

Air quality prediction based on the weather conditions in selected cities around the world

Team 2ql4sql – Marcel Wenka, Wojtek Replin,
Adam Kowalczyk and Adrianna Klimczak



Data sources

1. Air Quality - www.aqicn.org
2. Weather - www.openweathermap.org





Air Quality API



We are using following fields from AQ API:

- air quality index (scale)
- Unix timestamp of the measurement
- geographic coordinates of the requested location
- daily forecast of pollution for the requested location

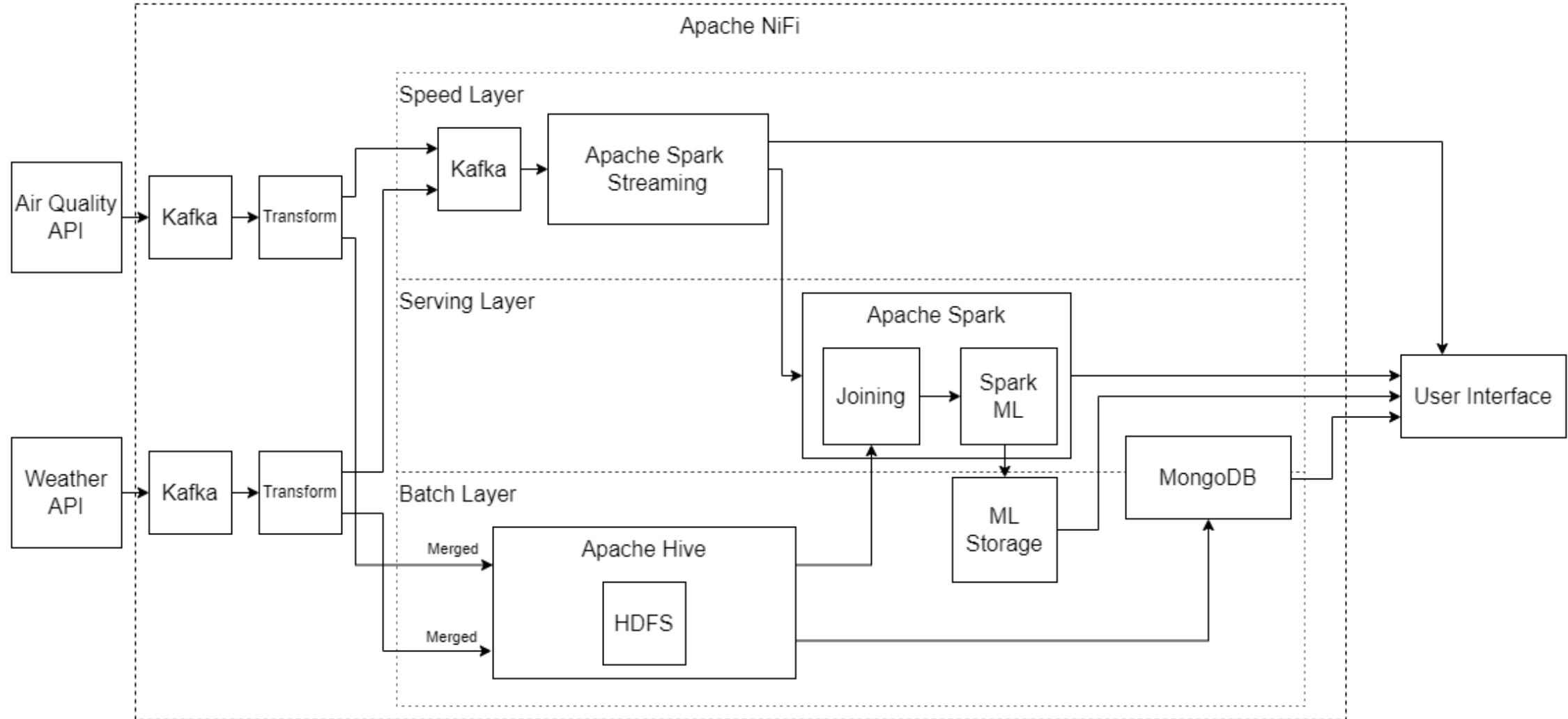
OpenWeatherMap API

We are using the following fields:

- coordinates of the requested location
- weather parameters
 - air temperature
 - air pressure
 - air humidity
- visibility in meters
- wind parameters
 - wind speed
 - wind angle in degrees
- Unix timestamp of the measurement



Architecture





Technologies

- Apache Nifi
- Apache Kafka
- Apache Hive
- Apache Spark Streaming
- Apache Spark + MLlib
- MongoDB





Data flow – current

1. Data is fetched from the APIs (Kafka).
2. NiFi collects the data by specifying the topic and transforms it.
3. Data is filtered to remove meaningless features (JOLT JSON)
4. Saved directly to HDFS to be accessed by HIVE via external tables

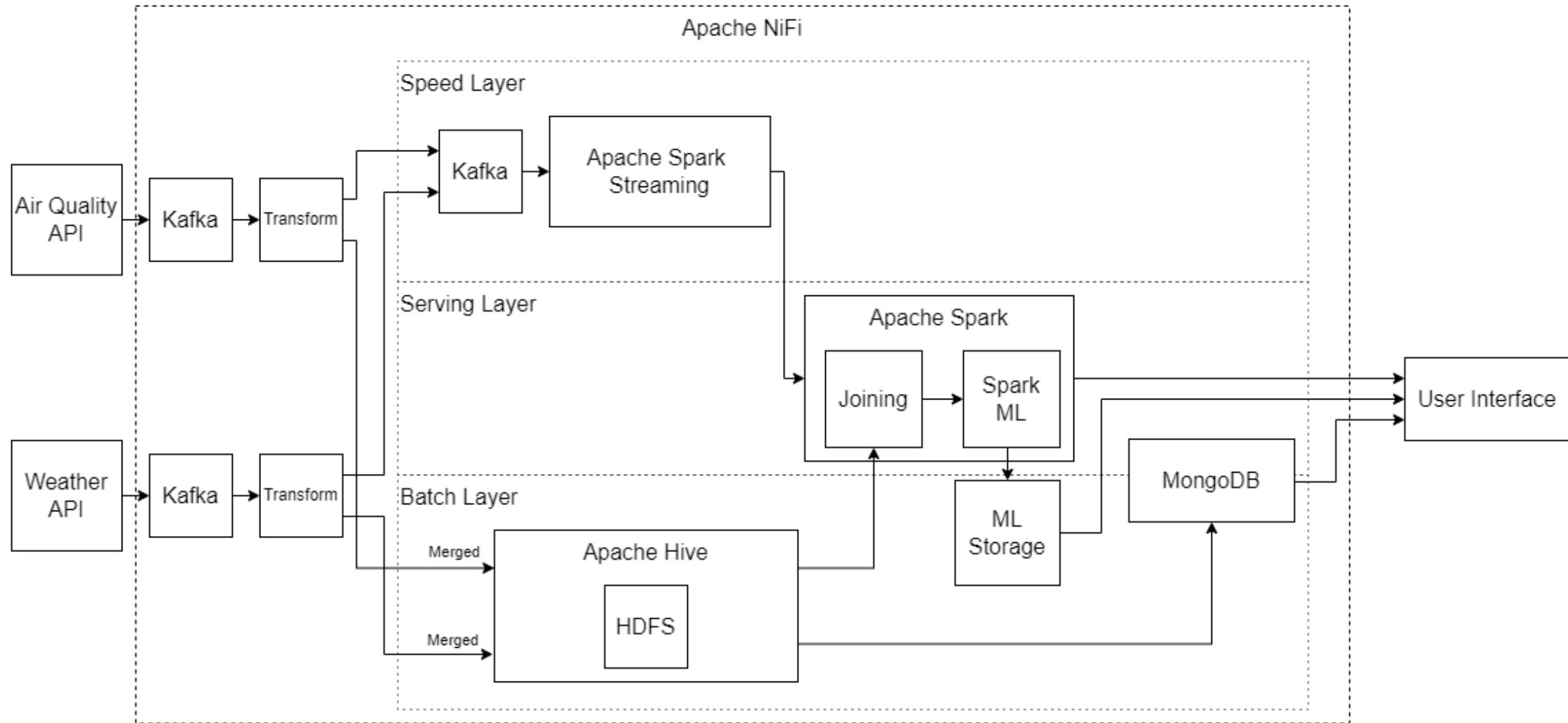




Data flow - plan

- 
1. Direct the data from Kafka to Spark components
 2. Feed to ML Model.
 3. Save the trained parameters of the model to ML Storage.
 4. Transfer to MongoDB database.
 5. Display the past measurements and weather on the UI

Lambda architecture



ML Model

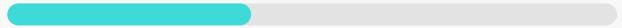
The data from the speed layer will be fed into Spark and a machine learning model will be trained on the joined weather and pollution data.



AIR POLLUTION FORECAST BASED ON WEATHER CONDITIONS IN: WARSAW

OK

Air status



57

Air quality index

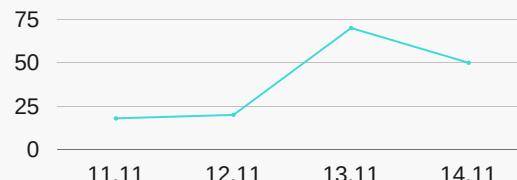


87%

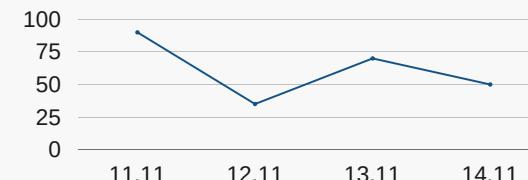
Prediction's accuracy



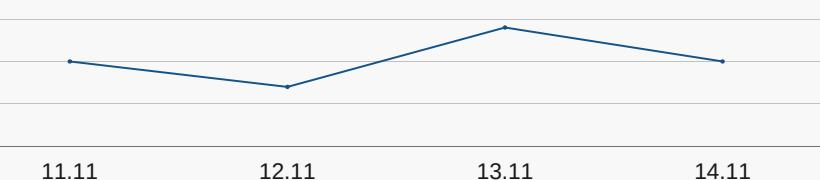
Air pollution history



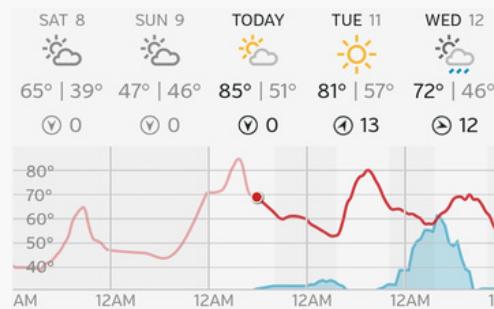
Air pollution prediction



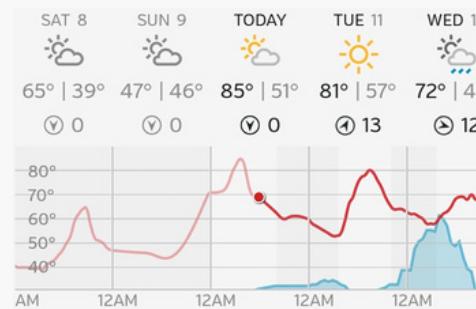
Model's accuracy in time



Weather history



Weather prediction



POSSIBLE CITIES





LIVE DEMO





QUESTIONS?

