

Hurtownie danych

Lista 5 – Tworzenie wymiaru czasowego, przygotowanie procesu ETL



Aleksander stepaniuk

nr. indeksu: 272644

Politechnika Wrocławska, Informatyka Stosowana

**Rozwiązania:**

Zadanie 1.

A screenshot of a computer screen

AI-generated content may be incorrect.

Zadanie 2.

Tworzenie DIM\_TIME

A screenshot of a computer

AI-generated content may be incorrect.

Tabele pomocnicze:

A screenshot of a computer program

AI-generated content may be incorrect.

Wypełniamy danymi:

A screenshot of a computer

AI-generated content may be incorrect.

Zadanie 3.

A screenshot of a computer code

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

Zadanie 4.

A screenshot of a diagram

AI-generated content may be incorrect.

1. A screenshot of a computer

   AI-generated content may be incorrect.
2. A screenshot of a computer

   AI-generated content may be incorrect.
3. A screenshot of a computer

   AI-generated content may be incorrect.
4. A screenshot of a computer

   AI-generated content may be incorrect.
5. A screenshot of a computer

   AI-generated content may be incorrect.
6. PRINT ‘Proces zakończony pomyślnie’;

A screenshot of a computer screen

AI-generated content may be incorrect.

Zadanie 5.

A diagram of a software project

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

# Wnioski:

Napisanie skryptu, który automatycznie usuwa wszystkie tabele DIM i FACT, pozwoliło na to, że za każdym razem mamy czyste środowisko do pracy.

Przy próbie implementacji importowania danych do DIM\_TIME w SSIS w ETL w formie noSQL moje pierwsze podejście obejmowało próbę obróbki dat ShipDate oraz OrderDate bez wielokrotnego pobierania źródła danych (np. za pomocą Unpivot), ale końcowo uznałem za prostsze (i działające) użycie Union All oraz dwóch źródeł danych.

W Derived Column nauczyłem się uważać na wartości NULL i najpierw je odfiltrować, żeby nie napotkać błędu przy rzutowaniu typów.

Korzystanie z Lookup, gdzie można było dołączhć nazwy miesięcy i dni pokazało, jak łatwo uzupełnić atrybuty bez pisania skomplikowanych zapytań.

Czyste zapytania SQL w Execute SQL Task dały nam pewność, że każda kolejna instrukcja wykona się w ustalonej kolejności, a ewentualny błąd zatrzyma proces dokładnie tam, gdzie trzeba – można też było łatwiej obserwować przebieg takiego procesu i jego ewentualna naprawa w przypadku błędu była o wiele prostsza niż dla klasycznego kodu w SQL.

Event Handlers służą do globalnego łapania wyjątków i logowania ich do plików, jest to rozwiązanie, które w niektórych sytuacjach może być wygodne i praktyczne, jednak brak domyślnego error outputu w konsoli mnie osobiście irytował i nie był zbyt intuicyjny przy próbie naprawy kodu.

Przy pracy nad tymi zadaniami przekonać się można, że komponenty SSIS potrafią naprawdę dużo, zwłaszcza gdy chcemy zbudować wygodny do wykorzystania prototyp ETL. Takie rozwiązanie pozwala łączyć szerokie możliwości języka SQL z wizualną wygodą graficznej reprezentacji w SSIS, co sprawia że końcowy projekt staje się bardziej solidny, elastyczny i odporny na typowe błędy danych. Jednak warto wspomnieć że robienie tego typu zapytań graficznie jest o wiele bardziej czasochłonne niż napisanie krótkiego kodu w czystym SQL.