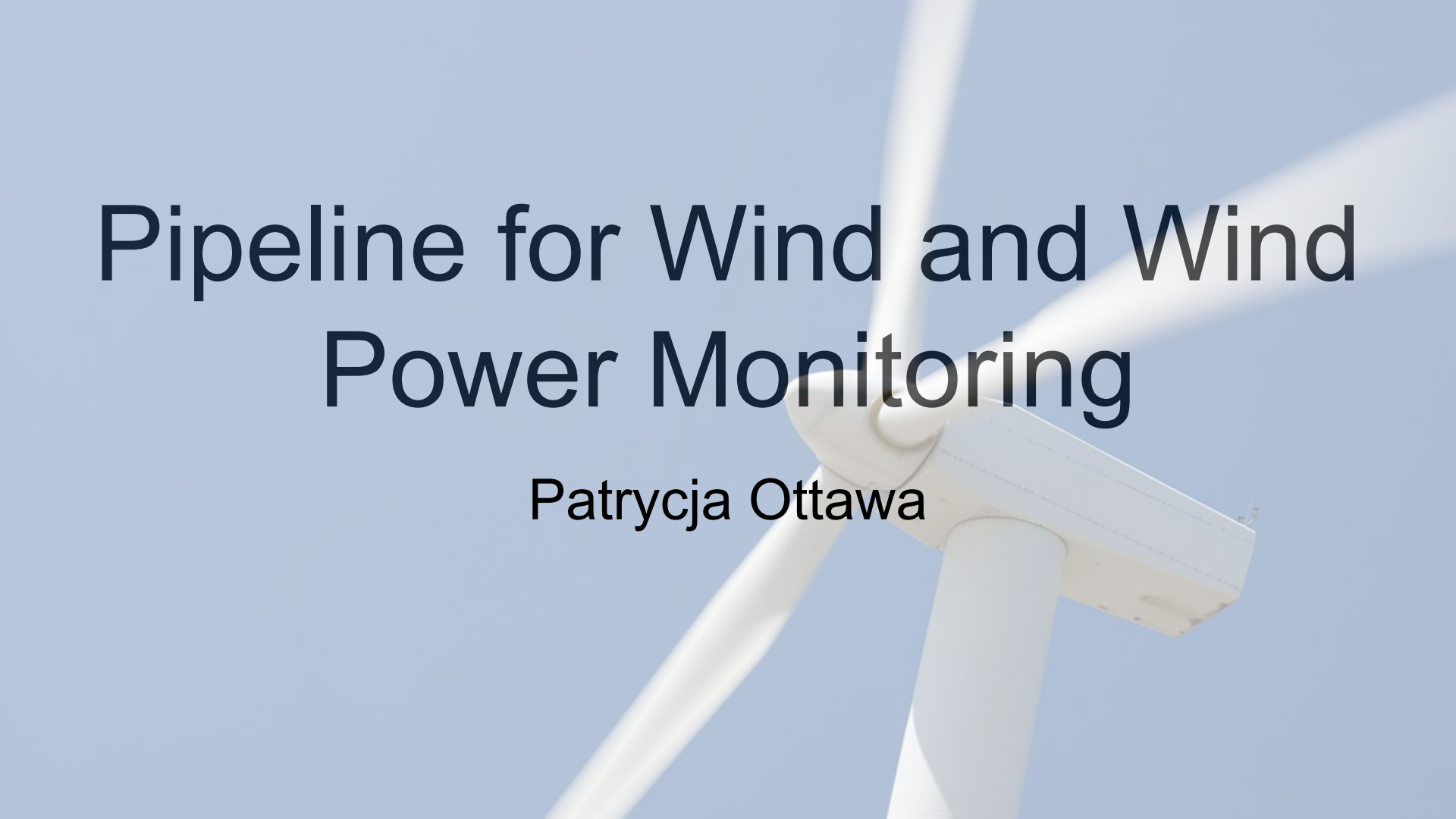
Pipeline for Wind and Wind Power Monitoring

Patrycja Ottawa



Agenda

Why?

The Project

Database

Challenges/Lessons Learned

Stack & Architecture

ToDo

Why Track Wind And Wind Power Production?

Possible use cases:

- wind turbines need to be switched off, in case of wind speed exceeding threshold set by wind turbines rotor limitations
- overproduction of energy could
  - overload the network
  - or be distributed differently (e. g. energy storage)
- overall, better energy management

The Project

Monitoring pipeline

Data Acquisition	Transform Data	Load Data into a DB	Serve Data
hourly: wind speed data scraping wind power data scraping via Github Actions **daily:** wind speed forecast data API acquisition	.json to .csv with python	**hourly/daily** with python to postgres hosted on Python Anywhere	Streamlit backed by postgres

Database

lookup table

regionID	region	lat	lon
9188113	Berg	47.9686	11.356
9375	Regensburg	49.047094	12.108713
9186	Pfaffenhofen	48.529853	11.508729

live monitor data

timestamp	regionID	windPower_kWh	solarPower_kWh	windSpeed_m/s	cloudCover_pct	temperature_C
1633521600	9188113	303.4414080414484	219.49882688921116	5.1	100.0	10.3
1633521600	9375	886.4224747790514	26327.061356436843	2.4	100.0	12.9
1633521600	9186	729.4201091332043	10486.623822515914	3.5	99.4	11.3

weather forecast data

timestamp	regionID	temperature_C	wind_speed_m/s	wind_direction	wind_gust	cloud_cover_pct	weathercode
1633564800	9188113	7.51	1.27	231.06	2.99	97.54	1001
1633568400	9188113	7.32	1.17	242.51	2.5	97.6	1001
1633572000	9188113	7.23	0.94	246.56	2.16	78.75	1001
1633575600	9188113	7.13	0.93	261.85	1.56	92.72	1001
1633579200	9188113	7.36	0.74	239.36	2.1	99.68	1001

Challenges/Lessons Learned

Data Acquisition

where to find “live” (?) energy power data or in at least hourly resolutions?

should I scrape?
→ GitHub Actions

different sources with different frequency
→ separation of scheduled cronjobs

Transform Data

nested dictionaries, in lists, in dictionaries, ∞

different date formats across APIs and scraping
→ unixtime

Are all units in sync?

how to organize the acquired .json files and .csv files
→ naming convention

scraped weather .json has no regionID?
→ lookup table

should I keep files on GitHub?

where to store the data?
→ Github/Webhook to Python Anywhere

makefile struggles
→ cronjobs

Load Data into a DB

how to connect the dots
→ python glue

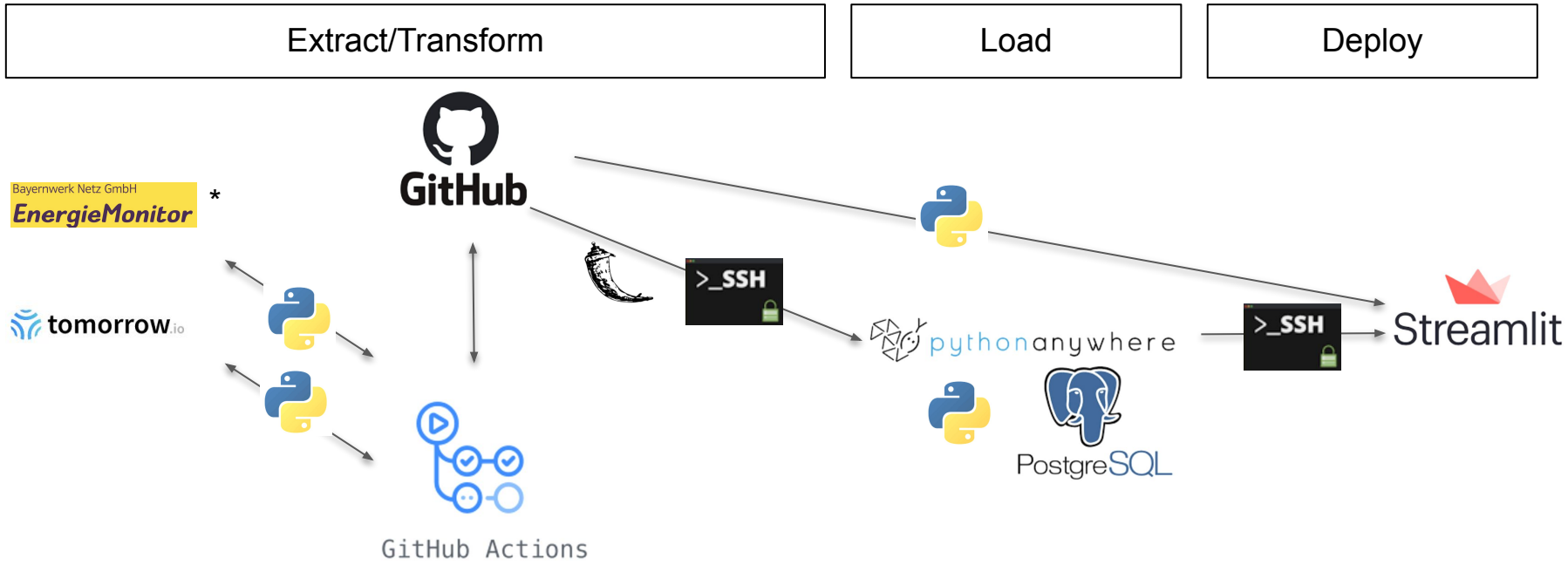
use the column names wisely
“wind_speed_m/s, regionID”

secrets !

Serve Data

how to establish the connection to streamlit
→ ssh

Stack & Architecture



* https://energiemonitor.bayernwerk.de/regensburg-landkreis

ToDo

- improve streamlit deployment
- feed in a lot more data
- integrate sunshine forecast API
- implement github secrets
- integrate forecasting for solar and wind power
- Great Expectations
- Logging & Error Handling
- Migration to AWS + Prefect
- expand tool to more regions

Streamlit App

https://share.streamlit.io/patrycjao/pipelineproject/main

Thank you!