

Zadania kwalifikacyjne

Wstęp

Zadania proszę przysyłać przez stronę warsztatów, najlepiej w formacie PDF. Pytania dotyczące treści i rozwiązań można wysyłać na mojego maila aurbala@gmail.com, tytuł maila dobrze jest rozpocząć od [WWW] – zminimalizuje to szansę, że zignoruję wiadomość (InPost – Potwierdzenie nadania przesyłki też działa całkiem dobrze, ale proszę nie próbować). Nie bójcie się także pytać o pojęcia, których nie rozumiecie – zdaję sobie sprawę z tego, że nie każdy miał do czynienia np. z PostgreSQL-em, choć oczywiście zachęcam do samodzielnego zapoznania się z *nieznanymi* narzędziami, bo nigdy nie wiadomo, kiedy mogą się przydać ;)

Zadanie 1 (5 punktów)

Czy dany proces jest procesem ETL?

- Pobieranie postów z Twittera Janusza Piechocińskiego i zapisywanie ich do plików.
- Pobieranie postów z Twittera Janusza Piechocińskiego, przydzielenie im kategorii, zapis do odpowiedniej tabeli w bazie danych.
- Wyciągnięcie danych odnośnie nowych zakupów z bazy danych MySQL, wyliczenie statystyk (np. liczba zamówień, średnia cena zamówienia) i ich zapis do bazy PostgreSQL.
- Ładowanie plików .csv do Google Data Studio i stworzenie na ich podstawie dashboardów.
- Ładowanie danych z hurtowni danych, denormalizacja schemy, zapis danych do bazy biznesowej.

Zadanie 2 (5 punktów)

Dla każdego procesu z zadania 1. wyodrębnij fragment TRANSFORM.

Np. dla procesu: *Pobranie wierszy z pliku i wstawienie unikalnych do bazy wystarczającą odpowiedzią jest usunięcie duplikatów.*

Zadanie 3 (5 punktów)

Wymień 5 źródeł danych i (opcjonalnie) sposób/narzędzie do interakcji z tym źródłem.

Przykład: Baza PostgreSQL (PostgreSQL JDBC Driver).

Zadanie 4 (5 punktów)

Zaproponuj sposób postępowania, gdy nie uda się załadować jakichś rekordów do bazy docelowej. Kiedy ta metoda się sprawdza, a kiedy nie?