

Hurtownie danych Laboratorium Czw 11:15

Lista 5

Kajetan Pynka 254495

Zad 1

```
IF (EXISTS (SELECT *
            FROM INFORMATION_SCHEMA.TABLES
            WHERE TABLE_SCHEMA = 'Pynka'
            AND TABLE_NAME = 'FACT_SALES'))
BEGIN
    DROP TABLE Pynka.FACT_SALES
END;

IF (EXISTS (SELECT *
            FROM INFORMATION_SCHEMA.TABLES
            WHERE TABLE_SCHEMA = 'Pynka'
            AND TABLE_NAME = 'DIM_TIME'))
BEGIN
    DROP TABLE Pynka.DIM_TIME
END;

IF (EXISTS (SELECT *
            FROM INFORMATION_SCHEMA.TABLES
            WHERE TABLE_SCHEMA = 'Pynka'
            AND TABLE_NAME = 'ENUM_MONTHS'))
BEGIN
    DROP TABLE Pynka.ENUM_MONTHS
END;

IF (EXISTS (SELECT *
            FROM INFORMATION_SCHEMA.TABLES
            WHERE TABLE_SCHEMA = 'Pynka'
            AND TABLE_NAME = 'ENUM_WEEKDAYS'))
BEGIN
    DROP TABLE Pynka.ENUM_WEEKDAYS
END;

IF (EXISTS (SELECT *
            FROM INFORMATION_SCHEMA.TABLES
            WHERE TABLE_SCHEMA = 'Pynka'
            AND TABLE_NAME = 'DIM_CUSTOMER'))
BEGIN
    DROP TABLE Pynka.DIM_CUSTOMER
END;

IF (EXISTS (SELECT *
            FROM INFORMATION_SCHEMA.TABLES
            WHERE TABLE_SCHEMA = 'Pynka'
            AND TABLE_NAME = 'DIM_PRODUCT'))
BEGIN
    DROP TABLE Pynka.DIM_PRODUCT
END;
```

```

IF (EXISTS (SELECT *
            FROM INFORMATION_SCHEMA.TABLES
            WHERE TABLE_SCHEMA = 'Pynka'
            AND TABLE_NAME = 'DIM_SALESPERSON'))
BEGIN
    DROP TABLE Pynka.DIM_SALESPERSON
END;

```

Zad 2.

```

CREATE TABLE Pynka.ENUM_MONTHS (
    Miesiac INT PRIMARY KEY,
    Miesiac_slownie VARCHAR(11) NOT NULL
);

INSERT INTO Pynka.ENUM_MONTHS VALUES(1, 'styczeń');
INSERT INTO Pynka.ENUM_MONTHS VALUES(2, 'luty');
INSERT INTO Pynka.ENUM_MONTHS VALUES(3, 'marzec');
INSERT INTO Pynka.ENUM_MONTHS VALUES(4, 'kwiecień');
INSERT INTO Pynka.ENUM_MONTHS VALUES(5, 'maj');
INSERT INTO Pynka.ENUM_MONTHS VALUES(6, 'czerwiec');
INSERT INTO Pynka.ENUM_MONTHS VALUES(7, 'lipiec');
INSERT INTO Pynka.ENUM_MONTHS VALUES(8, 'sierpień');
INSERT INTO Pynka.ENUM_MONTHS VALUES(9, 'wrzesień');
INSERT INTO Pynka.ENUM_MONTHS VALUES(10, 'październik');
INSERT INTO Pynka.ENUM_MONTHS VALUES(11, 'listopad');
INSERT INTO Pynka.ENUM_MONTHS VALUES(12, 'grudzień');

CREATE TABLE Pynka.ENUM_WEEKDAYS (
    Dzień INT PRIMARY KEY,
    Dzień_tygodnia VARCHAR(12) NOT NULL
);

INSERT INTO Pynka.ENUM_WEEKDAYS VALUES(7, 'poniedziałek');
INSERT INTO Pynka.ENUM_WEEKDAYS VALUES(2, 'wtorek');
INSERT INTO Pynka.ENUM_WEEKDAYS VALUES(3, 'środa');
INSERT INTO Pynka.ENUM_WEEKDAYS VALUES(4, 'czwartek');
INSERT INTO Pynka.ENUM_WEEKDAYS VALUES(5, 'piątek');
INSERT INTO Pynka.ENUM_WEEKDAYS VALUES(6, 'sobota');
INSERT INTO Pynka.ENUM_WEEKDAYS VALUES(1, 'niedziela');

CREATE TABLE Pynka.DIM_TIME (
    PK_TIME INT PRIMARY KEY,
    Rok INT NOT NULL,
    Kwartał INT NOT NULL,
    Miesiac INT NOT NULL,

```

```

Miesiac_slownie VARCHAR(11) NOT NULL,
Dzien_tygodnia VARCHAR(12) NOT NULL,
Dzien_miesiaca INT NOT NULL
);

INSERT INTO Pynka.DIM_TIME
SELECT CAST(FORMAT(DateRec, 'yyyy') + FORMAT(DateRec, 'MM') + FORMAT(DateRec,
'dd') AS INT),
        YEAR(DateRec), DATEPART(quarter, DateRec), MONTH(DateRec),
EM.Miesiac_slownie,
        EW.Dzien_tygodnia, DAY(DateRec)
FROM (
        SELECT DISTINCT OrderDate "DateRec" FROM Sales.SalesOrderHeader
        UNION SELECT ShipDate FROM Sales.SalesOrderHeader) I
JOIN Pynka.ENUM_MONTHS EM ON EM.Miesiac=MONTH(DateRec)
JOIN Pynka.ENUM_WEEKDAYS EW ON EW.Dzien=DATEPART(WEEKDAY, DateRec);

```

Wynikowe tabele:

	PK_TIME	Rok	Kwartal	Miesiac	Miesiac_slownie	Dzien_tygodnia	Dzien_miesiaca
1	20110531	2011	2	5	maj	sroda	31
2	20110601	2011	2	6	czerwiec	czwartek	1
3	20110602	2011	2	6	czerwiec	piatek	2
4	20110603	2011	2	6	czerwiec	sobota	3
5	20110604	2011	2	6	czerwiec	poniedzialek	4
6	20110605	2011	2	6	czerwiec	niedziela	5
7	20110606	2011	2	6	czerwiec	wtorek	6
8	20110607	2011	2	6	czerwiec	sroda	7
9	20110608	2011	2	6	czerwiec	czwartek	8
10	20110609	2011	2	6	czerwiec	piatek	9
11	20110610	2011	2	6	czerwiec	sobota	10
12	20110611	2011	2	6	czerwiec	poniedzialek	11

	Miesiac	Miesiac_slownie
1	1	styczen
2	2	luty
3	3	marzec
4	4	kwieciec
5	5	maj
6	6	czerwiec
7	7	lipiec
8	8	sierpiec
9	9	wrzesien
10	10	pazdziernik
11	11	listopad
12	12	grudzien

	Dzien	Dzien_tygodnia
1	1	niedziela
2	2	wtorek
3	3	sroda
4	4	czwartek
5	5	piatek
6	6	sobota
7	7	poniedzialek

Zad 3.

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

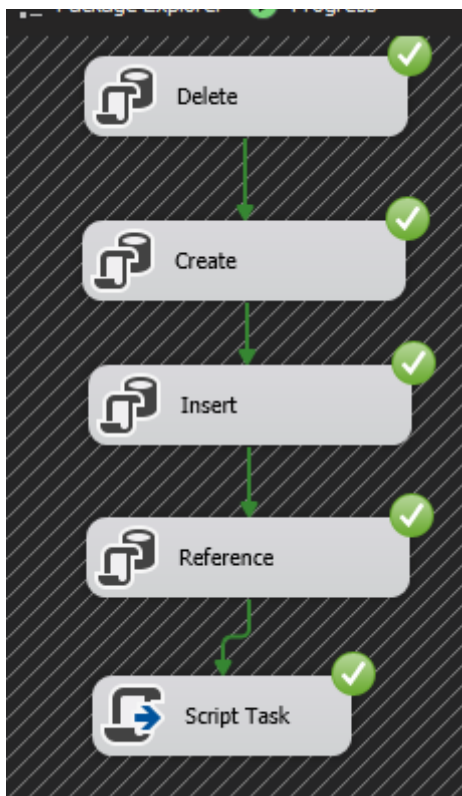
UPDATE Pynka.DIM_CUSTOMER SET CountryRegionCode='000' WHERE CountryRegionCode
IS NULL;
UPDATE Pynka.DIM_CUSTOMER SET [Group]='Unknown' WHERE [Group] IS NULL;

UPDATE Pynka.DIM SALESPERSON SET CountryRegionCode='000' WHERE
CountryRegionCode IS NULL;
UPDATE Pynka.DIM SALESPERSON SET [Group]='Unknown' WHERE [Group] IS NULL;
```

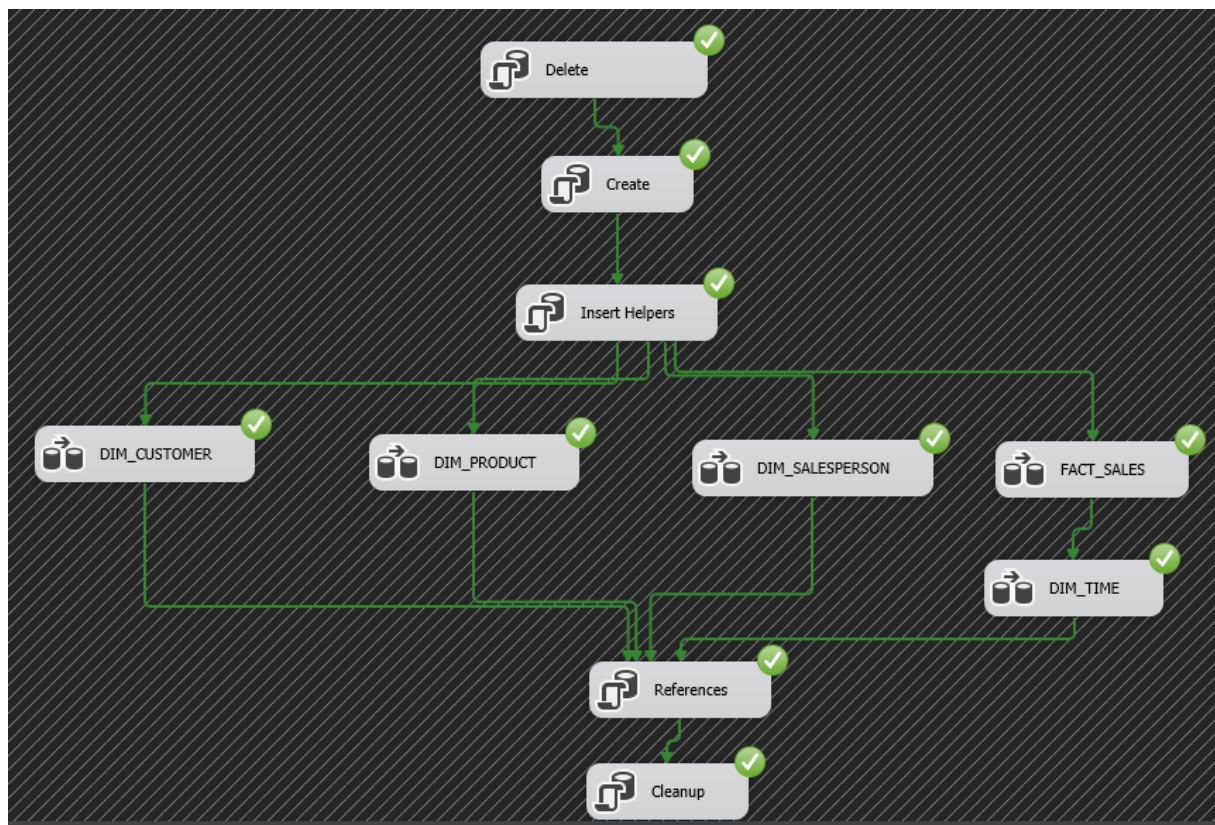
Po czyszczeniu:

	SalesPersonID	FirstName	LastName	Title	Gender	CountryRegionCode	Group
1	274	Stephen	Jiang	NULL	M	000	Unknown
2	275	Michael	Blythe	NULL	M	000	Unknown
3	276	Linda	Mitchell	NULL	F	000	Unknown
4	277	Jillian	Carson	NULL	F	000	Unknown
5	278	Garrett	Vargas	NULL	M	000	Unknown
6	279	Tsvi	Reiter	NULL	M	000	Unknown

Zad 4.



Zad 5.



Wnioski:

- W ramach listy przygotowaliśmy proces ETL korzystając z projektu „Integration Services” z Visual Studio, który pozwala nam w bardzo prosty sposób wywoływać kolejne kroki ekstrakcji, transformacji i wczytywania danych. W związku z tym mamy do czynienia w zasadzie z gotowym skryptem i za pomocą jednego kliknięcia możemy wykonać wszystkie wcześniej wymienione kroki (bez ponownego pytania zapytań czy wpisywania ich w dobrej kolejności). W przypadku zmian wystarczy zaktualizować to co potrzebne w ramach projektu „Integration Services”.
- Dodatkowo możemy rozbić poszczególne elementy zapytań SQL’owych na „bloczki” dostarczone przez VS w ramach projektu „Integration Services”. Może to pozwolić szybciej i łatwiej odczytywać poszczególne kroki procesu ETL a co za tym idzie w celu dokonania zmian nie trzeba „grzebać” w kodzie SQL tylko wystarczy „wyklikać” odpowiednie rzeczy w interfejsie graficznym.