**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.table\_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/datepart-transact-sql?view=sql-server-ver15>

**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 yyyymmdd – 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



*Przykładowe rozwiązanie – tabela DIM\_TIME.*

**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**

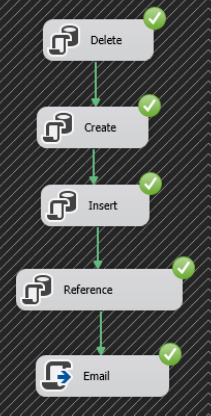
Używając Visual Studio utworzyć projekt typu Integration Services (wybierając z Menu File -> New Project) zawierający zapytania SQL opracowane w zadaniach 1-4.1 z listy 4 oraz w zadaniach 1-3 z listy 5.

*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:

1. Usunąć tabele z przedrostkiem DIM i FACT (oczywiście usunąć tylko te, które istnieją),
2. Utworzyć tabele z przedrostkiem DIM i FACT,
3. Wypełnić tabele danymi (instrukcje INSERT INTO),
4. Dodać więzy integralności z zadania 4.1 z listy 4 (bez sprawdzania poprawności integralności),
5. Obsłużyć błędy i wyjątki – zakładka Event Handlers,
6. Wysłać informację o pozytywnie zakończonym procesie na swój adres mailowy,

Uruchomić pakiet i sprawdzić poprawność jego działania.



*Przykładowe rozwiązanie.*

**Zad. 5. ETL (prawie) bez SQLa**

Przygotować proces ETL analogiczny do opisanego w zad. 4. Import danych przygotować korzystając z narzędzi dostępnych w zakładce Data Flow, m.in. OLE DB Source/Destination, Merge Join, Sort, Derived Column, Fuzzy Lookup, Fuzzy Grouping, itp.



*Przykładowe rozwiązanie*

**Rozwiązania:**

**Wnioski:**

***Uwaga!***

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