

Szybkie tramwaje

i dane po warszawsku



Cel

- zdobyć informacje o średnich prędkościach na różnych liniach
- dowiedzieć się, gdzie tramwaje są naprawdę szybkie



Użyte zbiory danych

- <https://api.um.warszawa.pl>
- JSON

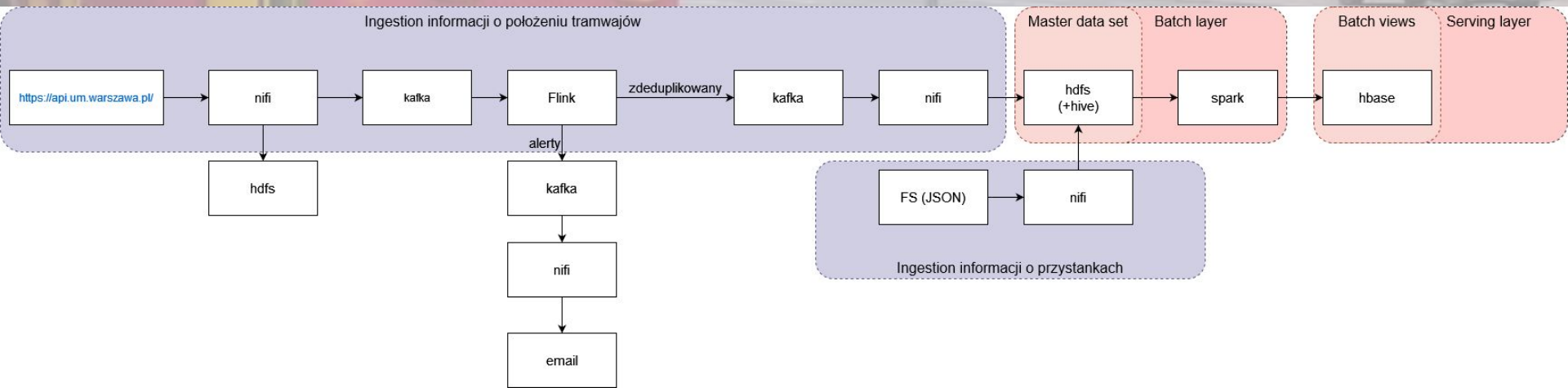
Trams:

```
{"Lines": "15", "Lon": 20.980919, "VehicleNumber": "1167+1166", "Time": "2022-01-19 21:50:24", "Lat": 52.215996, "Brigade": "11"}
```

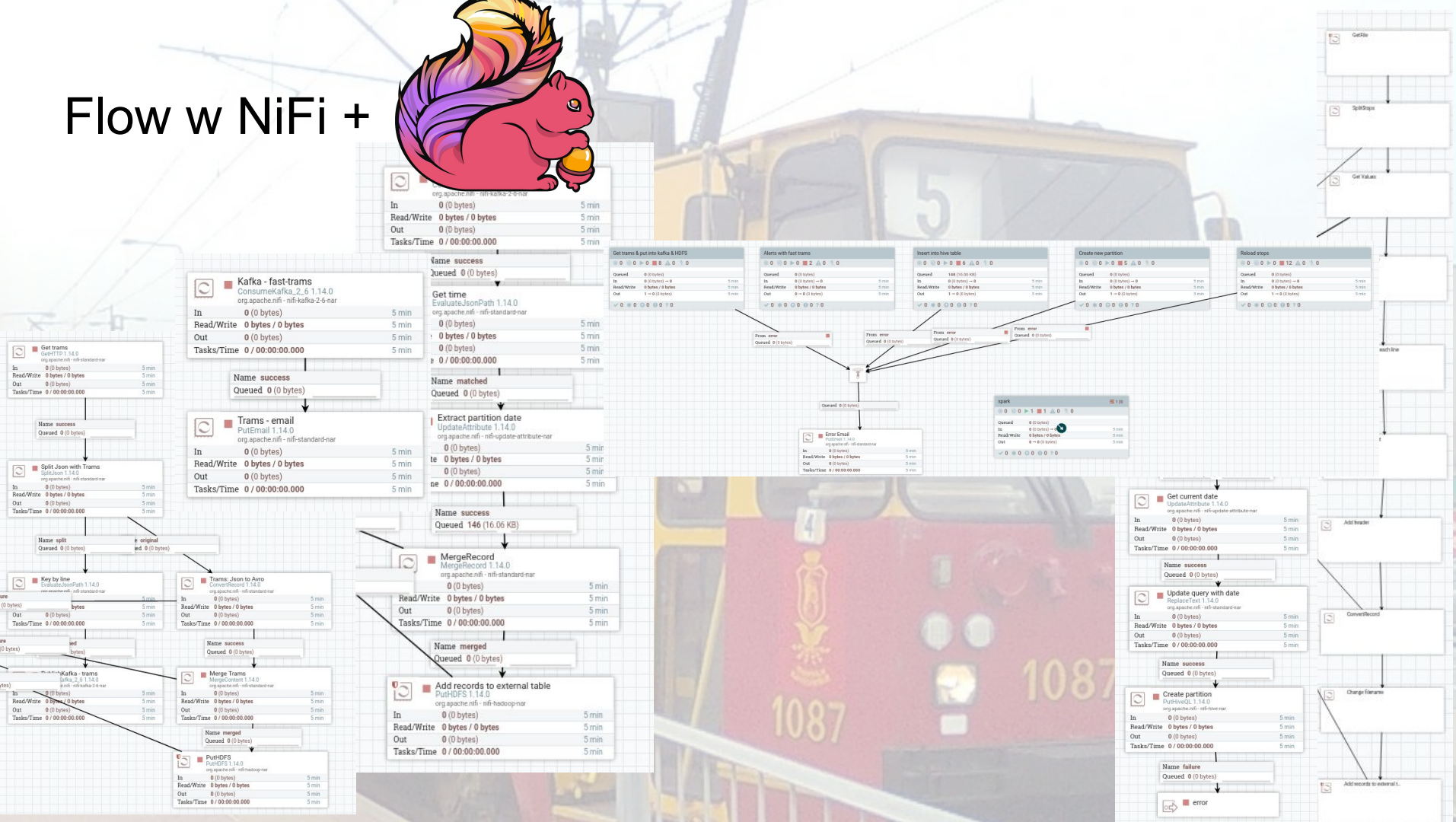
Stops:

```
{"values": [{"value": "1001", "key": "zespol"}, {"value": "01", "key": "slupek"}, {"value": "Kijowska", "key": "nazwa_zespołu"}, {"value": "2201", "key": "id_ulicy"}, {"value": "52.248455", "key": "szer_geo"}, {"value": "21.044827", "key": "dlug_geo"}, {"value": "al.Zieleniecka", "key": "kierunek"}, {"value": "2021-09-01 00:00:00.0", "key": "obowiazuje_od"}]}
```


architektura



Flow w NiFi +



MASTER DATASET

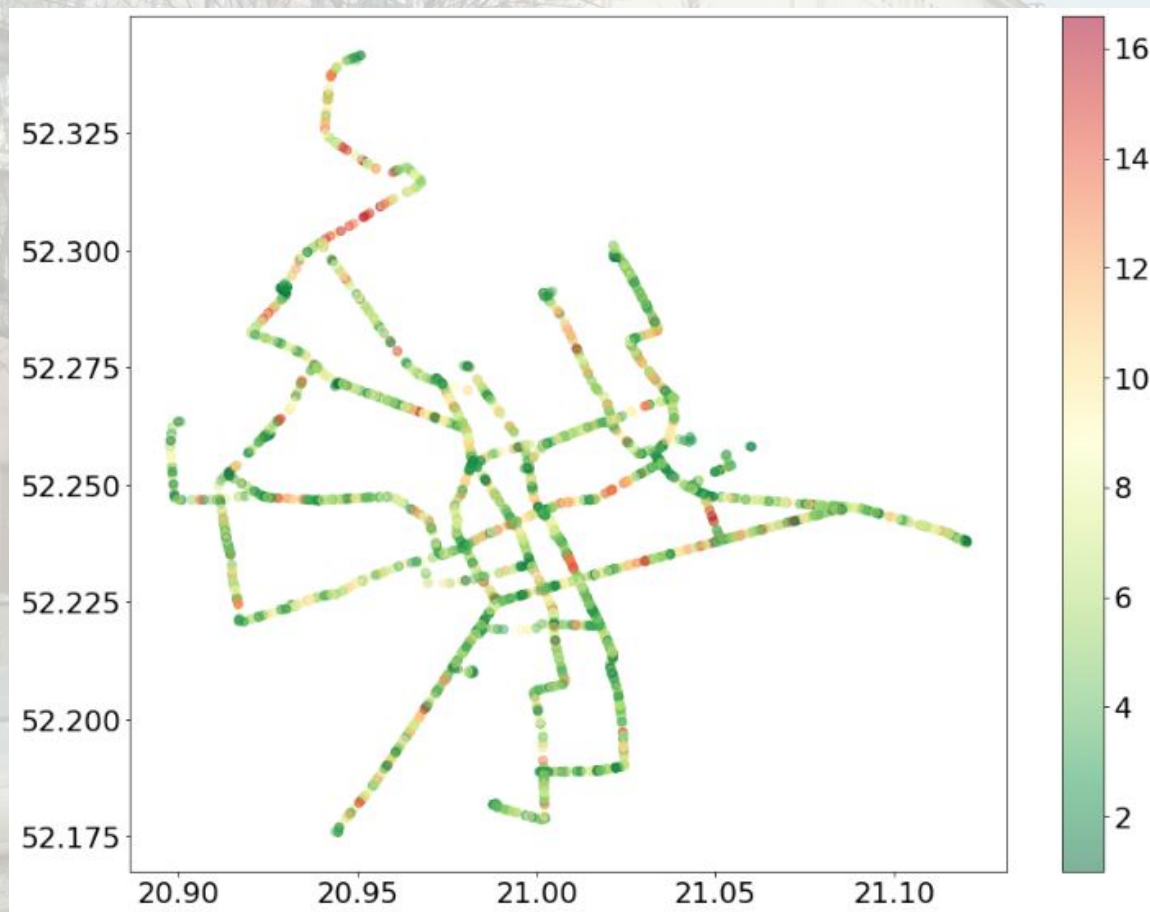
0: jdbc:hive2://localhost:10000> select * from trams limit 5;

trams.lines	trams.lon	trams.lat	trams.vehiclenumber	trams.brigade	trams.time	trams.day
4	21.034567	52.254936	1123+1122	015	1642031501000	2022-01-12
22	20.934433	52.298763	1281+1282	6	1642031501000	2022-01-12
24	21.063822	52.246525	1295+1296	017	1642031497000	2022-01-12
35	20.92479	52.26023	1313+1314	12	1642031497000	2022-01-12
15	20.956976	52.188877	1405+1406	2	1642031501000	2022-01-12

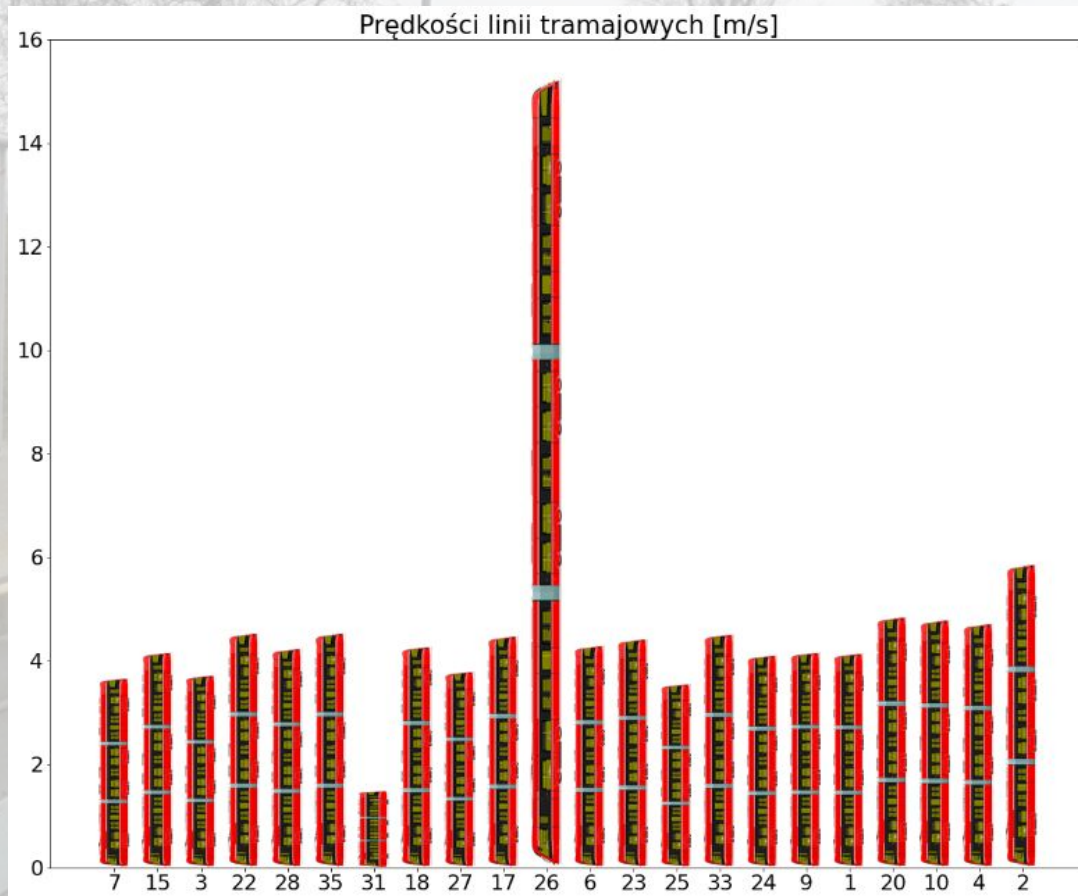
0: jdbc:hive2://localhost:10000> select * from stops limit 5;

stops.zespol	stops.slupek	stops.nazwa_zespolu	stops.id_ulicy	stops.szer_geo	stops.dlug_geo	stops.kierunek	stops.obowiazuje_od
1001	1	Kijowska	2201	52.24845504760742	21.04482650756836	al.Zieleniecka	2021-09-01 00:00:00.0
1001	2	Kijowska	2201	52.24907684326172	21.044443130493164	Z?bkowska	2021-09-01 00:00:00.0
1001	3	Kijowska	2201	52.24892807006836	21.04416847229004	al.Zieleniecka	2021-09-01 00:00:00.0
1001	4	Kijowska	2201	52.249969482421875	21.041587829589844	Z?bkowska	2021-09-01 00:00:00.0
1001	5	Kijowska	1203	52.25032043457031	21.043861389160156	al.Zieleniecka	2021-09-01 00:00:00.0

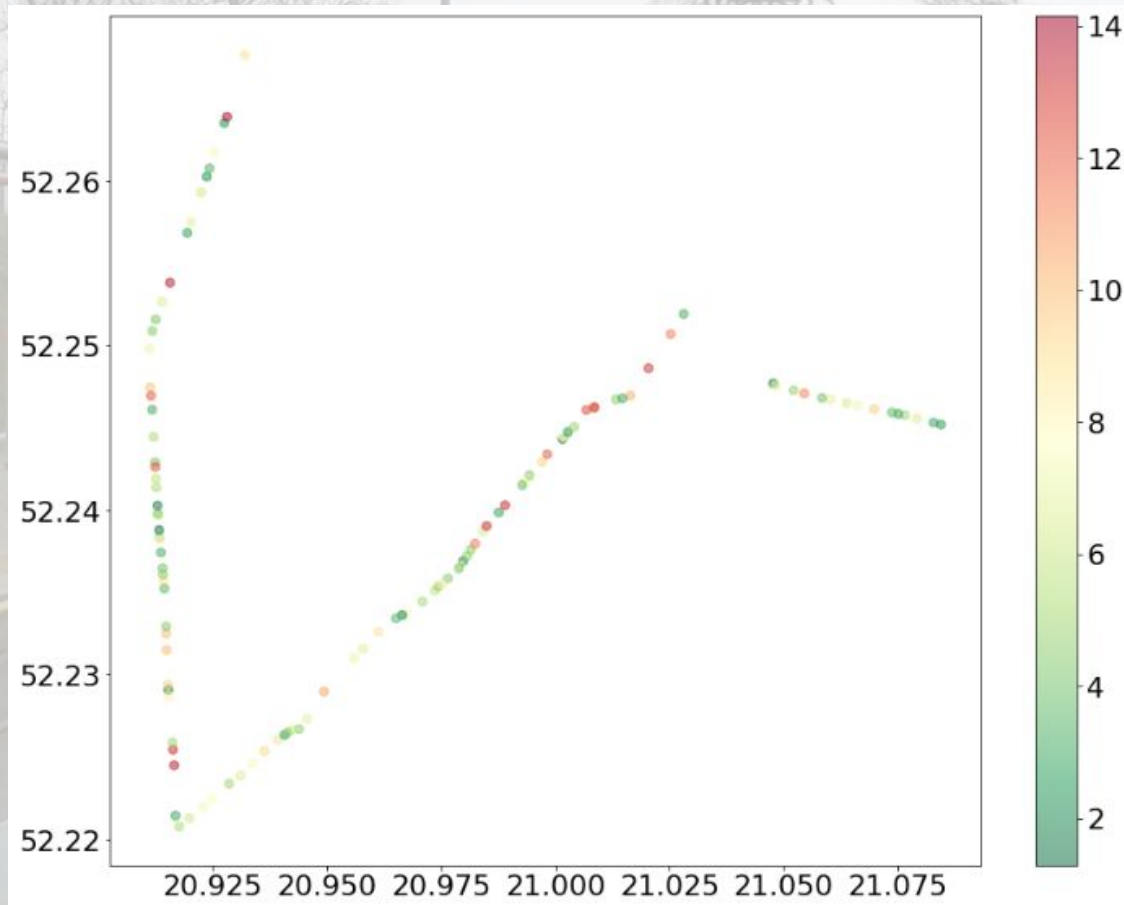
Wizualizacja #1



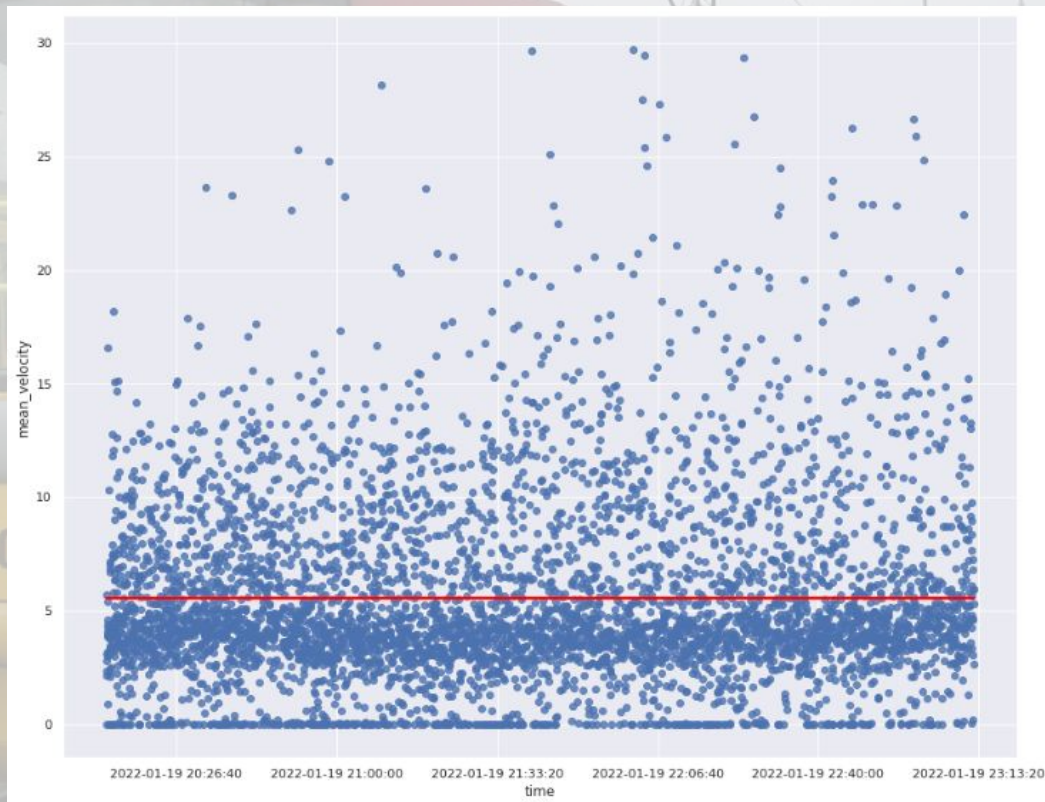
Wizualizacja #2



Wizualizacja #2

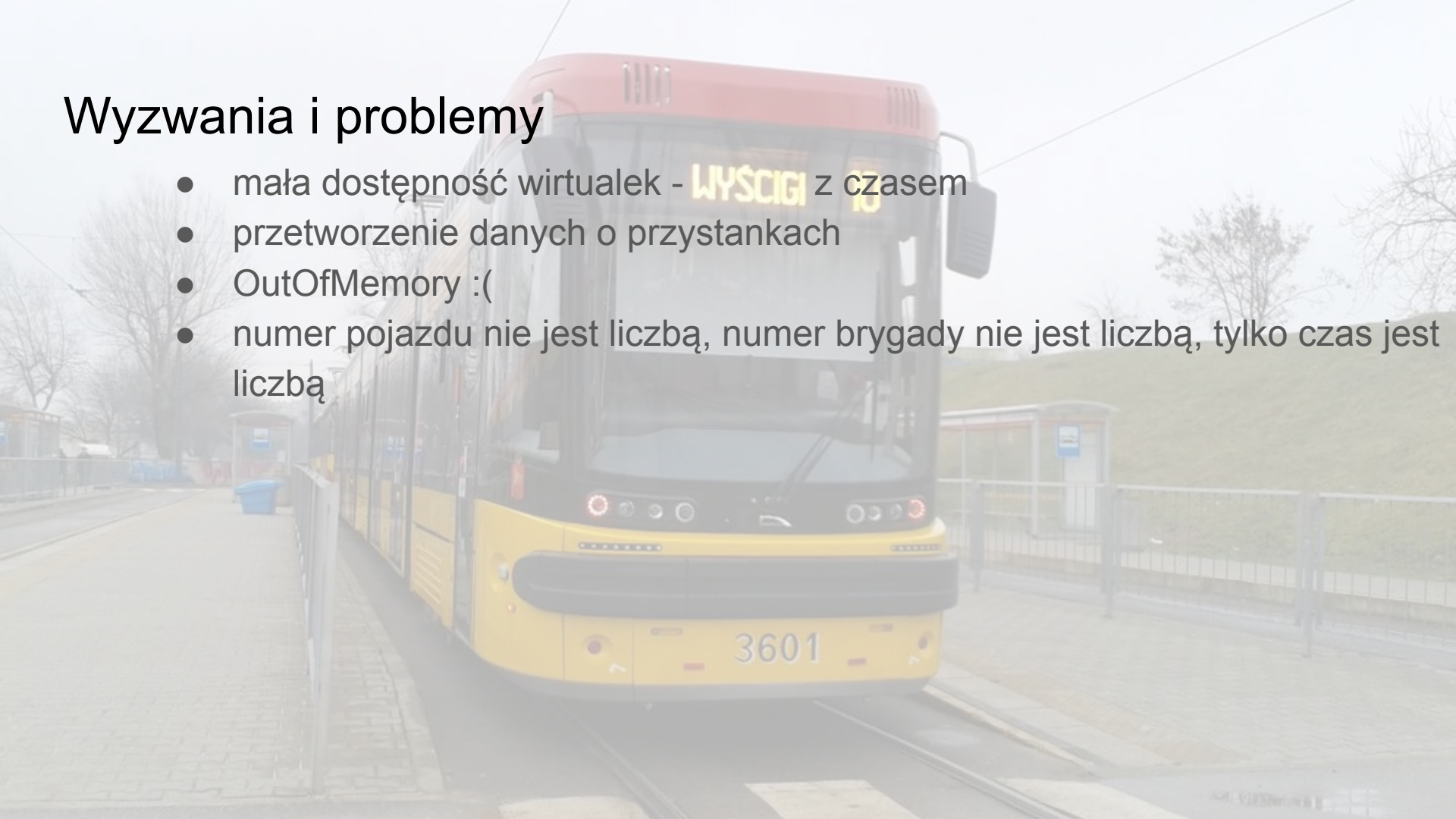


Wizualizacja #3



Wyzwania i problemy

- mała dostępność wirtualiek - WYSCIGI z czasem
- przetworzenie danych o przystankach
- OutOfMemory :(
- numer pojazdu nie jest liczbą, numer brygady nie jest liczbą, tylko czas jest liczbą



¿Query?

