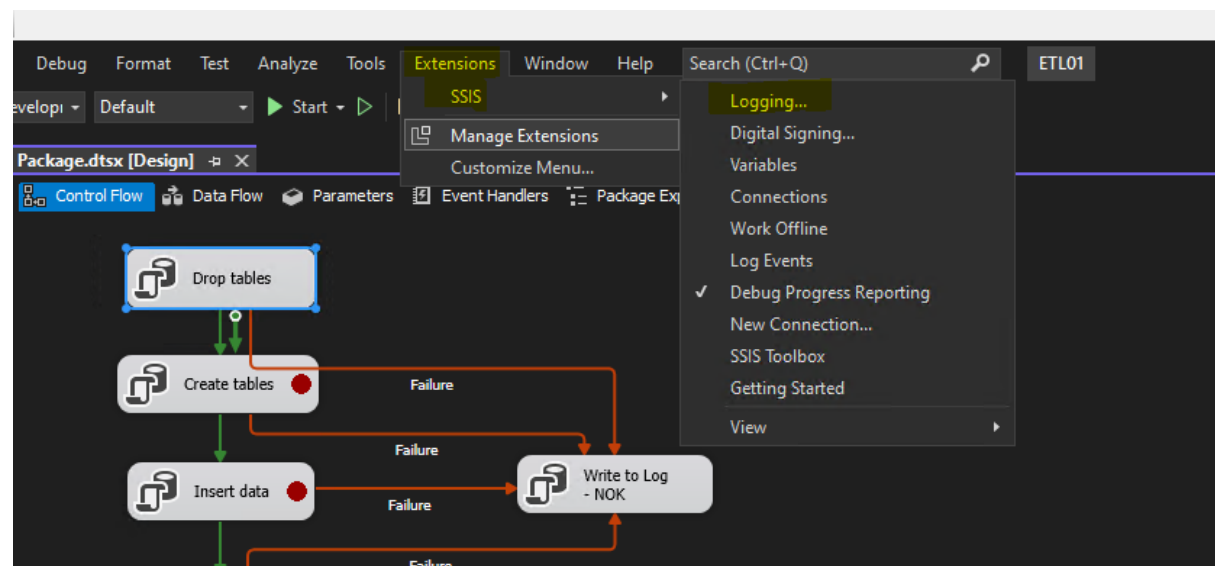
Tworząc zadania wykonanie instrukcji SQL należy wykorzystać stworzone wcześniej połączenie do bazy danych.

Instrukcje SQL można budować w Query Builder:



(Opcjonalne)

Logi procesu mogą być zapisywane do bazy danych poprzez dodatkową konfigurację:



Po jego konfiguracji np.:

Configure SSIS Logs: Package

Create and configure a new log to capture log-enabled events that occur at run time.

Containers:

- ☒ Package
 - ☒ Create tables
 - ☒ Drop tables
 - ☒ Insert data
 - ☒ Integrity
 - ☒ Write to Log - NOK
 - ☒ Write to Log - OK

Providers and Logs

Add a new log

Provider type: SSIS log provider for SQL Server

Select the logs to use for the container:

Name	Description	Configuration
<input checked="" type="checkbox"/> SSIS log provider f...	Writes log entries for even...	DESKTOP-VQAVUNA.Adv...

oraz określeniu co ma trafić do dziennika:

Create and configure a new log to capture log-enabled events that occur at run time.

Containers:

- ☒ Package
 - ☒ Create tables
 - ☒ Drop tables
 - ☒ Insert data
 - ☒ Integrity
 - ☒ Write to Log - NOK
 - ☒ Write to Log - OK

Providers and Logs

Select the events to be logged for the container:

Events	Description
<input checked="" type="checkbox"/> OnError	Handles error events. Use to define actions to p...
<input type="checkbox"/> OnExecStatusChanged	Handles changes of execution status. Use to def...
<input checked="" type="checkbox"/> OnInformation	Handles information events. The meanings of i...
<input type="checkbox"/> OnPostExecute	Handles post-execution events. Use to define p...
<input type="checkbox"/> OnPostValidate	Handles post-validation events. Use to define p...
<input type="checkbox"/> OnPreExecute	Handles pre-execution events. Use to define pre...

W bazie danych powstanie tabela z dziennikiem wykonywanych procesów:

Object Explorer

Connect

Tables

- System Tables
 - dbo.sysssislog
- File Tables
- External Tables
- Graph Tables
- dbo.AWBuildVersion
- dbo.DatabaseLog
- dbo.ErrorLog
- hd.DIM_CUSTOMER
- hd.DIM_PRODUCT
- hd.DIM_SALESPERSON
- hd.ETL_LOG
- HumanResources.Department
- HumanResources.Employee
- HumanResources.EmployeeDepartmentHistory
- HumanResources.JobCandidate
- HumanResources.Shift
- Person.Address
- Person.AddressType
- Person.BusinessEntity
- Person.BusinessEntityAddress
- Person.BusinessEntityContact

SQLQuery7.sql - DES...\arturwilczek (62)

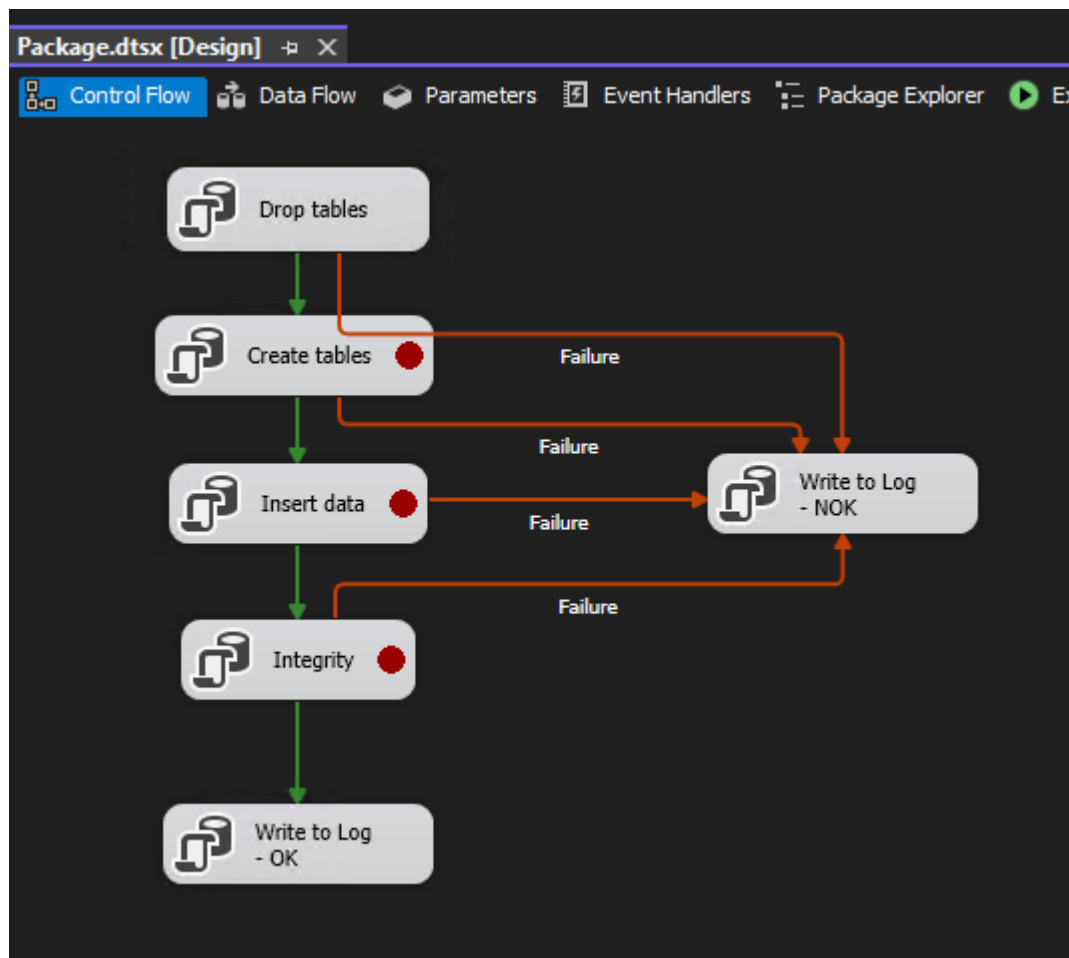
```

/***** Script for SelectTopNRows command from SSIS *****/
SELECT TOP (1000) [id]
, [event]
, [computer]
, [operator]
, [source]
, [sourceid]
, [executionid]
, [starttime]
, [endtime]
, [datacode]
, [databytes]
, [message]
FROM [AdventureWorks2014].[dbo].[sysssislog]
    
```

Results

id	event	computer	operator	source	sourceid	executionid	sta
1	PackageStart	DESKTOP-VQAVUNA	AzureAD\arturwilczek	Package	D3CFA8B5-A021-4C15-BF44-D7BFD4DFOAC7	7B709E9F-64CD-4373-85B7-99DC0409136B	2i
2	OnError	DESKTOP-VQAVUNA	AzureAD\arturwilczek	Drop tables	EBFB8F19-270B-4879-83ED-D1B5C9422985	7B709E9F-64CD-4373-85B7-99DC0409136B	2i
3	OnError	DESKTOP-VQAVUNA	AzureAD\arturwilczek	Package	D3CFA8B5-A021-4C15-BF44-D7BFD4DFOAC7	7B709E9F-64CD-4373-85B7-99DC0409136B	2i
4	PackageEnd	DESKTOP-VQAVUNA	AzureAD\arturwilczek	Package	D3CFA8B5-A021-4C15-BF44-D7BFD4DFOAC7	7B709E9F-64CD-4373-85B7-99DC0409136B	2i
5	PackageStart	DESKTOP-VQAVUNA	AzureAD\arturwilczek	Package	D3CFA8B5-A021-4C15-BF44-D7BFD4DFOAC7	F2028F0B-0678-4836-8196-E916F9277880	2i
6	PackageEnd	DESKTOP-VQAVUNA	AzureAD\arturwilczek	Package	D3CFA8B5-A021-4C15-BF44-D7BFD4DFOAC7	F2028F0B-0678-4836-8196-E916F9277880	2i

Przykładowe zadania w stworzonym pakiecie:



Rozwiązania:

Configure your new project

Integration Services Project

Project name
ETL01

Location
C:\Users\FilipStrózik(260377)\source\repos

Solution name ⓘ
ETL01

☐ Place solution and project in the same directory

Project will be created in "C:\Users\FilipStrózik(260377)\source\repos\ETL01\ETL01\"

Back Create

Connection Manager

Provider: Native OLE DB\SQL Server Native Client 11.0

Connection

Server name:
DESKTOP-B5JEBJO Refresh

Log on to the server

Authentication: Windows Authentication

User name: Password:

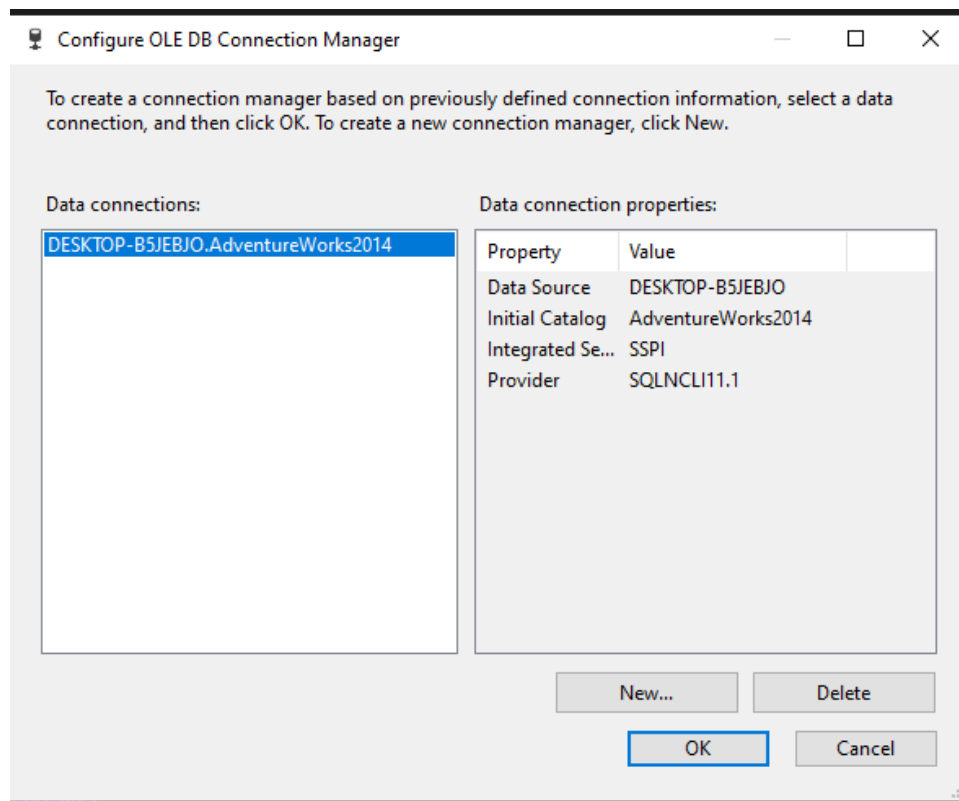
☐ Save my password

Connect to a database

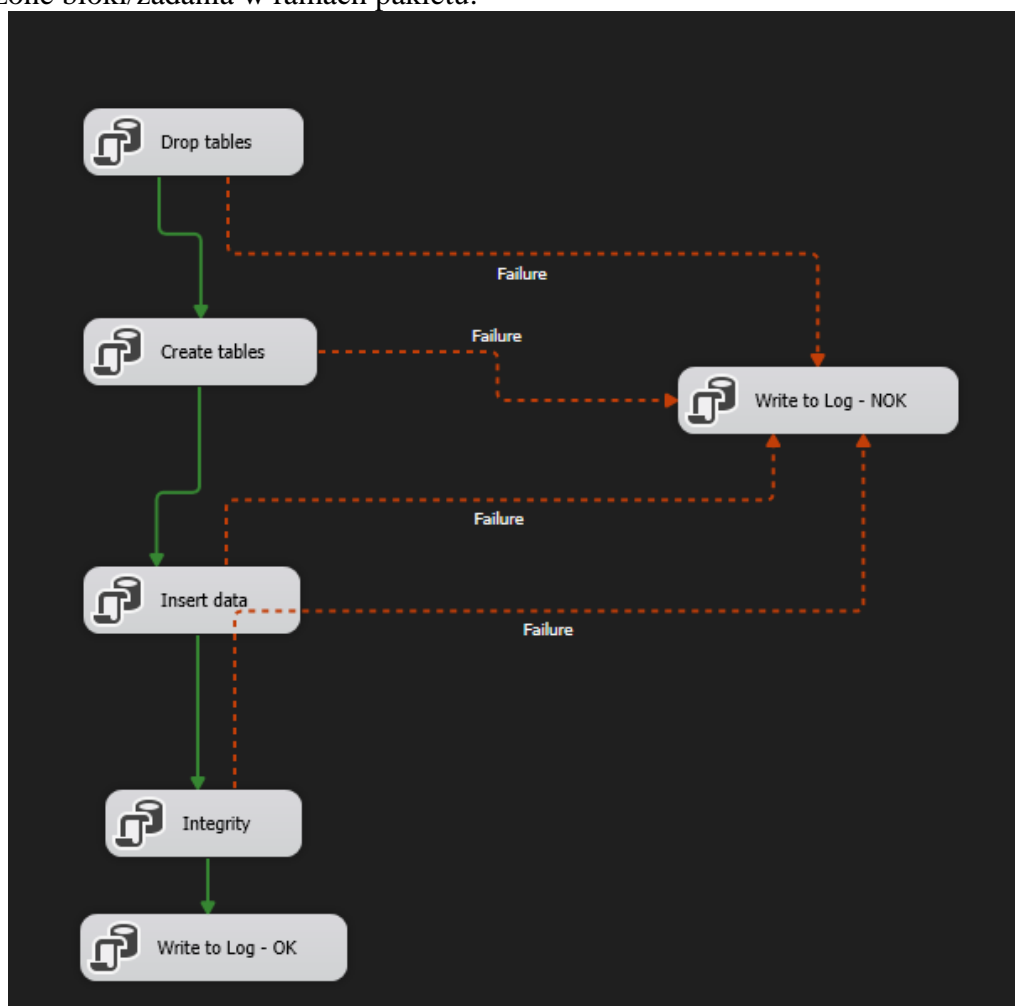
☒ Select or enter a database name:
AdventureWorks2014

☐ Attach a database file:
Browse... Logical name:

Test Connection OK Cancel Help



Stworzone bloki/zadania w ramach pakietu:



SQL w bloku „Drop tables“: (a)

```
DROP TABLE IF EXISTS STROZIK.FACT_SALES;  
  
DROP TABLE IF EXISTS STROZIK.DIM_PRODUCT;  
  
DROP TABLE IF EXISTS STROZIK.DIM_CUSTOMER;  
  
DROP TABLE IF EXISTS STROZIK.DIM SALESPERSON;  
  
DROP TABLE IF EXISTS STROZIK.DIM_TIME;  
  
DROP TABLE IF EXISTS STROZIK.months_names;  
  
DROP TABLE IF EXISTS STROZIK.weekday_names;
```

SQL w bloku „Create tables“: (b)

```
CREATE TABLE STROZIK.DIM_CUSTOMER  
(  
    CustomerID INT IDENTITY(1,1) PRIMARY KEY NOT NULL,  
    FirstName NVARCHAR(50) NOT NULL,  
    LastName NVARCHAR(50) NOT NULL,  
    Title NVARCHAR(8),  
    City NVARCHAR(30) NOT NULL,  
    TerritoryName NVARCHAR(50) NOT NULL,  
    CountryRegionCode NVARCHAR(3) NOT NULL,  
    [Group] NVARCHAR(50) NOT NULL  
);  
GO
```

```
--2.2  
CREATE TABLE STROZIK.DIM_PRODUCT  
(  
    ProduktID INT NOT NULL PRIMARY KEY,  
    Name NVARCHAR(50) NOT NULL,  
    ListPrice MONEY NOT NULL,  
    Color NVARCHAR(15),  
    SubCategoryName NVARCHAR(50),  
    CategoryName NVARCHAR(50),  
    Weight DECIMAL(8,2),  
    Size NVARCHAR(5),  
    IsPurchased BIT  
);  
GO
```

```
--2.3  
CREATE TABLE STROZIK.DIM SALESPERSON  
(  
    SalesPersonID INT NOT NULL PRIMARY KEY,  
    FirstName NVARCHAR(50) NOT NULL,  
    LastName NVARCHAR(50) NOT NULL,  
    Title NVARCHAR(8),  
    Gender NCHAR(1),  
    CountryRegionCode NVARCHAR(3),  
    [Group] NVARCHAR(50)  
);  
GO
```

```
--2.4  
CREATE TABLE STROZIK.FACT_SALES  
(  
    ProductID INT NOT NULL,  
    CustomerID INT NOT NULL,  
    SalesPersonID INT,  
    OrderDate INT NOT NULL,  
    ShipDate INT NOT NULL,  
    OrderQty INT NOT NULL,  
    UnitPrice MONEY NOT NULL,  
    UnitPriceDiscount MONEY NOT NULL,  
    LineTotal NUMERIC(38, 6) NOT NULL  
);
```

```
CREATE TABLE STROZIK.DIM_TIME (
    PK_TIME INT PRIMARY KEY,
    Rok INT,
    Kwartal INT,
    Miesiac INT,
    Miesiac_slownie VARCHAR(20),
    Dzień_tyg_slownie VARCHAR(20),
    Dzień_miesiaca INT
);
GO

CREATE TABLE STROZIK.months_names (
    month_number INTEGER,
    month_name VARCHAR(20)
);

INSERT INTO STROZIK.months_names (month_number, month_name)
VALUES (1, 'Styczeń'),
       (2, 'Luty'),
       (3, 'Marzec'),
       (4, 'Kwiecień'),
       (5, 'Maj'),
       (6, 'Czerwiec'),
       (7, 'Lipiec'),
       (8, 'Sierpień'),
       (9, 'Wrzesień'),
       (10, 'Październik'),
       (11, 'Listopad'),
       (12, 'Grudzień');

CREATE TABLE STROZIK.weekday_names (
    weekday_number INTEGER,
    day_name VARCHAR(20)
);

INSERT INTO STROZIK.weekday_names (weekday_number, day_name)
VALUES (1, 'Poniedziałek'),
       (2, 'Wtorek'),
       (3, 'Środa'),
       (4, 'Czwartek'),
       (5, 'Piątek'),
       (6, 'Sobota'),
       (7, 'Niedziela');
```

Zawartość poleceń jest taka sama jak w przypadku listy 4 i 5, aczkolwiek stwierdzono, że klucz główny dla wymiaru czasowego jednak powinien być reprezentowany jako liczba całkowita.

Kod SQL dla bloku „Insert data”: (c)

```
SET IDENTITY_INSERT STROZIK.DIM_CUSTOMER ON
INSERT INTO
    STROZIK.DIM_CUSTOMER (
        CustomerID,
        FirstName,
        LastName,
        Title,
        City,
        TerritoryName,
        CountryRegionCode,
        [Group]
    )
SELECT
    DISTINCT C.CustomerID,
    P.FirstName,
    P.LastName,
    P.Title,
    A.City,
    T.[Name],
    T.CountryRegionCode,
    T.[Group]
```

Filip Strózik 260377

```
FROM Sales.Customer C JOIN Person.Person P on C.PersonID = P.BusinessEntityID
JOIN Sales.SalesTerritory T on C.TerritoryID = T.TerritoryID
JOIN Sales.SalesOrderHeader H on C.CustomerID = H.CustomerID
JOIN Person.Address A on A.AddressID=H.ShipToAddressID
SET IDENTITY_INSERT STROZIK.DIM_CUSTOMER OFF;
```

```
INSERT INTO
    STROZIK.DIM_PRODUCT
SELECT
    p.ProductID,
    p.Name,
    p.ListPrice,
    p.Color,
    sc.Name,
    c.Name,
    p.Weight,
    p.Size,
    ~p.MakeFlag
FROM Production.Product p
LEFT JOIN Production.ProductSubcategory sc ON p.ProductSubcategoryID = sc.ProductSubcategoryID
LEFT JOIN Production.ProductCategory c ON sc.ProductCategoryID = c.ProductCategoryID;
```

```
INSERT INTO STROZIK.DIM_SALESPERSON
SELECT SP.BusinessEntityID, FirstName, LastName, Title, Gender, CountryRegionCode, [Group]
FROM Sales.SalesPerson SP
JOIN HumanResources.Employee E ON E.BusinessEntityID = SP.BusinessEntityID
JOIN Person.Person P ON E.BusinessEntityID = P.BusinessEntityID
LEFT JOIN Sales.SalesTerritory st ON sp.TerritoryID = st.TerritoryID;
```

```
INSERT INTO STROZIK.FACT_SALES
SELECT
    DISTINCT
    SOD.ProductID,
    SOH.CustomerID,
    SOH.SalesPersonID,
    YEAR(SOH.OrderDate) * 10000 + MONTH(SOH.OrderDate) * 100 + DAY(SOH.OrderDate),
    YEAR(SOH.ShipDate) * 10000 + MONTH(SOH.ShipDate) * 100 + DAY(SOH.ShipDate),
    SOD.OrderQty,
    SOD.UnitPrice,
    SOD.UnitPriceDiscount,
    SOD.LineTotal
FROM Sales.SalesOrderHeader SOH
LEFT JOIN Sales.SalesOrderDetail SOD ON SOD.SalesOrderID = SOH.SalesOrderID;
```

```
INSERT INTO STROZIK.DIM_TIME
SELECT DISTINCT
    YEAR(SOH.OrderDate) * 10000 + MONTH(SOH.OrderDate) * 100 + DAY(SOH.OrderDate) AS PK_TIME,
    YEAR(SOH.OrderDate) AS Rok,
    DATEPART(Q, SOH.OrderDate) AS Kwartal,
    MONTH(SOH.OrderDate) AS Miesiac,
    mn.month_name AS Miesiac_slownie,
    wdn.day_name AS Dzień_tyg_slownie,
    DAY(SOH.OrderDate) AS Dzień_miesiaca
FROM Sales.SalesOrderHeader SOH
    JOIN STROZIK.weekday_names wdn
    ON DATEPART(DW, OrderDate) = wdn.weekday_number
    JOIN STROZIK.months_names mn
    ON MONTH(OrderDate) = mn.month_number;
```

```
UPDATE STROZIK.DIM_PRODUCT
SET Color = 'Unknown'
WHERE Color IS NULL;
```

```
UPDATE STROZIK.DIM_PRODUCT
SET SubCategoryName = 'Unknown'
WHERE SubCategoryName IS NULL;
```

```
UPDATE STROZIK.DIM_CUSTOMER
SET CountryRegionCode = '000'
WHERE CountryRegionCode IS NULL;
```

```
UPDATE STROZIK.DIM_SALESPERSON
SET CountryRegionCode = '000'
WHERE CountryRegionCode IS NULL;
```



```
UPDATE STROZIK.DIM_CUSTOMER  
SET [Group] = 'Unknown'  
WHERE [Group] IS NULL;
```

```
UPDATE STROZIK.DIM SALESPERSON  
SET [Group] = 'Unknown'  
WHERE [Group] IS NULL;
```

Tzreba tu zauważyć, że skoro już zdecydowaliśmy się na utrzymanie typu danych przedstawiających daty to powinniśmy się tego trzymać. Dodatkowo, nie mogliśmy po prostu wprowadzić dat do wymiaru DIM_TIME bazując na datach z tabeli faktów, ponieważ nie odtworzylibyśmy potrzebnych dla nas wartości, które są oferowane przez właściwości typu DATETIME, dlatego łączamy odpowiednie tabele, by móc pobrać daty w właśnie typie danych DATETIME.

Kod SQL dla bloku „Integrity“: (d)

```
ALTER TABLE STROZIK.FACT_SALES  
ADD CONSTRAINT FK_ProductID FOREIGN KEY (ProductID) REFERENCES STROZIK.DIM_PRODUCT(ProductID),  
    CONSTRAINT FK_CustomerID FOREIGN KEY (CustomerID) REFERENCES STROZIK.DIM_CUSTOMER(CustomerID),  
    CONSTRAINT FK_SalesPersonID FOREIGN KEY (SalesPersonID) REFERENCES STROZIK.DIM_SALESPERSON(SalesPersonID),  
    CONSTRAINT FK_TIMEID FOREIGN KEY (OrderDate) REFERENCES STROZIK.DIM_TIME(PK_TIME);
```

Tutaj trzeba zauważyć, że zawarcie wszystkich więzłów integralności jest zwarte w jednej operacji / połączeniu, ponieważ już mamy wszystkie potrzebne wymiary. W przypadku listy 4 i 5 doszedł stopniowo do na wymiar DIM_TIME.

Kod SQL dla bloku „Write to Log - OK“: (e)

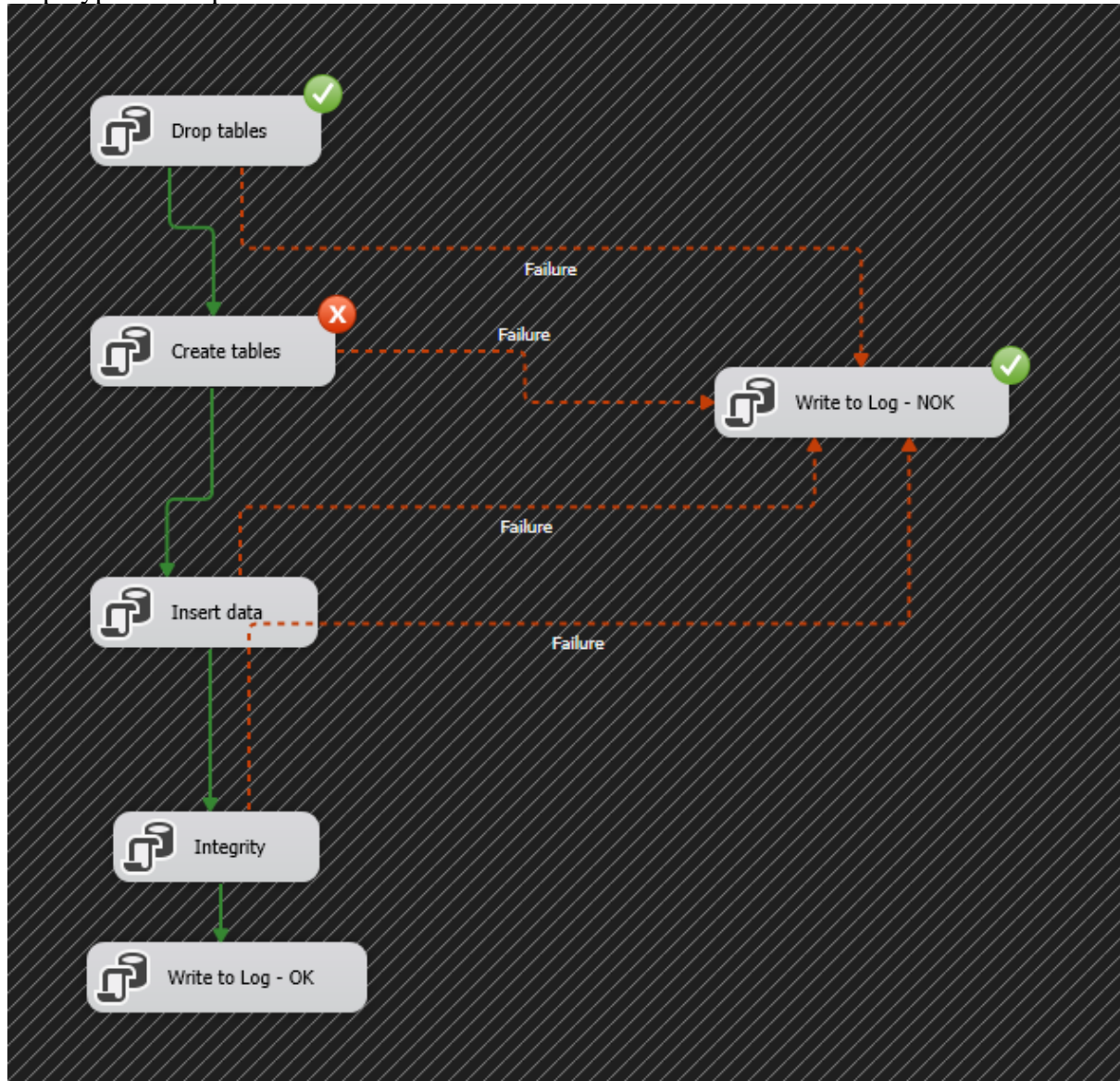
```
INSERT INTO STROZIK.ETL_LOG(ProcessID,ExecDayTime,ExecStatus)  
SELECT DISTINCT sourceid, endtime, datacode  
FROM dbo.sysssislog syslog LEFT JOIN STROZIK.ETL_LOG etllog on syslog.id = etllog.LogID  
    WHERE etllog.ProcessID is null;
```

Kod SQL dla bloku „Write to Log - NOK“: (e)

```
INSERT INTO STROZIK.ETL_LOG(ProcessID,ExecDayTime,ExecStatus)  
SELECT DISTINCT sourceid, endtime, datacode  
FROM dbo.sysssislog syslog LEFT JOIN STROZIK.ETL_LOG etllog on syslog.id = etllog.LogID  
    WHERE etllog.ProcessID is null;
```

Warto tutaj zauważyć że kod esquivel dla obu bloków logujących informacje do dziennika logów jest taki sam, ponieważ jesteśmy w stanie z opcjonalnie skonfigurowanej tabeli sysssislog otrzymać dane które są wpisywane podczas wykonywania kolejnych operacji w ramach uruchomienia pakietu. Skutkuje to tym że dynamicznie nadajemy wartość i atrybut potem które będziemy wpisywać do dziennika logów, a nie stałe wartości.

W przypadku niepowodzenia:



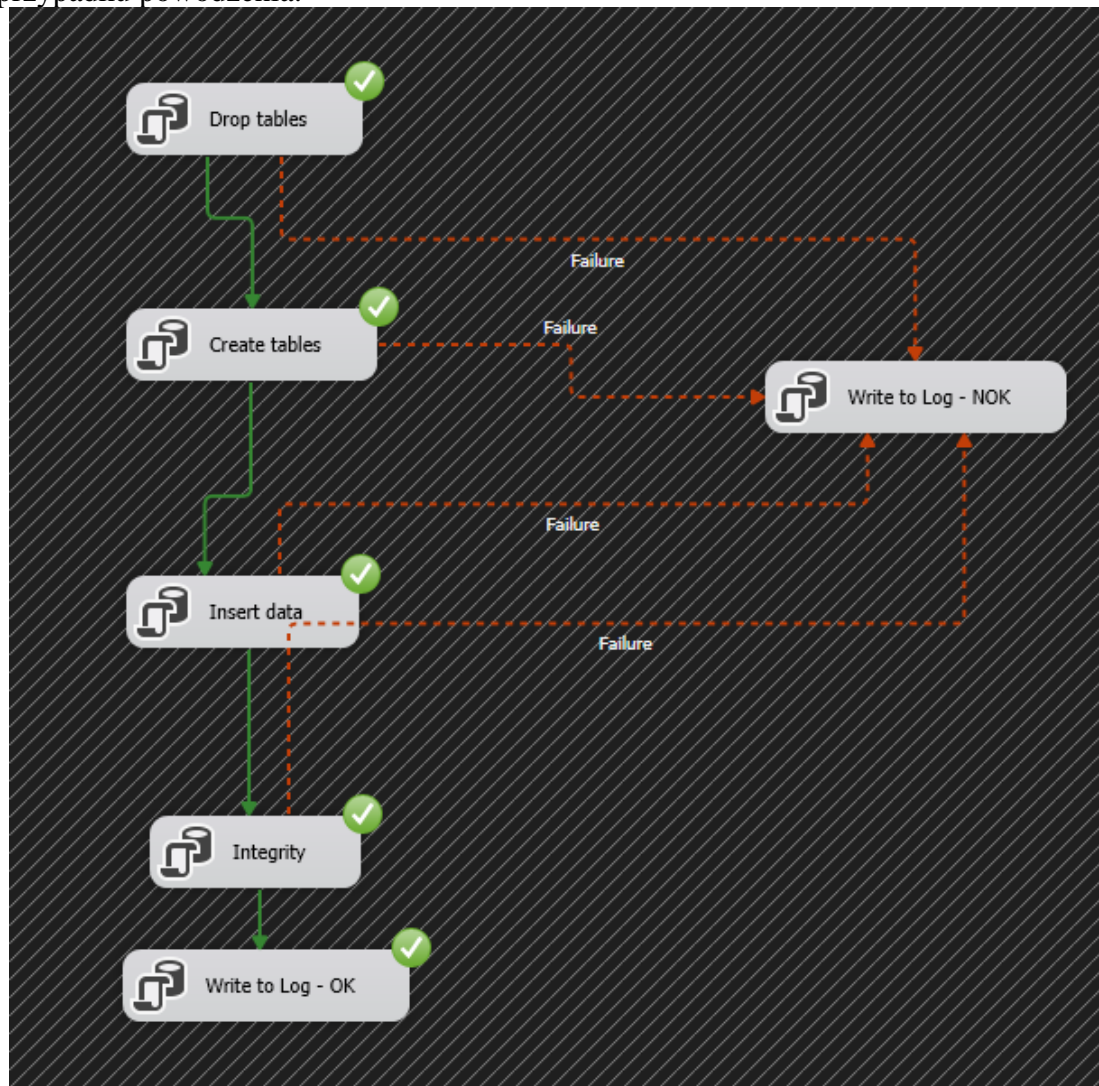
```
SELECT * FROM STROZIK.ETL_LOG;
```

116 %

Results Messages

	LogID	ProcessID	ExecDayTime	ExecStatus
1	1	AAB12417-1CC0-4B72-9BD9-79D35D6CB0D9	2023-05-20 10:21:42.000	-1073548784
2	2	46CFA758-5744-4EC2-A59F-CD33288F3121	2023-05-20 10:21:42.000	-1073548784
3	3	46CFA758-5744-4EC2-A59F-CD33288F3121	2023-05-20 10:21:42.000	0

W przypadku powodzenia:



```
SELECT * FROM STROZIK.ETL_LOG;
```

116 %

Results Messages

	LogID	ProcessID	ExecDayTime	ExecStatus
1	1	AAB12417-1CC0-4B72-9BD9-79D35D6CB0D9	2023-05-20 10:21:42.000	-1073548784
2	2	46CFA758-5744-4EC2-A59F-CD33288F3121	2023-05-20 10:21:42.000	-1073548784
3	3	46CFA758-5744-4EC2-A59F-CD33288F3121	2023-05-20 10:21:42.000	0
4	4	46CFA758-5744-4EC2-A59F-CD33288F3121	2023-05-20 10:21:42.000	1
5	5	46CFA758-5744-4EC2-A59F-CD33288F3121	2023-05-20 10:22:42.000	0

Na koniec możemy sprawdzić systemowy dziennik logów sysssislog:

id	event	computer	operator	source	sourceid	executionid	starttime	endtime	datacode	databytes	message
1	PackageStart	DESKTOP...	AzureAD\Filip Strózik...	Package	46CFA758-5744-4EC2-A5...	32B9A7A5-70...	2023-05-20 10:21:42.000	2023-05-20 10:21:42.000	0	0x	Beginning of package execution.
2	OnError	DESKTOP...	AzureAD\Filip Strózik...	Create t...	AAB12417-1CC0-4B72-9B...	32B9A7A5-70...	2023-05-20 10:21:42.000	2023-05-20 10:21:42.000	-1073548784	0x	Executing the query " -2.2 CREATE TABLE STROZIK...
3	PackageEnd	DESKTOP...	AzureAD\Filip Strózik...	Package	46CFA758-5744-4EC2-A5...	32B9A7A5-70...	2023-05-20 10:21:42.000	2023-05-20 10:21:42.000	-1073548784	0x	Executing the query " -2.2 CREATE TABLE STROZIK...
4	PackageEnd	DESKTOP...	AzureAD\Filip Strózik...	Package	46CFA758-5744-4EC2-A5...	32B9A7A5-70...	2023-05-20 10:21:42.000	2023-05-20 10:21:42.000	1	0x	End of package execution.
5	PackageStart	DESKTOP...	AzureAD\Filip Strózik...	Package	46CFA758-5744-4EC2-A5...	FEB754CB-7E...	2023-05-20 10:22:42.000	2023-05-20 10:22:42.000	0	0x	Beginning of package execution.
6	PackageEnd	DESKTOP...	AzureAD\Filip Strózik...	Package	46CFA758-5744-4EC2-A5...	FEB754CB-7E...	2023-05-20 10:22:44.000	2023-05-20 10:22:44.000	0	0x	End of package execution.

Wnioski:

Trzeba zwrócić uwagę, że lista była dobrym wprowadzeniem do procesu ETL, ponieważ pokazywała kluczowe kroki do skonfigurowania projektu oraz uruchomienia pakietu składającego się z wielu etapów. Lista głównie opierała się na procesie przekształcania oraz ładowania danych z tabel już istniejących, czyli mowa tu o T - Transform oraz L – Load. E czyli Extract odnosi się do wszelkiego wyodrębniania danych z różnego rodzaju plików, zawierających surowe dane w różnych formatach. Proces przekształcania był zauważalny w momencie, gdzie czyściliśmy tabele z wartości NULL na określona wartość. A Load w momencie wpisywania danych.

Proces ETL poniekąd automatyzuje kod SQL i opakuje go w skończone etapy, które mogą się wykonywać zależnie od sposobu połączenia z innymi etapami. Jest to bardzo wizualne podejście do kreowania takich złożonych przebiegów bloków SQL, porównując PL/SQL i tworzenie procedur.

Trzeba tu zwrócić uwagę na poziom zaawansowania wszelkich opcji. Podczas listy wystarczyło zaimportować gotowe polecenia, aczkolwiek potencjał kreowania bardzo złożonych zapytań wewnątrz pakietu jest niewyobrażalny.

Aby poprawnie obsługiwać logowanie błędów do dziennika, trzeba było ustalić połączenie Failure na „OR“, ponieważ chcemy logować błędy w przypadku co najmniej jednego niepowodzenia.

Opcjonalna funkcja dodająca logi do tabeli sysssislog była kluczowa aby móc wprowadzać do własnego dziennika informacje o powodzeniu lub niepowodzeniu procesu. W prawdziwym świecie mając systemowe logi potrzeba istnienia kolejnej tabeli przedstawiającej te same informacje spada, jednak stworzenie tabeli reprezentującej taki dziennik było jednym z głównych zadań na liście, dlatego zdecydowano pozostać przy dwóch.

Ważne jest, aby zauważyć, że proces ETL może być złożony i wymagać elastyczności w obsłudze różnych typów danych i różnych scenariuszy biznesowych. Dzięki narzędziom takim jak Visual Studio możemy postawić „breakpointy” w odpowiednich miejscach i próbować analizować proces w intuicyjny sposób.

Dzięki procesowi ETL możliwe jest wykonywanie bardzo złożonych operacji, które możemy dzielić na mniejsze i mniejsze kawałki, można by rzec „atomowe” czynności. Przypomina to podejście „Divide & Conquer”, w którym duży problem jest dzielony na mniejsze, bardziej zarządzalne podproblemy, a następnie rozwiązywane są one niezależnie, aby ostatecznie złożyć wyniki w celu uzyskania rozwiązania całego problemu. Jest to szczególnie ważne przy wysokiej złożoności problemu.

Uwaga!

- **Sprawozdanie, bez wniosków podsumowujących aspekt zagadnień analizowanych na zajęciach laboratoryjnych i zawartych w sprawozdaniu, jest automatycznie oceniane negatywnie!**