



Predicting public transport delays

Transatlantic Scooters' BD Project

Laura Bąkała
Wojciech Kretowicz
Karol Pysiak
Mateusz Szysz



Goals

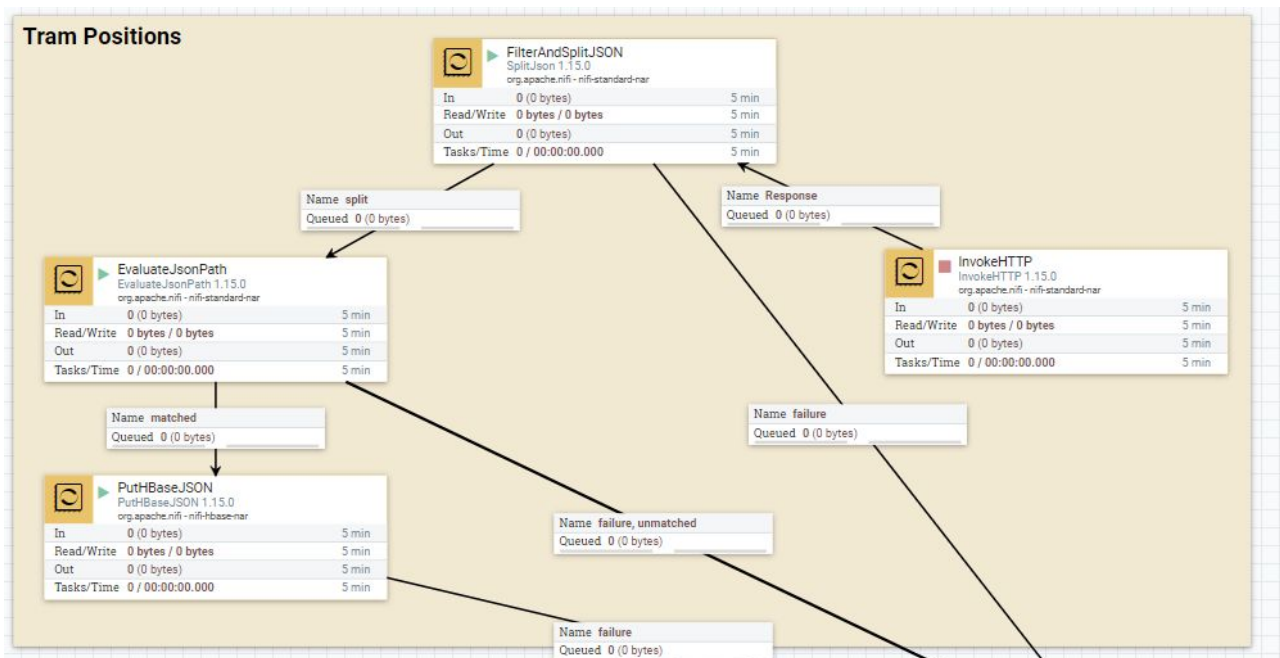
- Predicting public transport delays based on weather prognoses, traffic and past delays.
- Online monitoring of delays in Warsaw.
- Facilitating choosing appropriate means of transport at given moment and planning travels.



Data sources

- Warszawskie Dane (on-line data about trams and buses including their positions in given moment of time)
- Timetable for buses and trams in Warsaw
- Open-source meteorological data source with prognoses for specific amount of days
- ...*hopefully nothing else...*

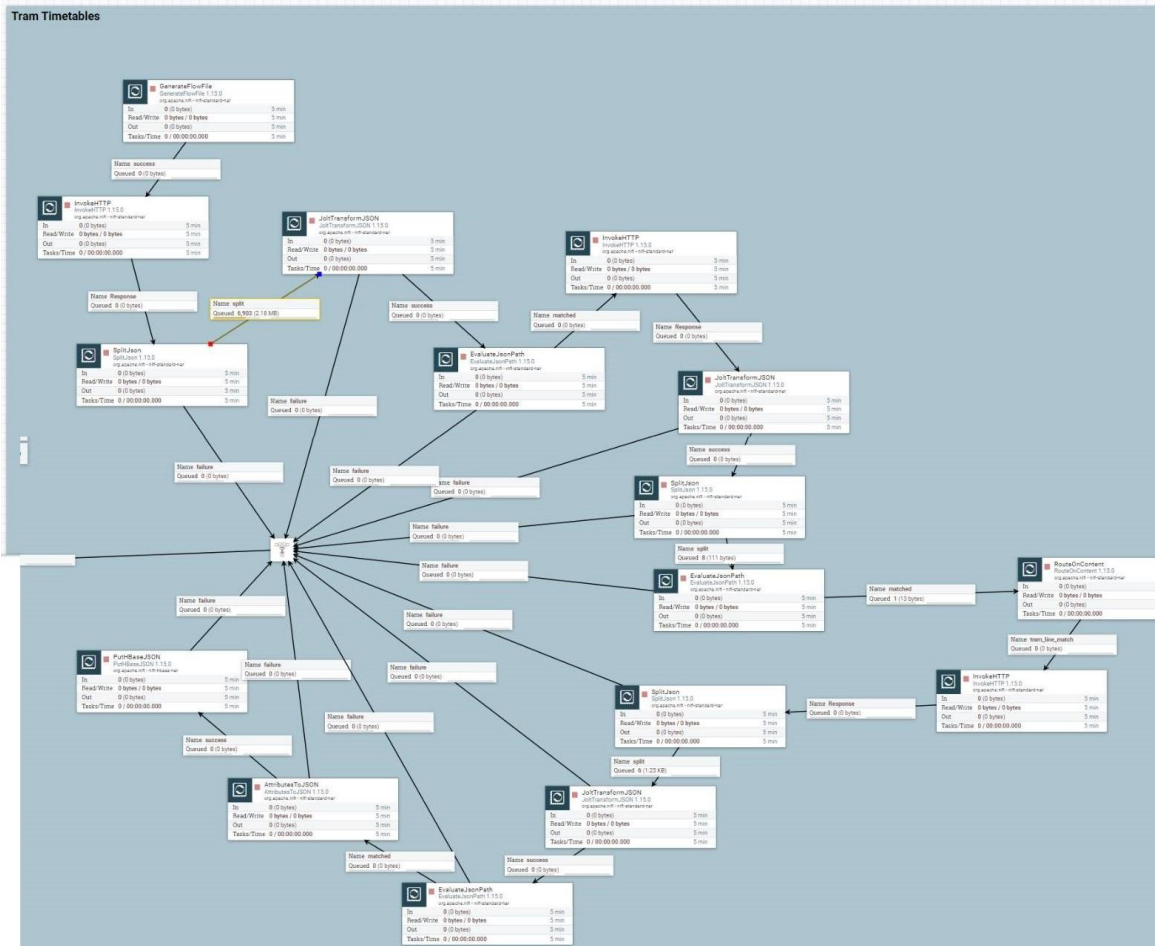
Warszawskie Dane





Warszawskie Dane: trams' locations

1. Retrieve data about trams' positions
2. Filter out positions (roughly) outside the boundaries of Warsaw
3. Save data to HBase





Warszawskie Dane: schedules

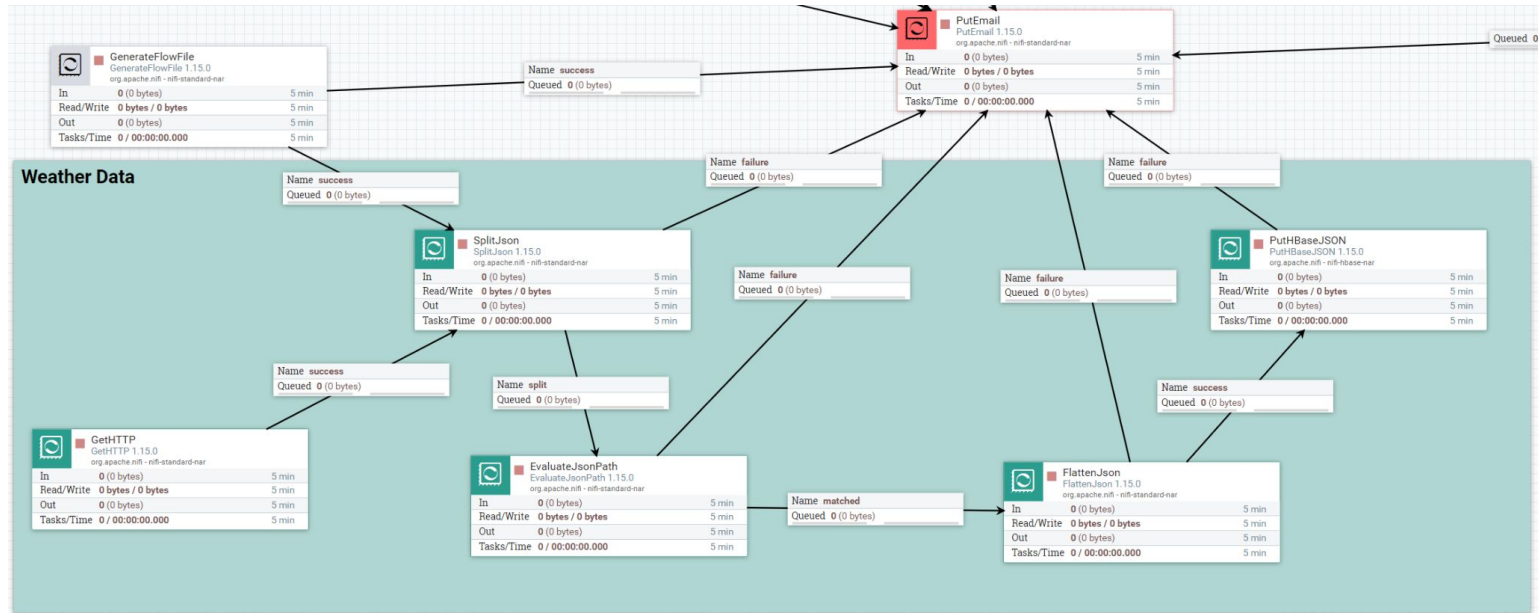
1. Retrieve the list of bus/tram stops
2. For each stop:
 - a. Retrieve the list of lines stopping there
 - b. Filter tram lines (1- & 2-digit codes)
 - c. For each tram line:
 - i. Retrieve timetable
 - ii. Split timetable into separate departure times
 - iii. Add relevant attributes from previous queries
 - iv. Save to data HBase

OpenWeather

```
1 {
2   "cod": "200",
3   "message": 0,
4   "cnt": 40,
5   "list": [
6     {
7       "dt": 1637506800,
8       "main": {
9         "temp": 281.4,
10        "feels_like": 278.73,
11        "temp_min": 281.4,
12        "temp_max": 281.84,
13        "pressure": 1004,
14        "sea_level": 1004,
15        "grnd_level": 993,
16        "humidity": 85,
17        "temp_kf": -0.44
18      },
19      "weather": [
20        {
21          "id": 803,
22          "main": "Clouds",
23          "description": "broken clouds",
24          "icon": "04n"
25        }
26      ],
27      "clouds": {
28        "all": 75
29      },
30      "wind": {
31        "speed": 4.59,
32        "deg": 270,
33        "gust": 8.64
34      },
35      "visibility": 10000,
36      "pop": 0,
37      "sys": {
38        "type": 1,
39        "id": 5281,
40        "country": "DE",
41        "sunrise": 1637500800,
42        "sunset": 1637500800
43      },
44      "dt_txt": "2021-11-21 15:00:00"
45    },
46    {
47      "dt": 1637517600,
48      "main": {
49        "temp": 281.4,
50        "feels_like": 278.73,
51        "temp_min": 281.4,
52        "temp_max": 281.84,
53        "pressure": 1004,
54        "sea_level": 1004,
55        "grnd_level": 993,
56        "humidity": 85,
57        "temp_kf": -0.44
58      },
59      "weather": [
60        {
61          "id": 803,
62          "main": "Clouds",
63          "description": "broken clouds",
64          "icon": "04n"
65        }
66      ],
67      "clouds": {
68        "all": 75
69      },
70      "wind": {
71        "speed": 4.59,
72        "deg": 270,
73        "gust": 8.64
74      },
75      "visibility": 10000,
76      "pop": 0,
77      "sys": {
78        "type": 1,
79        "id": 5281,
80        "country": "DE",
81        "sunrise": 1637500800,
82        "sunset": 1637500800
83      },
84      "dt_txt": "2021-11-21 18:00:00"
85    },
86    {
87      "dt": 1637528400,
88      "main": {
89        "temp": 281.4,
90        "feels_like": 278.73,
91        "temp_min": 281.4,
92        "temp_max": 281.84,
93        "pressure": 1004,
94        "sea_level": 1004,
95        "grnd_level": 993,
96        "humidity": 85,
97        "temp_kf": -0.44
98      },
99      "weather": [
100       {
101         "id": 803,
102         "main": "Clouds",
103         "description": "broken clouds",
104         "icon": "04n"
105       }
106     ],
107     "clouds": {
108       "all": 75
109     },
110     "wind": {
111       "speed": 4.59,
112       "deg": 270,
113       "gust": 8.64
114     },
115     "visibility": 10000,
116     "pop": 0,
117     "sys": {
118       "type": 1,
119       "id": 5281,
120       "country": "DE",
121       "sunrise": 1637500800,
122       "sunset": 1637500800
123     },
124     "dt_txt": "2021-11-21 21:00:00"
125   }
126 ]
127 }
```

```
6 {
7   "dt": 1637506800,
8   "main": {
9     "temp": 281.4,
10    "feels_like": 278.73,
11    "temp_min": 281.4,
12    "temp_max": 281.84,
13    "pressure": 1004,
14    "sea_level": 1004,
15    "grnd_level": 993,
16    "humidity": 85,
17    "temp_kf": -0.44
18  },
19  "weather": [
20    {
21      "id": 803,
22      "main": "Clouds",
23      "description": "broken clouds",
24      "icon": "04n"
25    }
26  ],
27  "clouds": {
28    "all": 75
29  },
30  "wind": {
31    "speed": 4.59,
32    "deg": 270,
33    "gust": 8.64
34  },
35  "visibility": 10000,
36  "pop": 0,
37  "sys": {
38    "pod": "n"
39  },
40  "dt_txt": "2021-11-21 15:00:00"
41 }
```


OpenWeather





OpenWeather

1	1638435600	column=weather_data:clouds.all, timestamp=1638019987294, value=100
2	1638435600	column=weather_data:dt, timestamp=1638019987294, value=1638435600
3	1638435600	column=weather_data:dt_txt, timestamp=1638019987294, value=2021-12-02 09:00:00
4	1638435600	column=weather_data:main.feels_like, timestamp=1638019987294, value=273.01
5	1638435600	column=weather_data:main.grnd_level, timestamp=1638019987294, value=980
6	1638435600	column=weather_data:main.humidity, timestamp=1638019987294, value=81
7	1638435600	column=weather_data:main.pressure, timestamp=1638019987294, value=994
8	1638435600	column=weather_data:main.sea_level, timestamp=1638019987294, value=994
9	1638435600	column=weather_data:main.temp, timestamp=1638019987294, value=277.7
10	1638435600	column=weather_data:main.temp_kf, timestamp=1638019987294, value=0
11	1638435600	column=weather_data:main.temp_max, timestamp=1638019987294, value=277.7
12	1638435600	column=weather_data:main.temp_min, timestamp=1638019987294, value=277.7
13	1638435600	column=weather_data:pop, timestamp=1638019987294, value=0.36
14	1638435600	column=weather_data:rain.3h, timestamp=1638019987294, value=0.42
15	1638435600	column=weather_data:sys.pod, timestamp=1638019987294, value=d
16	1638435600	column=weather_data:visibility, timestamp=1638019987294, value=10000
17	1638435600	column=weather_data:weather[0].description, timestamp=1638019987294, value=light rain
18	1638435600	column=weather_data:weather[0].icon, timestamp=1638019987294, value=10d
19	1638435600	column=weather_data:weather[0].id, timestamp=1638019987294, value=500
20	1638435600	column=weather_data:weather[0].main, timestamp=1638019987294, value=Rain
21	1638435600	column=weather_data:wind.deg, timestamp=1638019987294, value=266
22	1638435600	column=weather_data:wind.gust, timestamp=1638019987294, value=14.55
23	1638435600	column=weather_data:wind.speed, timestamp=1638019987294, value=7.21



Analysis - modelling

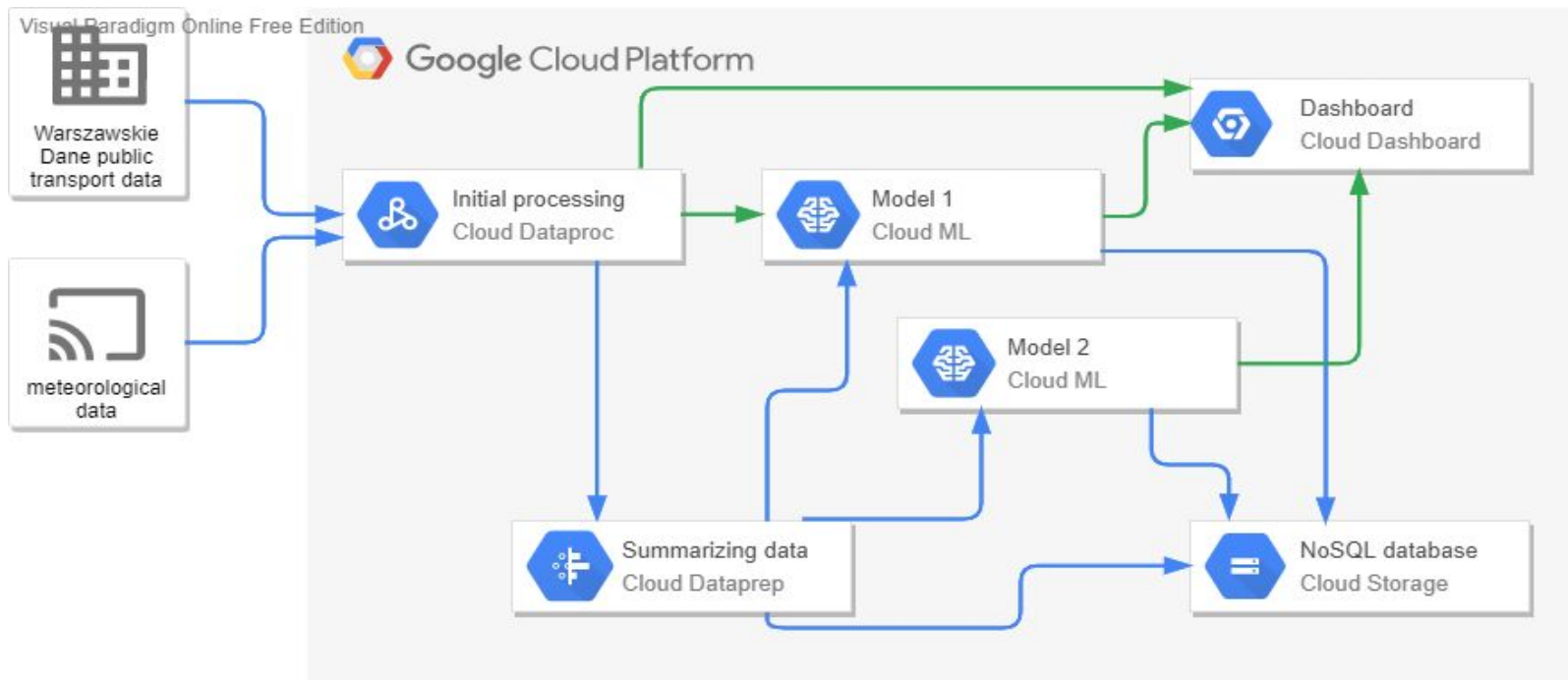
- Goal: predicting delays in the nearest future
- Model will be updated constantly with incoming data
- Model will utilize all data sources
- We plan to use following variables:
 - time of day
 - line number
 - current vehicle position
 - last delay
 - next stop
 - direction of the line
 - temperature
 - pressure
 - humidity
 - cloudiness
 - wind
 - visibility
 - precipitation




Dashboard

- Offline:
 - average delays against time of day / week day / month
 - average delays against the particular line
 - heatmap of Warsaw with average delays
- Online:
 - current / future average delay
 - current / future delay for each line
 - current / future heatmap of Warsaw with information about delays

Architecture





Thank you for your attention

Laura Bąkała
Wojciech Kretowicz
Karol Pysiak
Mateusz Szysz