

Hurtownie Danych - laboratorium Lista 5

Tworzenie wymiaru czasowego, przygotowanie procesu ETL

Uwaga: do realizacji zadań niezbędne jest przygotowanie listy 4.

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 ]
```

Ekstrakcja elementów daty – wynikiem jest liczba całkowita:

```
DATEPART ( datepart , date )
```

Możliwe argumenty:

datepart	skrót
year	yy, yyyy
quarter	qq, q
month	mm, m
dayofyear	dy, y
day	dd, d
week	wk, ww
weekday	dw
hour	hh
minute	mi, n
second	ss, s
millisecond	ms

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>

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

„ZPR PWr – Zintegrowany Program Rozwoju Politechniki Wrocławskiej”

Zadania do wykonania

Baza danych: **AdventureWorks**

Zad. 1. Przygotowanie powtarzalności procesu ETL

Przygotować instrukcję usuwającą każdą z tabel utworzonych w trakcie pracy nad listą 4.

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

Należy sprawdzić, czy dana tabela istnieje, używając instrukcji IF oraz informacji zawartych w widoku systemowym INFORMATION_SCHEMA.TABLES.

Zad. 2. Wymiar czasowy

Przygotować wymiar czasowy: utworzyć i wypełnić danymi tabelę DIM_TIME. Tabela DIM_TIME powinna być tabelą zawierającą wymiar czasowy (klucze obce do tej tabeli znajdują się w tabeli faktów).

Tabela DIM_TIME powinna zawierać następujące kolumny:

- PK_TIME (klucz główny – liczba całkowita postaci yyyyymmdd – format taki sam jak kolumn OrderDate, ShipDate)
- Rok
- Kwartał
- Miesiąc
- Miesiąc słownie (wykorzystać tabelę pomocniczą z 12 rekordami dokonać odpowiedniego złączenia)
- Dzień tygodnia słownie (wykorzystać tabelę pomocniczą z 7 rekordami dokonać odpowiedniego złączenia)
- Dzień miesiąca

	PK_Time	Rok	Kwartał	Miesiąc	Miesiąc_słownie	Dzień_tyg_słownie	Dzień_miesiąca
1	20110531	2011	2	5	maj	wtorek	31
2	20110601	2011	2	6	czerwiec	środa	1
3	20110602	2011	2	6	czerwiec	czwartek	2
4	20110603	2011	2	6	czerwiec	piątek	3
5	20110604	2011	2	6	czerwiec	sobota	4
6	20110605	2011	2	6	czerwiec	niedziela	5
7	20110606	2011	2	6	czerwiec	poniedziałek	6
8	20110607	2011	2	6	czerwiec	wtorek	7
9	20110608	2011	2	6	czerwiec	środa	8
10	20110609	2011	2	6	czerwiec	czwartek	9
11	20110610	2011	2	6	czerwiec	piątek	10
12	20110611	2011	2	6	czerwiec	sobota	11
13	20110612	2011	2	6	czerwiec	niedziela	12
14	20110613	2011	2	6	czerwiec	poniedziałek	13
15	20110614	2011	2	6	czerwiec	wtorek	14
16	20110615	2011	2	6	czerwiec	środa	15
17	20110616	2011	2	6	czerwiec	czwartek	16
18	20110617	2011	2	6	czerwiec	piątek	17

Przykładowe rozwiązanie – tabela DIM_TIME.

Czy przygotowanie tej tabeli można zoptymalizować?

Zad. 3. Elementarne czyszczenie danych

Zamienić wszystkie wartości NULL:

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

Zad. 4. Proces Extact – Transform - Load

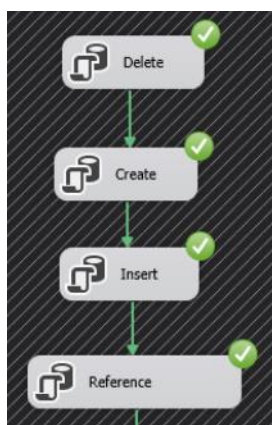
Używając Visual Studio utworzyć projekt typu Integration Services (wybierając z Menu File - > New Project) zawierający zapytania SQL opracowane w zadaniach 1-4.1 z listy 4 oraz w zadaniach 1-3 z listy 5.

Uwaga: Do umieszczenia instrukcji SQL w treści pakietu użyć zadania Execute SQL Task (Menu – View – Toolbox).

Utworzony pakiet powinien działać sekwencyjnie i wykonywać następujące zadania:

- Usunąć tabele z przedrostkiem DIM i FACT (oczywiście usunąć tylko te, które istnieją),
- Utworzyć tabele z przedrostkiem DIM i FACT,
- Wypełnić tabele danymi (instrukcje INSERT INTO),
- Dodać więzy integralności z zadania 4.1 z listy 4 (bez sprawdzania poprawności integralności),
- Obsłużyć błędy i wyjątki – zakładka Event Handlers,
- Wyświetlić informację o pozytywnie zakończonym procesie.

Uruchomić pakiet i sprawdzić poprawność jego działania.



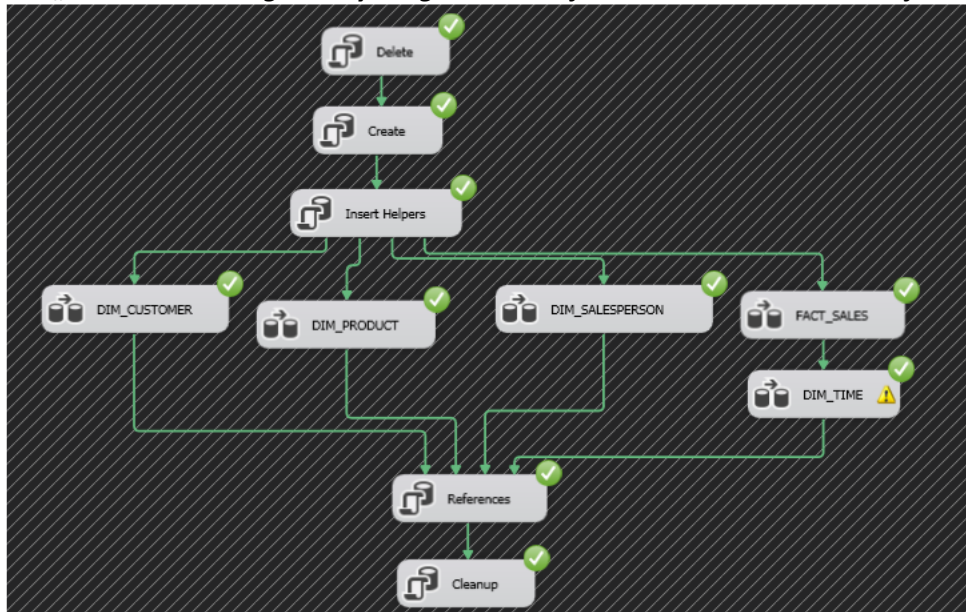
Przykładowe rozwiązanie.

Zad. 5. ETL (prawie) bez SQLa

Przygotować proces ETL analogiczny do opisanego w zad. 4. **Dla wymiaru czasowego i co najmniej jednego innego wymiaru** przygotować import danych korzystając z narzędzi dostępnych w zakładce Data Flow, m.in. OLE DB Source/Destination, Merge Join, Sort, Transformations, Derived Column, Fuzzy Lookup, Fuzzy Grouping, itp.

<https://learn.microsoft.com/en-us/sql/integration-services/data-flow/transformations/integration-services-transformations?view=sql-server-ver16>

„ZPR PWr – Zintegrowany Program Rozwoju Politechniki Wrocławskiej”



Przykładowe rozwiązanie

Rozwiązania:

Wnioski: