

Analiza prędkości pojazdów transportu publicznego w Warszawie

w zależności od warunków pogodowych

Zespół KPK:
Paulina Jaszczuk, Kacper Grzymkowski, Jakub Fołtyn

Potencjalne korzyści:

- kwestie bezpieczeństwa
 - analiza prędkości pojazdu w zależności od warunków drogowych
- zarządzanie transportem publicznym
 - czasu w trasie i dystansu pokonywanego przez pojazdy poszczególnych linii

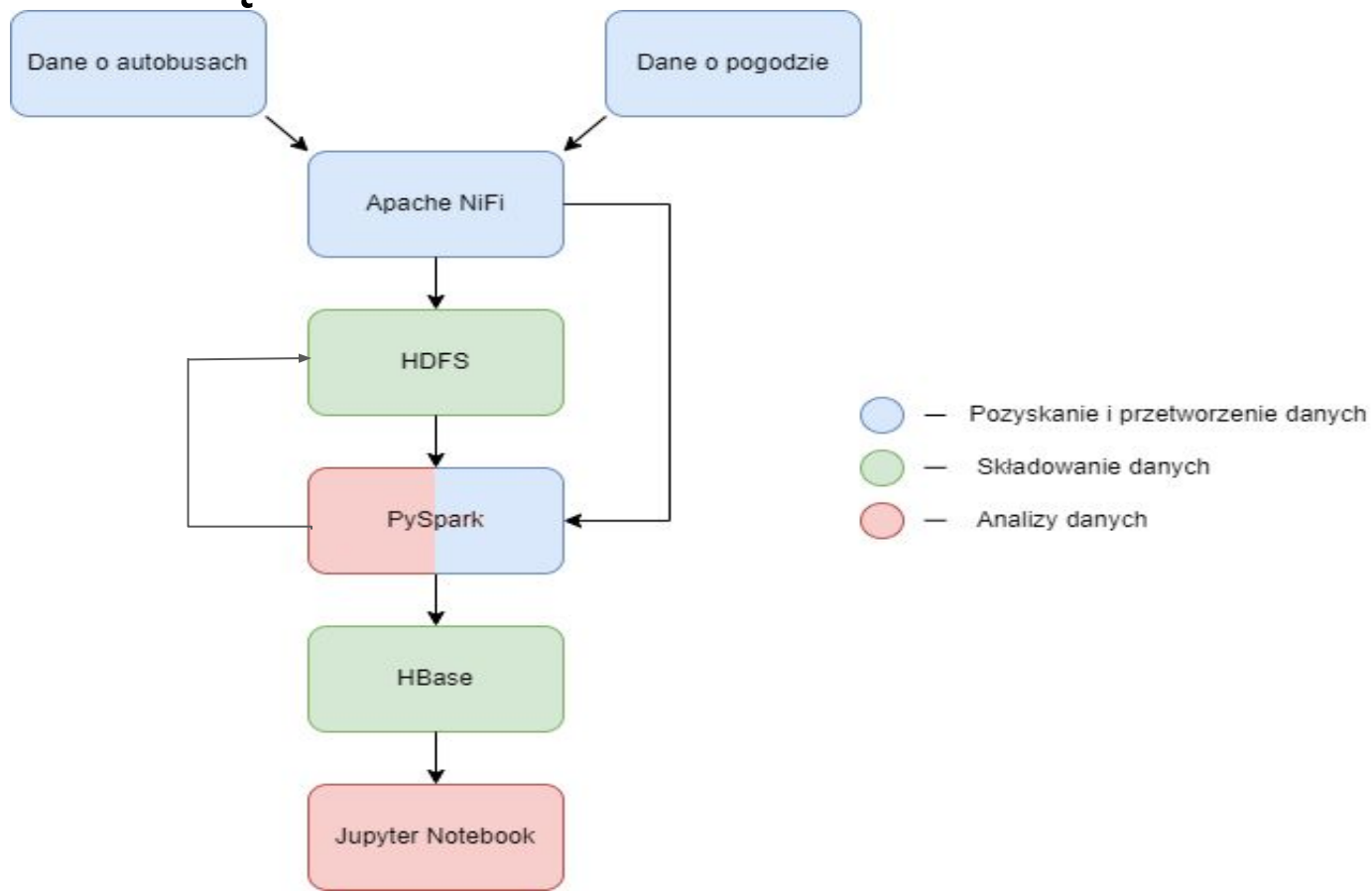
Opis danych o autobusach

- Dane pochodzą z API [Dane po warszawsku](#)
- Zdecydowaliśmy się na dane o autobusach
- Informacje takie jak: długość szerokość geograficzna, czas sygnału GPS, numer linii i brygady
- Dane pobieramy bezpośrednio z API (JSON) oraz z pliku CSV z danymi za lipiec

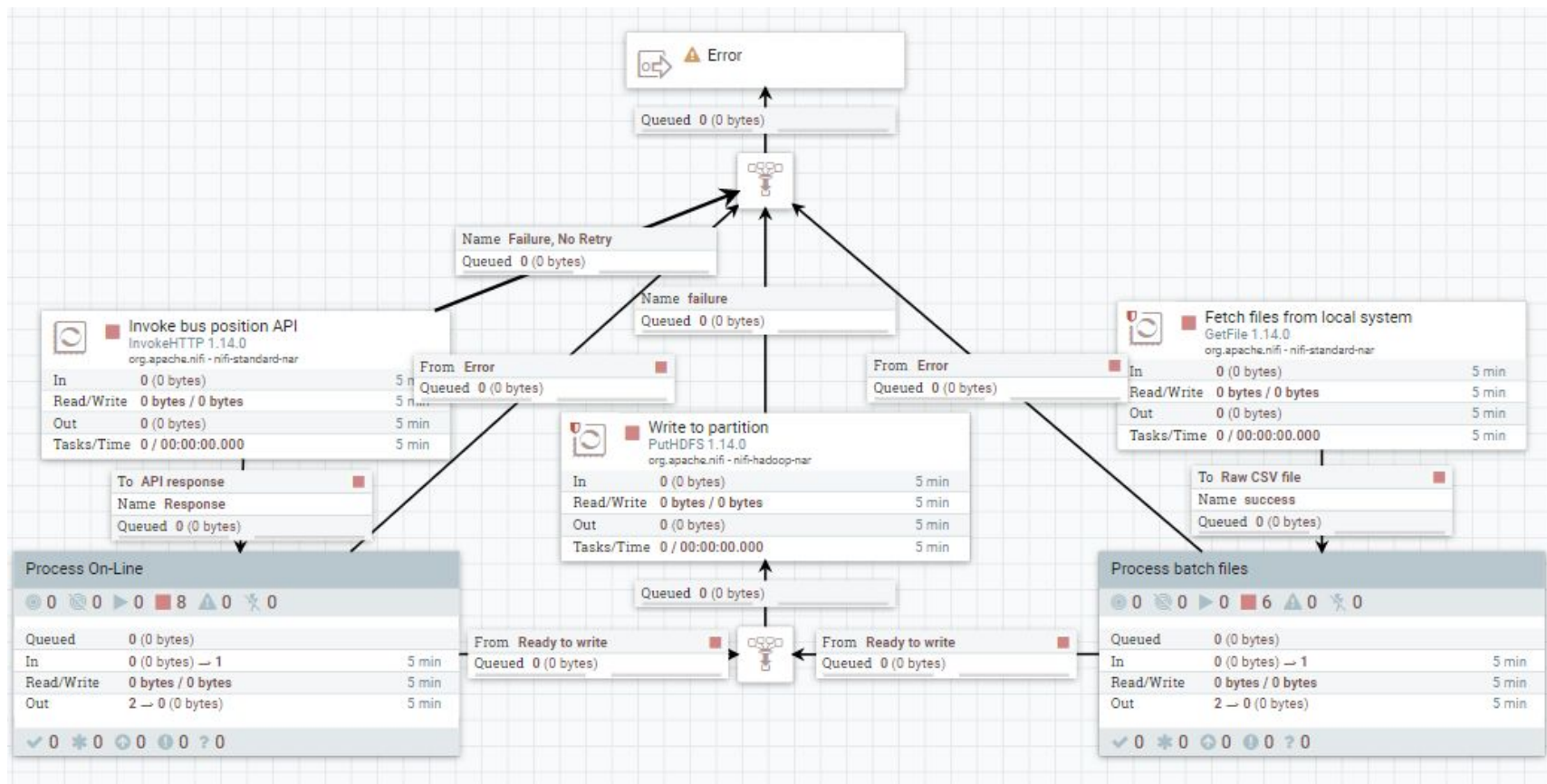
Opis danych o pogodzie

- Dane pochodzą z serwisu [Meteostat](#)
- Dane ze stacji Warszawa- Okęcie
- Dane zbierane co godzinę
- Informacje takie jak: data i godzina pomiaru, opady, wilgotność, prędkość i kierunek wiatru, ogólna jakość pogody (kod COCO)
- Ładowanie pliku z API (pełna historia od 1928) oraz “napływ danych” przy pomocy skryptu i API Python (ostatni tydzień)

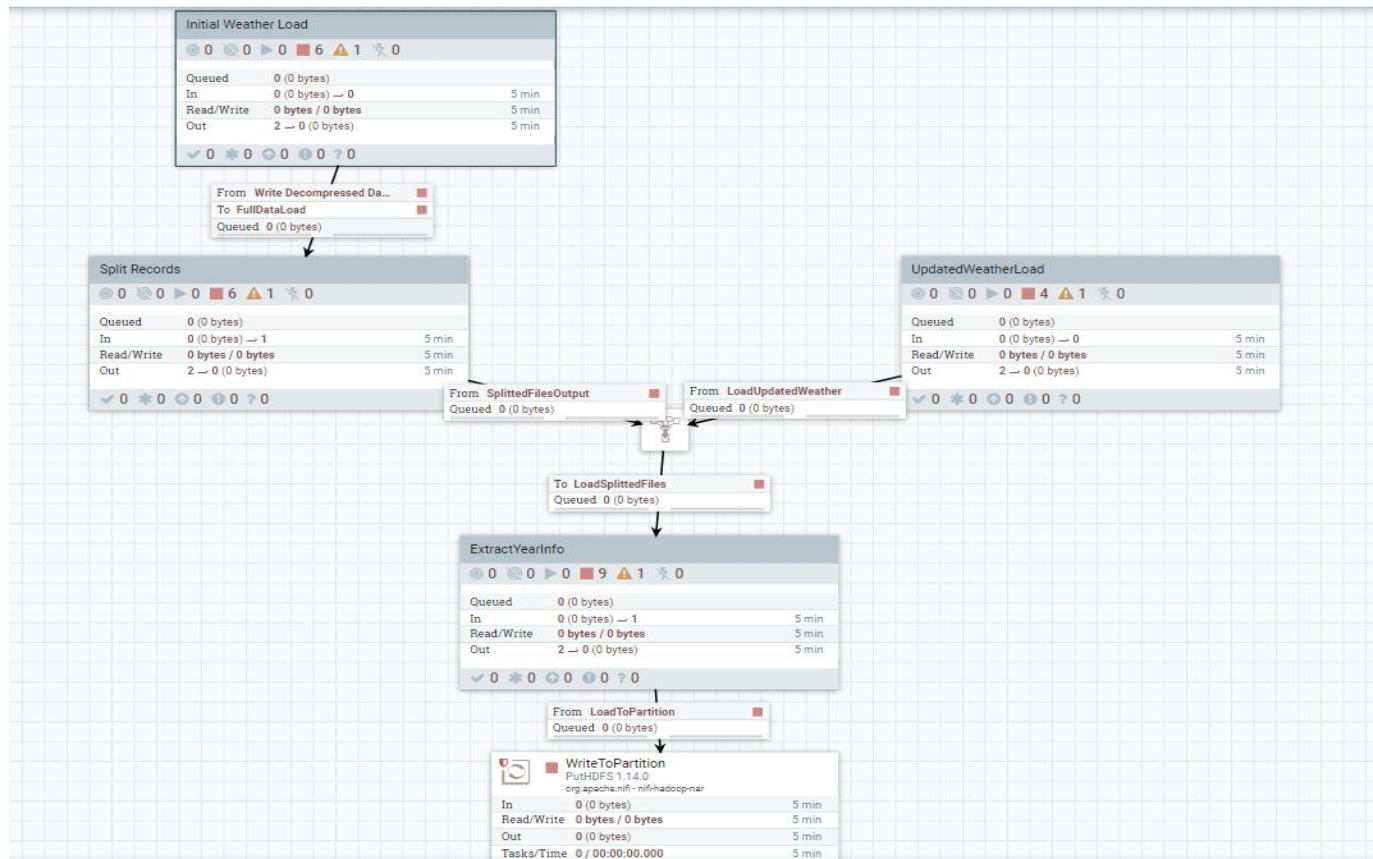
Schemat rozwiązania



Przetwarzanie w Nifi - dane o autobusach



Przetwarzanie w Nifi- dane o pogodzie



Składowanie danych

- **Dane pogodowe** - foldery na HDFS, przechowujące dane z odpowiedniego roku (2020, 2021 itd), składowane jako CSV
- **Dane o autobusach** - foldery na HDFS, zbierane przez pewien czas, i dopiero zapisywane, składowane jako AVRO, partycje po dacie

Wzbogacenie danych

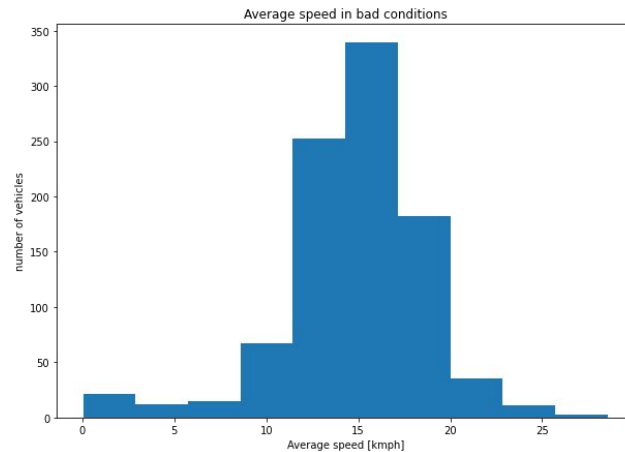
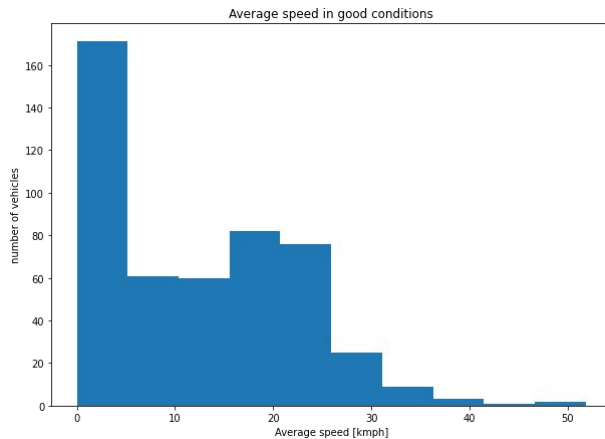
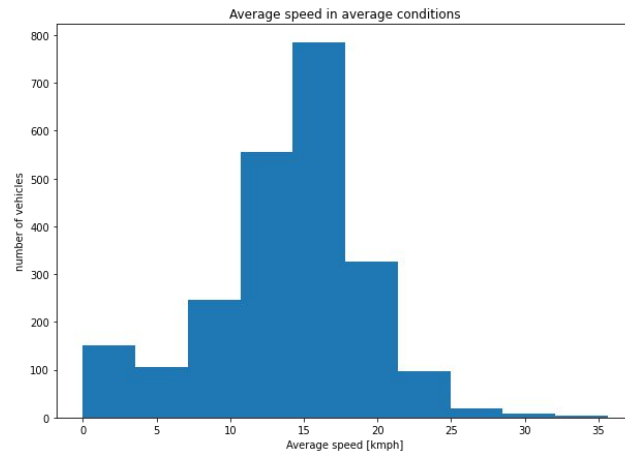
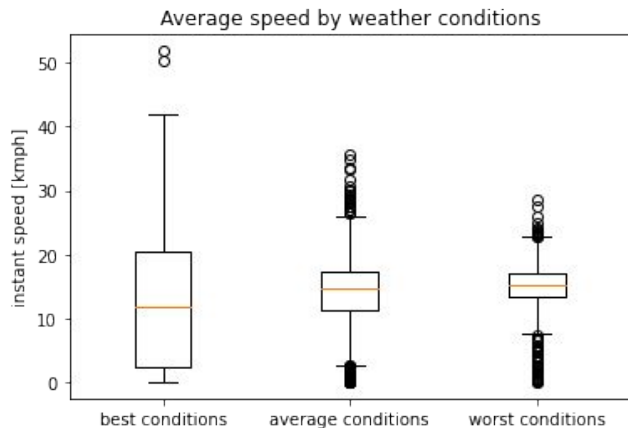
- Dane o autobusach zostały ubogacone przy użyciu PySpark
- Inferencja poprzednich obserwacji dla danego pojazdu
- Dodanie kolumn: prędkość chwilowa, dystans przebyty od ostatniego pomiaru, czas od ostatniego pomiaru

Analiza w Pyspark

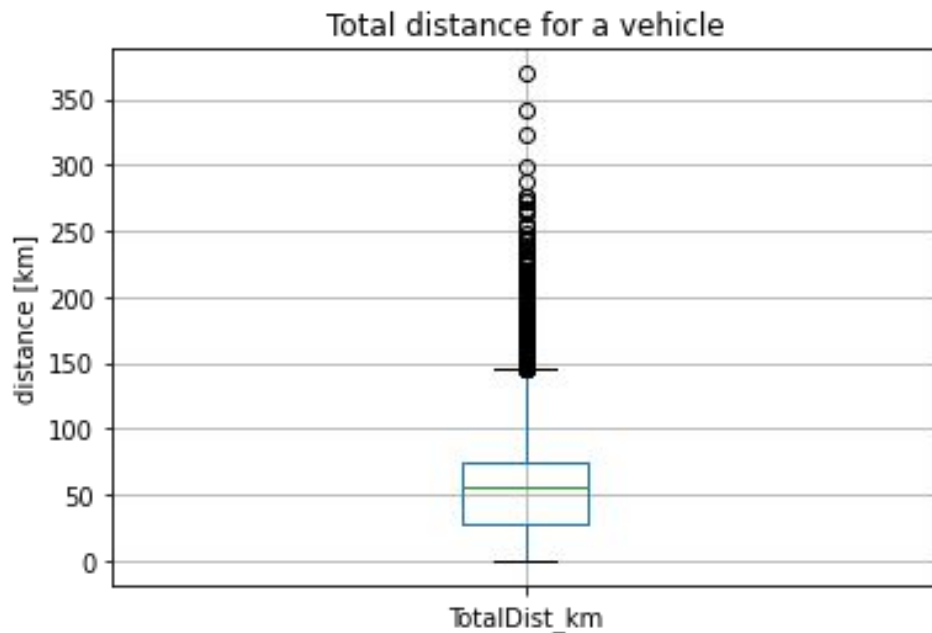
Przeprowadzono następujące analizy:

- Łączny przejechany dystans, czas w trasie oraz średnie prędkości dla pojazdów.
- Łączny przejechany dystans, czas w trasie oraz średnie prędkości dla linii.
- Łączny przejechany dystans, czas w trasie oraz średnie prędkości wszystkich pojazdów w zależności od kondycji pogodowej
- Łączny przejechany dystans, czas w trasie oraz średnie prędkości linii w zależności od ilości opadów
- Łączny przejechany dystans, czas w trasie oraz średnie prędkości pojazdów w zależności od prędkości wiatru.

Analizy końcowe - wpływ warunków pogodowych na prędkość



Analizy końcowe - analiza dystansu pokonanego przez pojazdy



| | Lines | mean |
|-----|-------|------------|
| 169 | 733 | 188.055087 |
| 176 | 743 | 164.255865 |
| 171 | 736 | 143.600393 |
| 183 | L-3 | 142.989906 |
| 119 | 264 | 136.499688 |
| 210 | L40 | 131.712477 |
| 167 | 730 | 121.156442 |
| 158 | 719 | 119.527083 |
| 181 | L-1 | 115.605433 |
| 152 | 712 | 112.140171 |

Podsumowanie projektu

- Udało się spełnić praktycznie wszystkie pierwotne założenia projektowe.
- Z punktu widzenia biznesowego wyniki mogą okazać się pomocne w podejmowaniu decyzji zarządczych dot. transportu miejskiego.
- Analizy można rozszerzać o kolejne statystyki oraz o dane np. z innych okresów i pór roku.