

KONCEPCJA PROCESU ETL ZADANIE ONWELO

Poniżej opisano w szczegółowy sposób zastosowaną metodologię pobierania, przetwarzania oraz kontrolowania stanu na określoną datę danych ekonomicznych dla konkretnych krajów przypisanych kolejno do regionów.

1. Pobranie danych ze strony <https://data.nasdaq.com/> z wykorzystaniem API

API pozwala na pobranie danych w formie tabelarycznej. Dane w dalszej kolejności możemy przefiltrować oraz stworzyć zapytanie DDL w celu łatwego wprowadzenia danych do lokalnej bazy Oracle. Warto zauważyć, że pobieranie danych odbywa się poprzez podanie *Data Link Code* skojarzonego z określonym zbiorem danych, które zapisane są przy pomocy określonej formuły, zakończonej skrótem nazwy danego kraju. Pozwala nam to w prosty sposób stworzyć klucze używane do ściągnięcia danych. Finalny kod zapisany został w pliku GET_DATA_NASDAQ.py w przesłanej wiadomości.

2. Stworzenie procesu monitorującego zmiany powiązań kraju z regionem

Cały proces został zbudowany w oparciu o język skryptowy PL/SQL. Przygotowane zapytanie przechowuje w oddzielnej tabeli dane aktualne powiązań oraz tworzy tabele historyczne w momencie wywołania (możliwość przyszłych analiz lub odtworzenia stanu z przeszłości). Finale dane z przechowywane są w oddzielnej tabeli, gdzie pola *start_date*, *end_date* wskazują okres rozpoczęcia/zakończenia przypisania kraju do określonego regionu (wartość NULL oznacza stan aktualny). Zapytanie w aktualnym stanie zostało zbudowane w oparciu o konkretne założenia:

- Zapytanie będzie wykonywane nie częściej niż raz dziennie
- Tabele z danymi historycznymi/aktualnymi pozostają zapisane w bazie ze względu na możliwość wykorzystania w analizach
- W momencie zmiany powiązania Kraju z regionem należy zmodyfikować plik excel dostarczony w wiadomości oraz zastąpić aktualne polecenie *INSERT ALL*