1. **Co to jest? Do czego służy?**  
     
   Skrypty: SQL – służy do utworzenia bazy PWG i wszystkich tabel wraz z więzami spójności;  
   Python – służy do otworzenia, rozprasowania, stworzenia tabel i przekazania ich do bazy PWG.
2. **Jak działa?**  
     
   Skrypt SQLowy został użyty raz, do stworzenia tabel w SQL.  
   Skrypt Pythonowy – otwiera wszystkie pliki z rozszerzeniem JSON z zadanej lokalizacji *JSON\_DIR* następnie rozparsowuje je do postaci dataframe’ów odpowiadających tabelom z bazy PWG. W kolejnym kroku nawiązuje się połączenie z MS SQL SERVER, skąd pobierane są informacje za pomocą zapytań:  
   - czy NIPy spółek, których informacje przekazujemy nie znajdują się już w bazie (jeśli występuje taka sytuacja, skrypt Python przekaże stare dane do tabel o schemacie .*hist*, natomiast nowe do tabel *.dbo*)  
   - liczbie rekordów w każdej z tabel, żeby w przekazywanych tabelach tworzyć indeks

Następnie, rozparsowane dane przekazywane są do bazy SQL.

Użytkownik musi w pliku.py sprecyzować lokalizację plików do przekazania tj. wpisać ich lokalizację do zmiennej *JSON\_DIR* oraz wpisać nazwę bazy danych *database\_name* (z prawdopodobieństwem bliskim 1 będzie to PWG)

1. **Uwagi. Możliwości poprawy/rozwoju.**

Problemy, które mogą wystąpić podczas dodawania nowych JSONów do bazy:

- w pierwszym kroku, przy otwieraniu plików JSON – może zdarzyć się, że plik będzie pusty lub będzie posiadał błędną strukturę – pliki te zostaną pominięte

- w którymś z pól plików JSON może pojawić się np. znak tekstowy w miejscu, gdzie powinna znajdować zmienna liczbowa – wówczas ze względu na typy danych w sql niektóre, z tabel mogą nie przyjąć nowych rekordów

Wszystkie występujące problemy można podejrzeć w generowanym przez skrypt logu *json\_to\_sql\_log\_dzisiejsza\_data\_godzina.txt,* którypo wykonaniu skryptu pojawi się w lokalizacji przekazywanych JSONów*.* Wewnątrz znajdziemy takie informacje jak:

**All JSON files:** – *liczba wszystkich znalezionych plików JSON*

**Faulty JSON files:** - *nazwy pustych (wówczas po nazwie dopisek ‘- empty’) lub błędnych plików*

**Duplicated files:** - *nazwy plików znajdujących się już w bazie (tych, które zostaną przekazane do .hist)*

**Tables failed entirely to pass to SQL SERVER**: - *nazwy tabel, dla których nie wszystkie rekordy zostały* *przekazane*

**Rows failed to pass to SQL SERVER**: - *dane z rekordów, których nie da się przekazać np.:*

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | id | ilosc | kapital | krs | nazwa | nip | procent | regon | wartosc | czas | tabela | | 23579 | 2.33054e+06 | 2.83358e+08 | 61745 | FRITO | 5291347721ddd | 82.25 | 12688286 | 2.33054e+08 | 03.12.2018 | ZALEZNE | |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |

Jeśli po ‘:’ nic nie ma - jest to **bardzo dobra informacja** – wszystko przebiegło **bez komplikacji**.