

Wstęp

Jednym z głównych elementów naszej pracy jest przetwarzanie surowych danych do postaci, która przynosi wartość biznesową i jest czytelna dla odbiorców. Integrujemy się z wieloma systemami zewnętrznymi jak i aplikacjami wewnętrznymi, dlatego nie ograniczamy się do jednego formatu plików.

Opis danych

Plik "*statuses.json*" zawiera sztucznie wygenerowane informacje na temat obsługi klientów.

<i>kontakt_id</i>	- unikalny identyfikator zdarzenia
<i>klient_id</i>	- identyfikator klienta
<i>pracownik_id</i>	- identyfikator pracownika, który nadał status danemu klientowi
<i>status</i>	- status jaki został nadany dla naszego klienta
<i>kontakt_ts</i>	- timestamp kontaktu z klientem w formacie

Zadania

1. Za pomocą języka Java, wczytaj plik "*statuses.json*", wyklucz rekordy z datą sprzed lipca 2017, posortuj dane w dostarczonym pliku po kolumnach "*klient_id*", "*kontakt_ts*", a następnie zapisz wynik w formacie **.csv** o strukturze:
kontakt_id | klient_id | pracownik_id | status | kontakt_ts
2. Zaproponuj zapytanie SQL zwracające **ostatni** status każdego klienta, z którym były co najmniej 3 próby kontaktu. Potraktuj podesłany plik jako tabelę.
3. Zaproponuj zapytanie SQL, które zwróci dane do faktu *f_docieralnosc* pokazującego globalną docieralność do klientów. Ma być on pogrupowany po dacie. Jego struktura wygląda następująco: *data | sukcesy | utraty | do_ponowienia | zainteresowani_utraty | niezainteresowani_sukcesy*
Fakt zawiera 1 wymiar:
 - *data* (nie timestamp!)oraz miary:
 - *sukcesy* - liczba klientów, których ostatnim statusem jest "zainteresowany"
 - *utraty* - liczba klientów, których ostatnim statusem jest "niezainteresowany"
 - *do_ponowienia* - liczba klientów, których ostatni status to "poczta_głosowa" lub "nie_ma_w_domu"
 - **(Bonus)** *zainteresowani_utraty* - liczba klientów, których ostatnim statusem jest "niezainteresowany", a poprzednio wystąpił status "zainteresowany"
 - **(Bonus)** *niezainteresowani_sukcesy* - liczba klientów, których ostatnim statusem jest "zainteresowany", a poprzednio wystąpił status "niezainteresowany"

4. **(Bonus)** Rozwiąż zadanie 1. w Talend Open Studio for Data Integration.
Podpowiedź: użyj blocków tFileInputJSON, tFilterRow, tSortRow i tFileOutputDelimited. Program do pobrania z <https://www.talend.com/products/talend-open-studio>

Podsumowanie

Do każdego z powyższych przygotuj krótki komentarz/notatkę w której wyjaśnisz nam swoje podejście do problemu. Zwróć uwagę na czytelność swojego kodu - ułatwi nam to weryfikację. W razie wątpliwości - pisz śmiało.

Przygotowane przez Ciebie rozwiązania prześlij nam w postaci archiwum **.zip** lub jeśli posiadasz konto na **github/bitbucket**- prześlij w odpowiedzi link do repozytorium.

Jeśli nie poradzisz sobie ze wszystkimi zadaniami, nie rezygnuj i prześlij nam stan swojej pracy. Pozwoli nam to dopasować stopień stanowiska (Junior/Regular/Senior) do Twojej wiedzy i umiejętności.

POWODZENIA !