

# PO12: Przetwarzanie Big Data 22/23-Z

Rozpoczęto	środa, 2 listopada 2022, 12:04
Stan	Ukończone
Ukończono	środa, 2 listopada 2022, 12:06
Wykorzystany czas	2 min. 8 sek.
Punkty	1,00/1,00
Ocena	10,00 pkt. na 10,00 pkt. możliwych do uzyskania (100%)

Zad 36 :)

Masz już gotowy skrypt, który może być użyty do wsadowego przetwarzania danych.

Niestety ma on zapisane w środku katalogi określające miejsca źródłowe oraz docelowe dla naszych danych.

Dokonaj parametryzacji tego skryptu tak aby katalog źródłowy był podawany podczas uruchamiania jako parametr `input_dir`, a katalog wynikowy jako parametr `output_dir`.

Jako rozwiązanie wklej całą zawartość Twojego skryptu.

```
pig -f analysis_1.pig -param input_dir=input -param output_dir=output1

%declare trip_path $input_dir/trips
%declare weather_path $input_dir/weather.csv

trip = LOAD '$trip_path' USING PigStorage(',');
weather = LOAD '$weather_path' USING PigStorage(',');
trip = LOAD '$trip_path' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',','NO_MULTILINE','NOCHANGE','SKIP_INPUT_HEADER')
as (trip_id:int, starttime:chararray,
    stoptime:chararray, bikeid:chararray,
    tripduration:double, from_station_name:chararray,
    to_station_name:chararray, from_station_id:chararray,
    to_station_id:chararray, usertype:chararray,
    gender:chararray, birthyear:int);
first_trips = LIMIT trip 3;
weather = LOAD '$weather_path' USING
org.apache.pig.piggybank.storage.CSVExcelStorage(',','NO_MULTILINE','NOCHANGE','SKIP_INPUT_HEADER')
as (Date:chararray,
    Max_Temperature_F:double, Mean_Temperature_F:double,
    Min_TemperatureF:double, Max_Dew_Point_F:double,
    MeanDew_Point_F:double, Min_Dewpoint_F:double,
    Max_Humidity:double, Mean_Humidity:double,
    Min_Humidity:double, Max_Sea_Level_Pressure_In:double,
    Mean_Sea_Level_Pressure_In:double, Min_Sea_Level_Pressure_In:double,
    Max_Visibility_Miles:double, Mean_Visibility_Miles:double,
    Min_Visibility_Miles:double, Max_Wind_Speed_MPH:double,
    Mean_Wind_Speed_MPH:double, Max_Gust_Speed_MPH:chararray,
    Precipitation_In:double,Events:chararray);
projected_trip = FOREACH trip
    GENERATE REGEX_EXTRACT(starttime, '(.*) (.)', 1) as startday,
    tripduration, gender;
projected_weather = FOREACH weather
    GENERATE Date, Events,
    SUBSTRING(Date, (int)SIZE(Date)-4, (int)SIZE(Date)) as year;
trip_with_weather = JOIN projected_trip BY startday, projected_weather BY Date;
trip_with_weather_simple = FOREACH trip_with_weather GENERATE
startday as startday,
    tripduration as tripduration,
    gender as gender,
    Events as events,
    year as year;
trip_with_weather_grouped = GROUP trip_with_weather_simple BY (gender, events, year);
final_result = FOREACH trip_with_weather_grouped {
    unique_days = DISTINCT trip_with_weather_simple.startday;
    GENERATE group,
    COUNT(trip_with_weather_simple)/COUNT(unique_days),
    SUM(trip_with_weather_simple.tripduration)/COUNT(unique_days); };
STORE final_result INTO '$output_dir' USING JsonStorage();
```

Current word count: 160

Ogólna informacja zwrotna

Tvoja odpowiedź jest poprawna.

Komentarz:

Pytanie 1

Poprawnie

Punkty: 1,00 z 1,00

Oflaguj pytanie

Nawigacja w teście

Zakończ przegląd

Zakończ przegląd

Jesteś zalogowany(a) jako [Andrzej Kapczyński](#) (Wyloguj)

[WliIT-Inf-st-I-sem7-PrzBigDat-22/23-Z](#)