**Raport – Projekt Zasadniczy HDiSE**

Hurtownia danych na potrzeby analizy konsumpcji treści cyfrowych

Prowadzący:   
Marcin Gorawski

Krzysztof Pasterak

Autorzy:

Cezary Pastor

Jakub Kuk

**Wprowadzenie**

W ramach projektu mieliśmy na celu stworzenie systemu do zbierania danych o firmach związanych z treściami cyfrowymi. Na przykład takiej analizy wzięliśmy przykład firmy NETFLIX. Dane wzięliśmy ze strony [Kaggle](https://www.kaggle.com/).

Dane dla hurtowni usługi streamingowej linki:  
  
1.Dane o subskrybentach <https://www.kaggle.com/datasets/mauryansshivam/netflix-ott-revenue-and-subscribers-csv-file>

2. Filmy i seriale

<https://www.kaggle.com/datasets/durgeshrao9993/netflix-shows-dataset>

3. Filmy i seriale

<https://www.kaggle.com/datasets/shivamb/netflix-shows>

4. Najnowsze filmy i seriale na rok 2022

<https://www.kaggle.com/datasets/senapatirajesh/netflix-tv-shows-and-movies>

5. Cena subskrypcji dla krajów

<https://www.kaggle.com/datasets/prasertk/netflix-subscription-price-in-different-countries>

6. Oryginalne produkcje NETFLIX oraz ich ocena na IMDB

<https://www.kaggle.com/datasets/luiscorter/netflix-original-films-imdb-scores>

7. Cena akcji NETFLIX 2002-2022

<https://www.kaggle.com/datasets/meetnagadia/netflix-stock-price-data-set-20022022>

Zacznijmy od struktury hurtowni danych, która reprezentuje osobne tabele.

Obraz zawierający tekst, zrzut ekranu, numer, Równolegle

Opis wygenerowany automatycznie

Obraz zawierający tekst, zrzut ekranu, Czcionka, numer

Opis wygenerowany automatycznie

Rysunek 1. Zdjęcie przestawia powstałe tabele

Działanie aplikacji:

W projekcie użyto oprogramowania Talend Open Studio for Data Integration 8.0.1 jako program ETL.

Obraz zawierający tekst, zrzut ekranu, oprogramowanie, wyświetlacz

Opis wygenerowany automatycznie

Możemy zauważyć że program posiada „jobs” są to zadania jakie program wykonuję w tym przypadku jest to ekstrakcja (extract) danych z plików typu exel a następnie transformowanie (transform) tych danych poprzez mapowanie zmiennych a następnie ładowanie (load) do bazy danych lub tworzenie nowych tabel z tymi danymi.

1. tDBConnection – łączenie się z bazą danych, poprzez podanie jej danych.  
   Obraz zawierający tekst, linia, numer, Czcionka

   Opis wygenerowany automatycznie
2. tFileInputDelimited – służy do odczytu plików, w tym przypadku pliki exel oraz danie jaki ma zostać użyty separator między danymi „ ; ” oraz jakie są pola i jakie mają typy .Obraz zawierający tekst, zrzut ekranu, wyświetlacz, numer

   Opis wygenerowany automatycznie  
     
   Obraz zawierający tekst, zrzut ekranu, linia, Czcionka

   Opis wygenerowany automatycznie
3. tMap – mapowanie danych z plików na zmienne tabeli.  
   Obraz zawierający tekst, zrzut ekranu, wyświetlacz, oprogramowanie

   Opis wygenerowany automatycznie
4. tDBOutput – akcja jak ma być wykonana na podłączonej bazie, w tym przypadku jest to tworzenie tabeli o nazwie „netflix\_tv\_shows” o ile nie istnieje już taka  
   Obraz zawierający tekst, zrzut ekranu, oprogramowanie, Czcionka

   Opis wygenerowany automatycznie
5. tDBCommit – komitowanie zmian na bazę.

Struktura aplikacji back-end:

Obraz zawierający tekst, zrzut ekranu, oprogramowanie, Oprogramowanie multimedialne

Opis wygenerowany automatycznie

1. Controllers – w tym folderze znajdują się kontrolery, dzięki którym możliwa jest interakcja z bazą danych postgres poprzez REST API. Dla każdej tabeli powstał osobny kontroler możliwością pobrania danych. Ze względu na to, że hurtownie danych są tylko do odczyty jedynie jest możliwość pobierania danych poprzez GET np. Obraz zawierający tekst, zrzut ekranu, Czcionka

   Opis wygenerowany automatycznie
2. Migrations – folder zawierający wszystkie migracje jakie zostały wprowadzone przez aplikacje, w tym utworzenie wszystkich tabel.
3. Models – folder zawierający wszystkie modele danych jakie są używane w aplikacji

Swagger:  
Obraz zawierający tekst, zrzut ekranu, oprogramowanie, numer

Opis wygenerowany automatycznie  
Obraz zawierający tekst, zrzut ekranu, numer, Czcionka

Opis wygenerowany automatycznie

Aplikacja front-end struktura:  
  
Obraz zawierający tekst, zrzut ekranu, oprogramowanie, Oprogramowanie multimedialne

Opis wygenerowany automatycznie

Wygląd aplikacji front-end:

Strona z globalnym przychodem platformy NETFLIX:Obraz zawierający tekst, zrzut ekranu, Wykres, Czcionka

Opis wygenerowany automatycznie  
  
Strona z cenami pakietów netflix oraz ilości seriali I filmów: Obraz zawierający tekst, mapa, atlas

Opis wygenerowany automatycznie

Strona z tytułami na platformie NETFLIX:Obraz zawierający tekst, zrzut ekranu, numer, Czcionka

Opis wygenerowany automatycznie  
Strona z danymi giełdowymi dla podanego roku platformy NETFLIX:  
Obraz zawierający tekst, linia, Czcionka, Wykres

Opis wygenerowany automatycznie  
  
**W ten sposób omówiliśmy całą architekturę aplikacji.**W ramach naszego systemu, na tę chwilę gromadzimy dane dotyczące danych statystycznych platformy NETFLIX z myślą o powiększenie o inne platformy jak HBO, Spotify, Disney oraz przy użyciu kostek wielowymiarowych przeprowadzać porównanie poszczególnych platform w konkretnych dziedzinach oraz tworzenie z tego raportów.   
  
W dzisiejszym czasach wiedza o działaniu spółek oraz ich danych są nie do ocenienia dla działań na giełdzie.