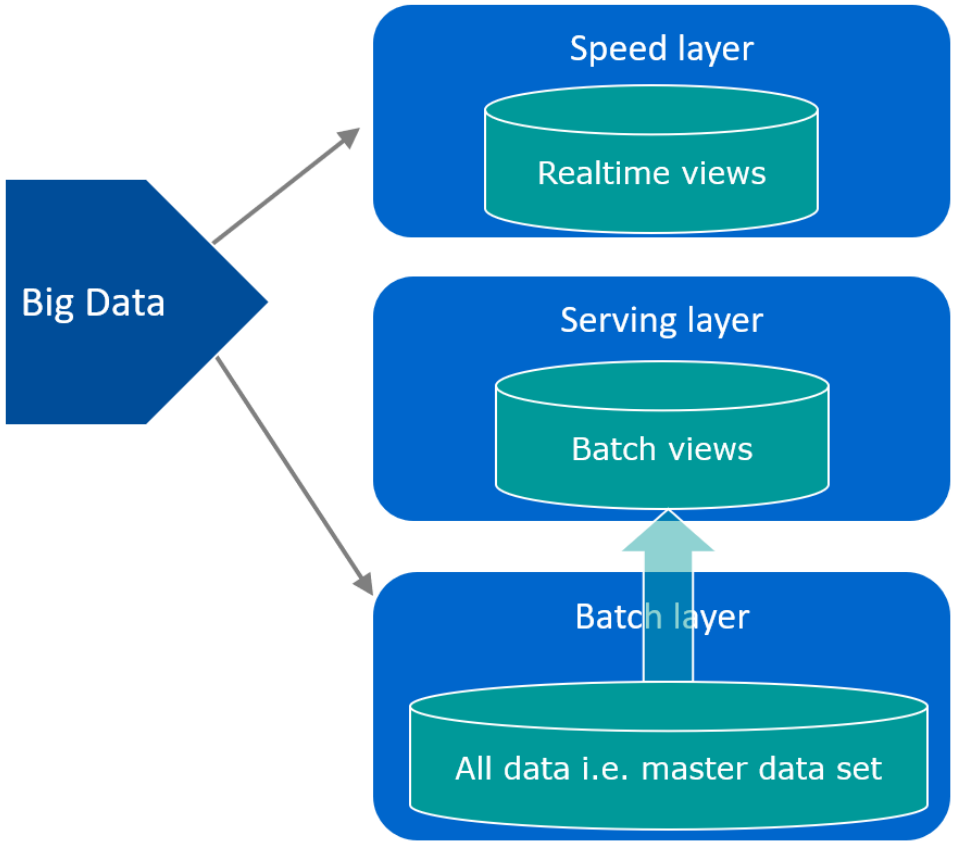
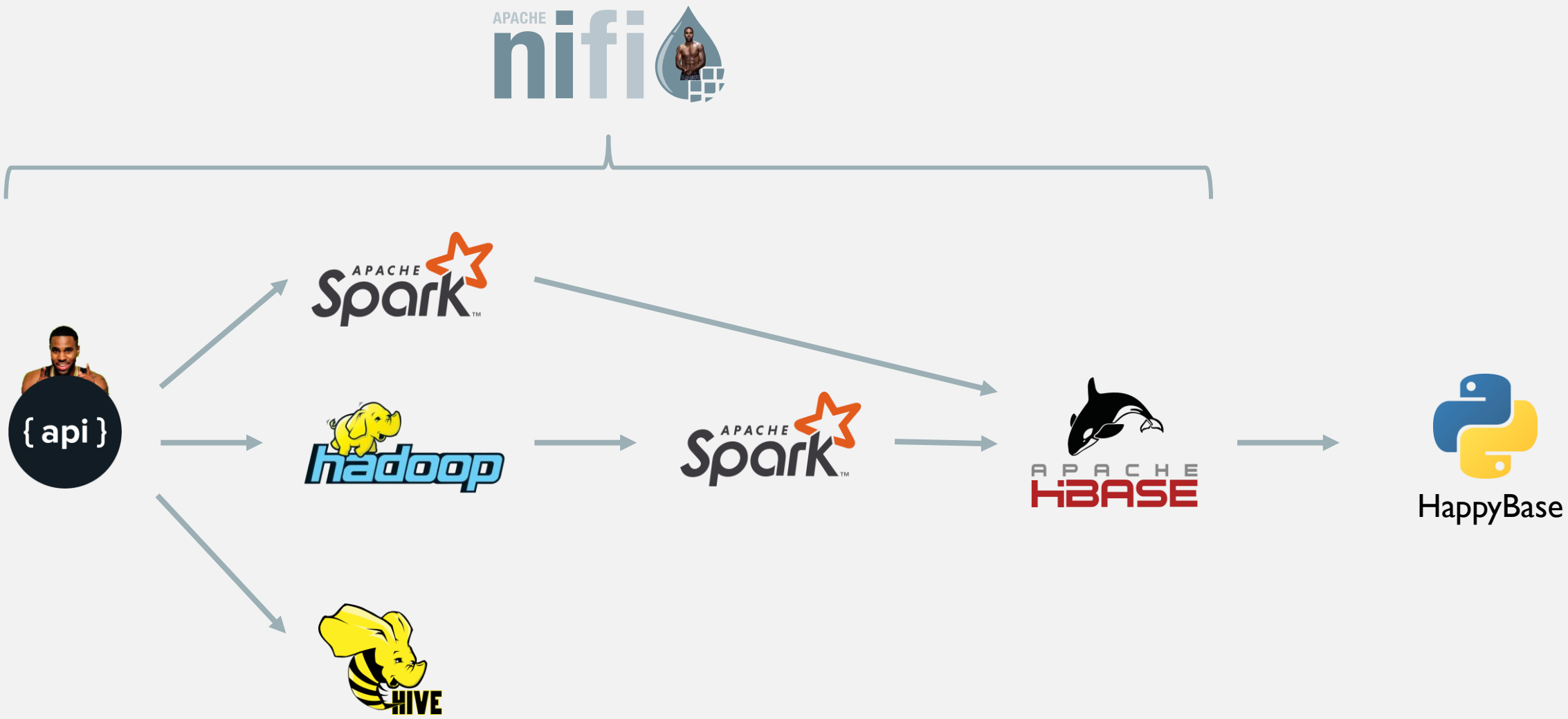
SKŁADOWANIE DANYCH W SYSTEMACH BIG DATA #BD3

Analiza pogody oraz zanieczyszczenia powietrza z wykorzystaniem systemów Big Data

ARCHITEKTURA LAMBDA

- Surowe dane dostępne, łatwa rekalkulacja widoków batchowych
- Appendowanie danych surowych zamiast transakcji
- Odporność na błędy ludzkie
- Lepsza niż Kappa, bo znajduje się później w greckim alfabecie



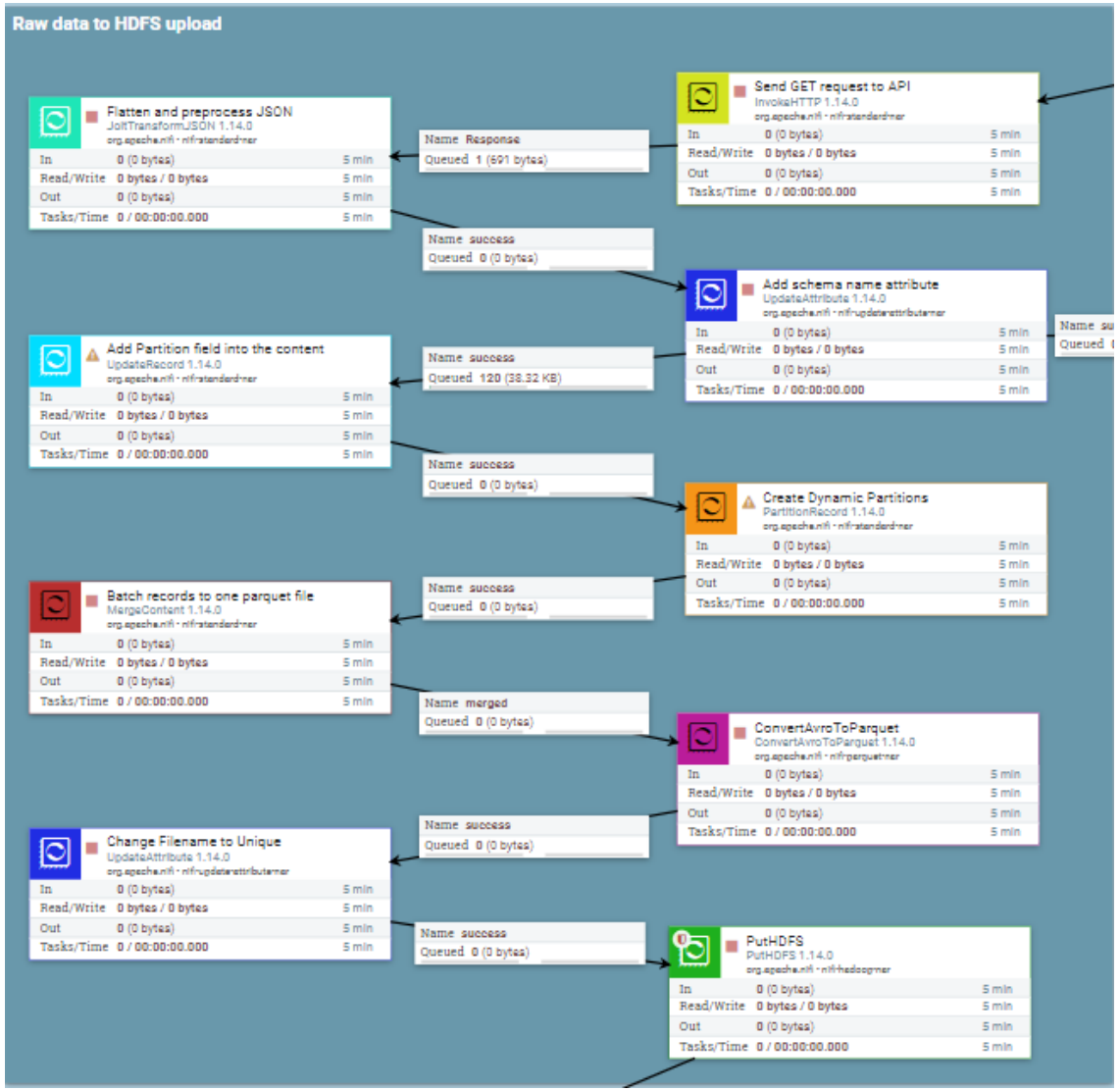

APACHE
nifi
APACHE
Spark
{ api }
hadoop
APACHE
Spark
APACHE
HBASE
HIVE
HappyBase

URUCHOMIENIE FLOW

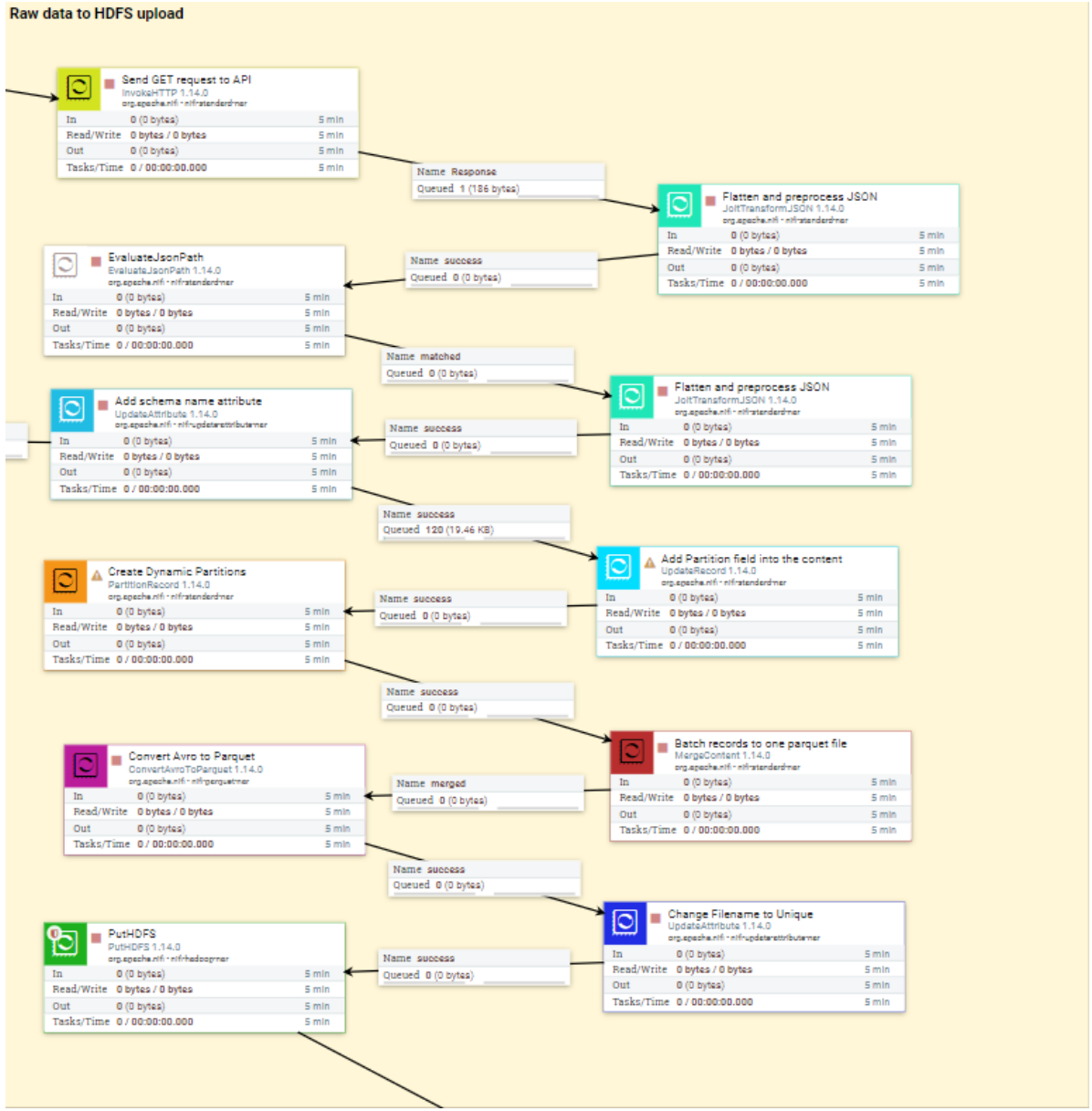
PRZYKŁADOWY CONFIG

```
{
   "requests":[
      {
         "city":"Radom",
         "lat":"50",
         "lon":"20"
      },
      {
         "city":"London",
         "lat":"50",
         "lon":"50"
      }
   ]
}
```

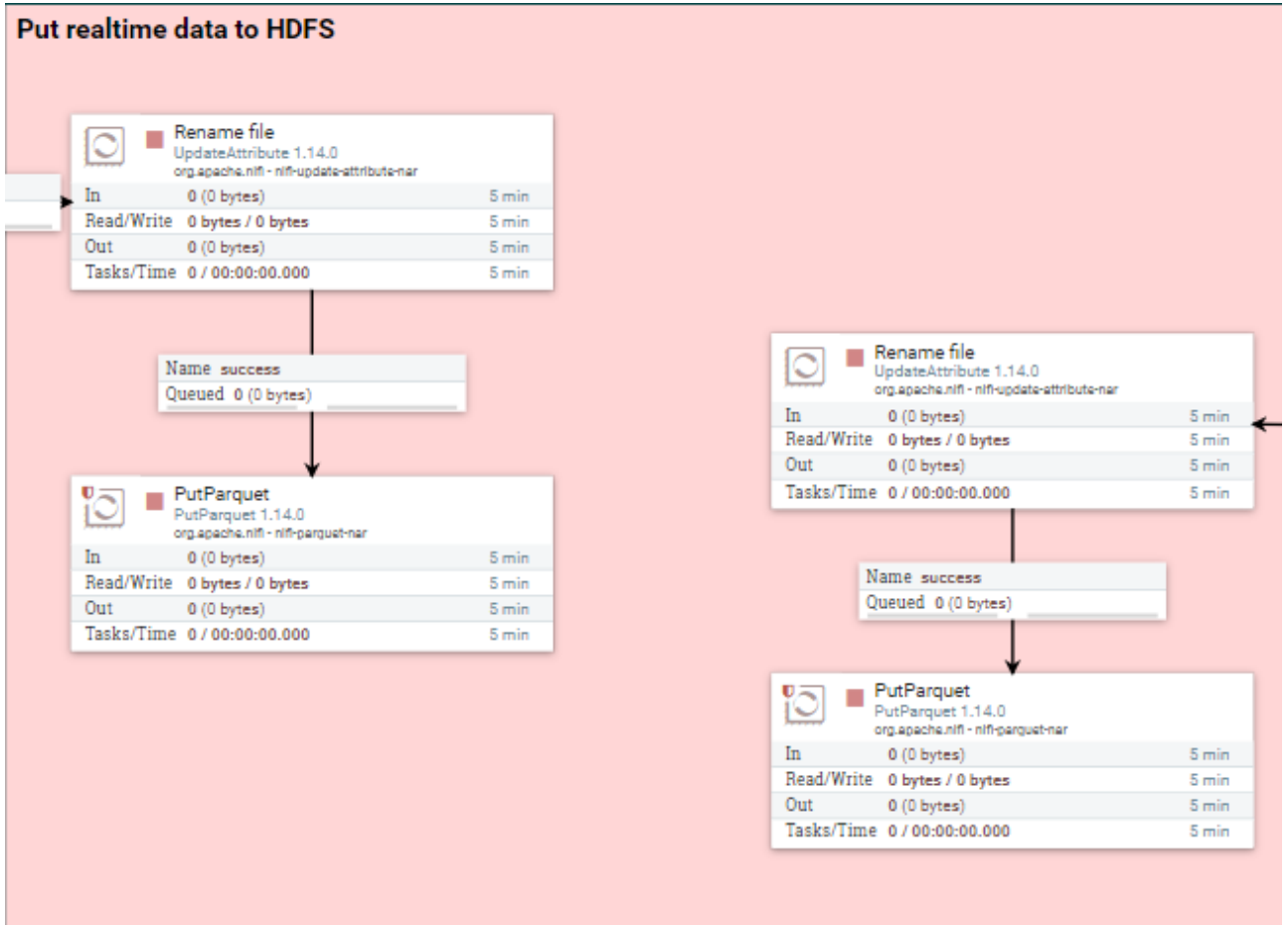
TRANSFORMACJA DANYCH O POGODZIE



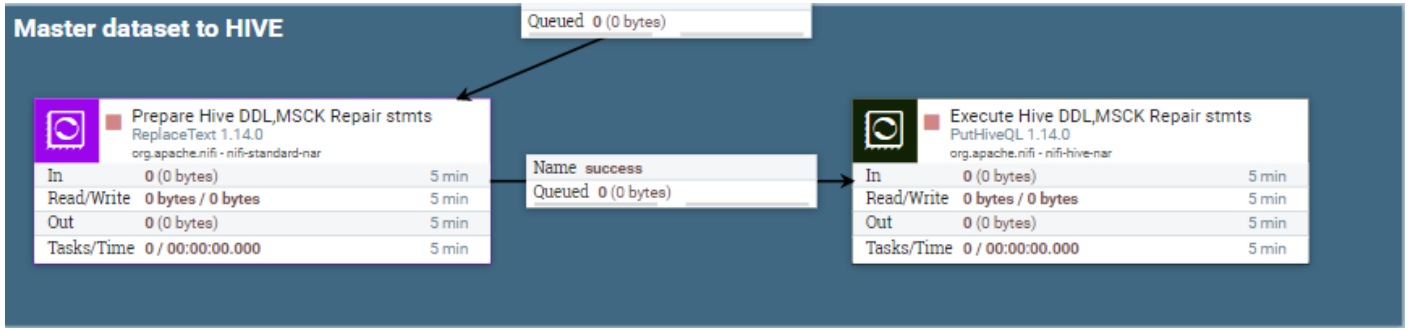
TRANSFORMACJA DANYCH O ZANIECZYSZCZENIACH

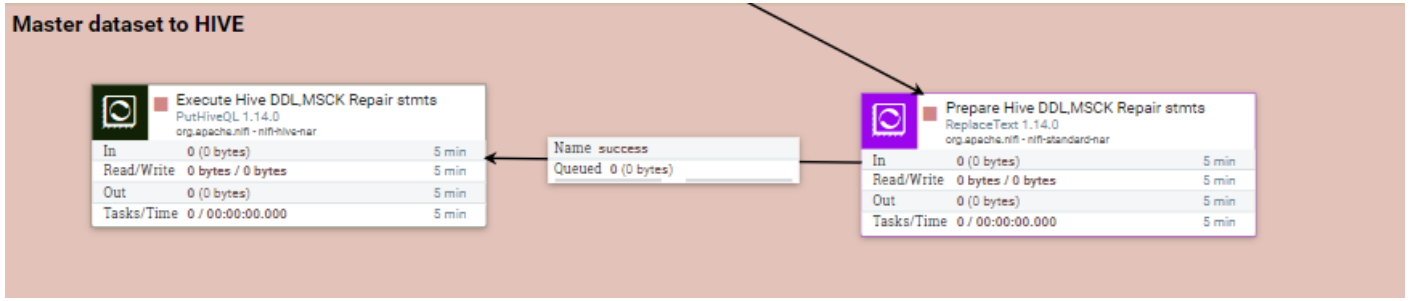


ZAPISANIE DANYCH BIEŻĄCYCH

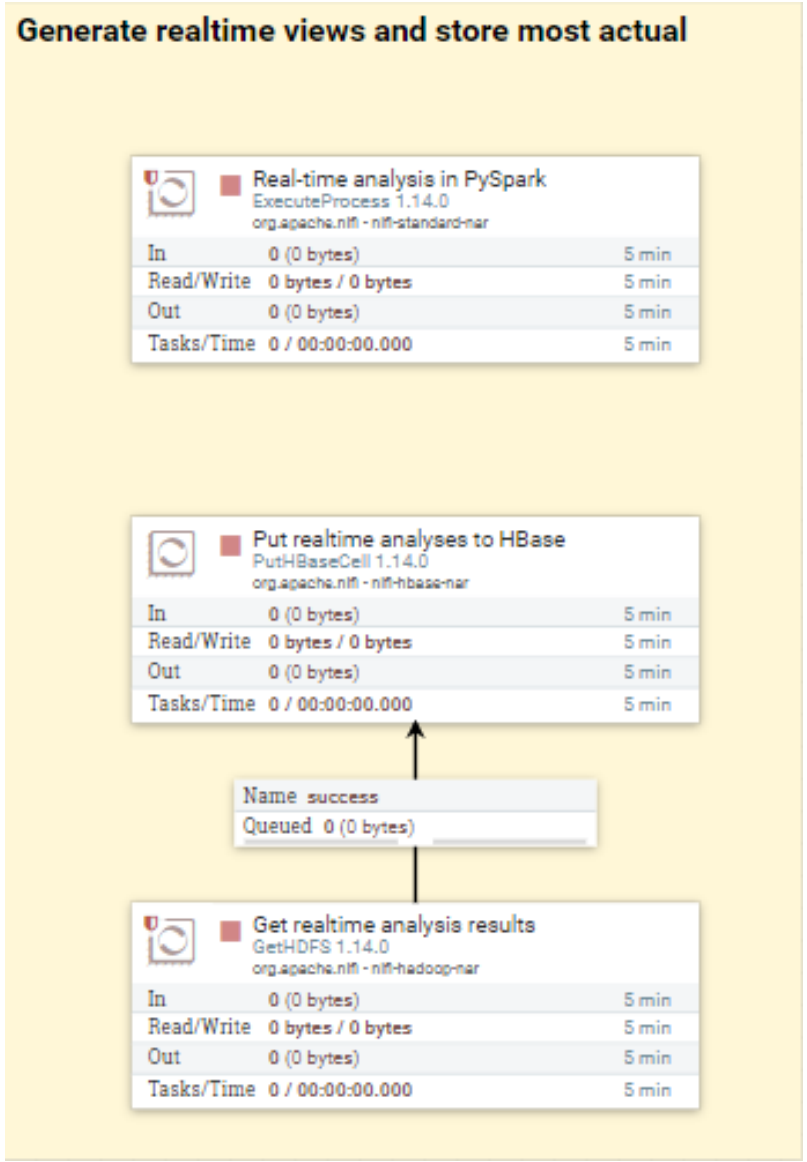


ZAPISANIE DANYCH DO HIVE

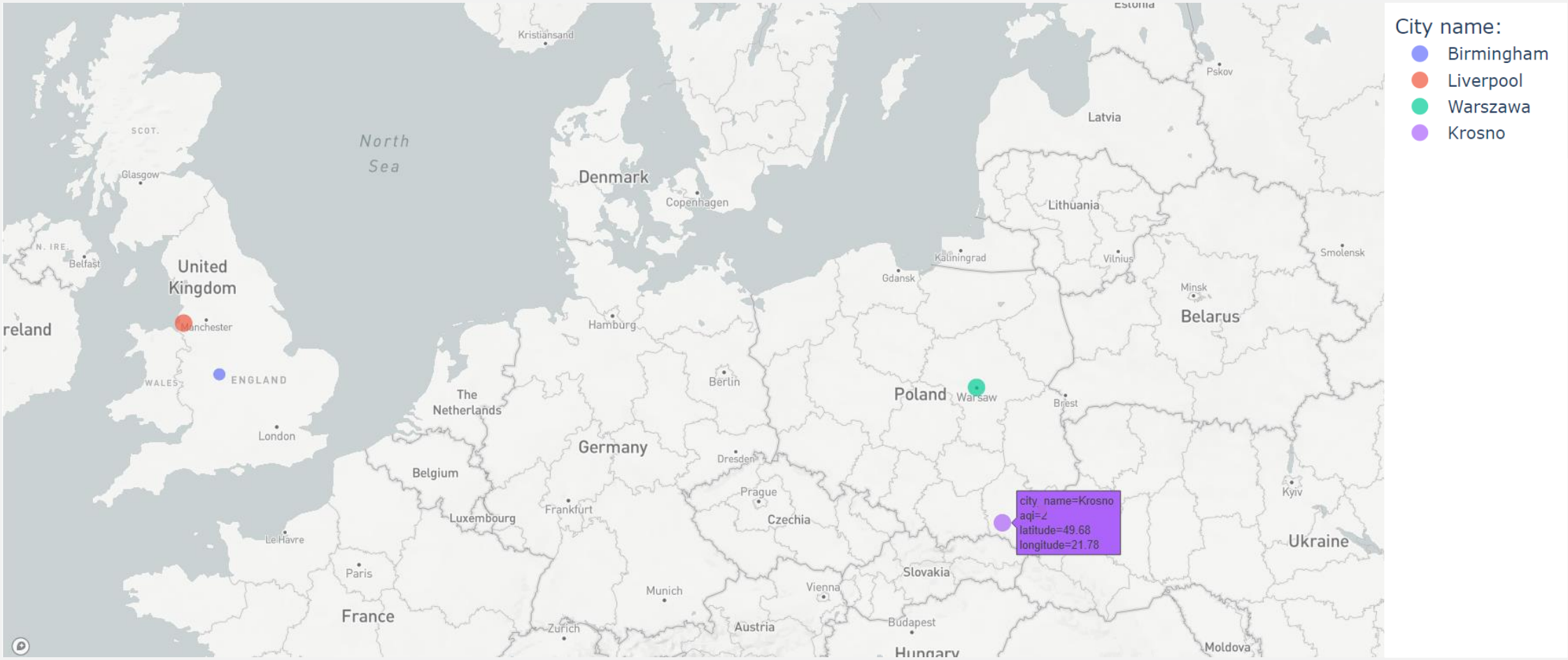




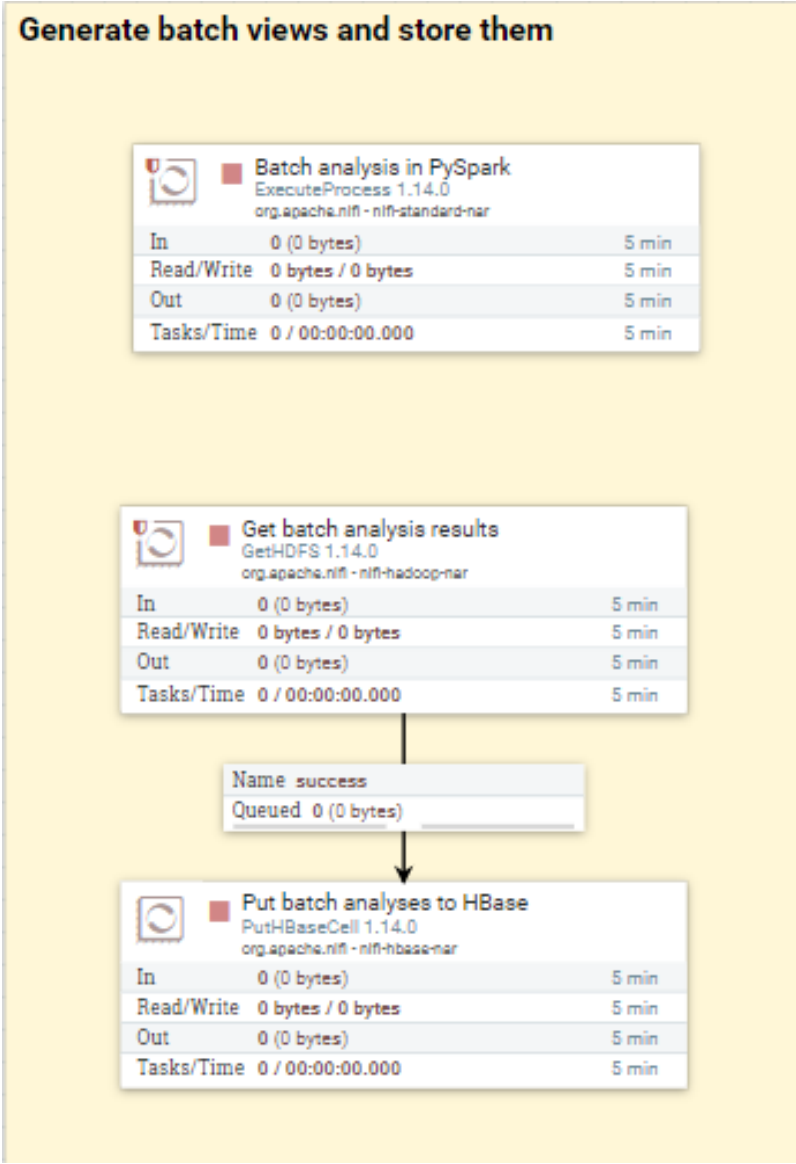
SPARK ANALIZA ONLINE



INTERAKTYWNA MAPA



SPARK ANALIZA BATCH



PRZYKŁADY WIDOKÓW BATCHOWYCH

	country	city_name	longitude	latitude	max_feelslike_temperature	max_temperature
0	Poland	Warszawa	21.000000	52.250000	1.3	6.0
1	United Kingdom	Liverpool	-2.940000	53.419998	10.0	10.1
2	United Kingdom	London	-0.110000	51.520000	10.5	10.5
3	United Kingdom	Birmingham	-1.860000	52.490002	9.7	9.5
4	Poland	Krosno	21.780001	49.680000	-1.4	4.1
5	Poland	Radom	21.150000	51.419998	-4.6	1.1

	country	city_name	max_aqi	max_co	max_no	max_no2	max_o3
0	United Kingdom	Birmingham	4	507.359985	65.269997	63.060001	75.820000
1	Poland	Warszawa	5	1428.599976	85.830002	54.840000	80.110001
2	Poland	Krosno	3	433.920013	1.730000	29.469999	77.959999
3	United Kingdom	London	5	794.409973	207.419998	71.290001	49.349998
4	United Kingdom	Liverpool	4	447.269989	65.269997	72.660004	90.120003

DZIĘKUJEMY ZA UWAGĘ!