

Energy vs. Meteo Data

Składowanie danych w systemach Big Data - projekt

Kacper Skonieczka, Grzegorz Zakrzewski

Cel:

- stworzenie systemu do gromadzenia, przetwarzania i analizowania danych meteorologicznych oraz informacji o energii wiatrowej

Próba odpowiedzi na pytania:

- jak prędkość wiatru wpływa na produkcję energii wiatrowej?
- jak produkcja energii wiatrowej wpływa na ceny energii?

Zamysł:

- wykorzystanie otwartych źródeł danych ENTSO-E i GEFS
- wykorzystanie technologii Big Data ze stosu Apache: HDFS, Hive, HBase, Spark



Dane opisujące:

- produkcję energii z elektrowni wiatrowych
- ceny energii elektrycznej (rynek dnia następnego)

Charakterystyka źródła:

- API dostępne za darmową rejestracją
- dane w formacie XML
- rekordy odpowiadają kolejnym godzinom w ciągu dnia



Dane opisujące:

- prognozy wartości zmiennych meteorologicznych, w tym składowych prędkości wiatru

Charakterystyka źródła:













- pliki z prognozami umieszczone na Amazon S3 Bucket
- dane w formacie GRIB2
- dane dostępne na każdy dzień w podziale na długość prognozy
- prognozy są przypisane do punktów na siatce współrzędnych

Źródło danych - GEFS





xarray.Dataset

► Dimensions: (latitude: 361, longitude: 720)

▼ Coordinates:

time	()	datetime64[ns]	2023-11-25	 
step	()	timedelta64[ns]	7 days	 
heightAboveGro...	()	float64	10.0	 
latitude	(latitude)	float64	90.0 89.5 89.0 ... -89.5 -90.0	 
longitude	(longitude)	float64	0.0 0.5 1.0 ... 358.5 359.0 359.5	 
valid_time	()	datetime64[ns]	2023-12-02	 

▼ Data variables:

u10	(latitude, longitude)	float32	...	 
v10	(latitude, longitude)	float32	...	 

Rysunek 1: Plik GRIB2 otwarty za pomocą biblioteki *xarray*

latitude	longitude	time	step	valid_time	u10	v10
90.0	0.0	2023-11-25	7 days	2023-12-02	2.43	0.66
90.0	0.5	2023-11-25	7 days	2023-12-02	2.44	0.64
90.0	1.0	2023-11-25	7 days	2023-12-02	2.44	0.62
90.0	1.5	2023-11-25	7 days	2023-12-02	2.45	0.59
90.0	2.0	2023-11-25	7 days	2023-12-02	2.45	0.57

Tabela 1: Przykładowe dane GEFS w formie tabelarycznej

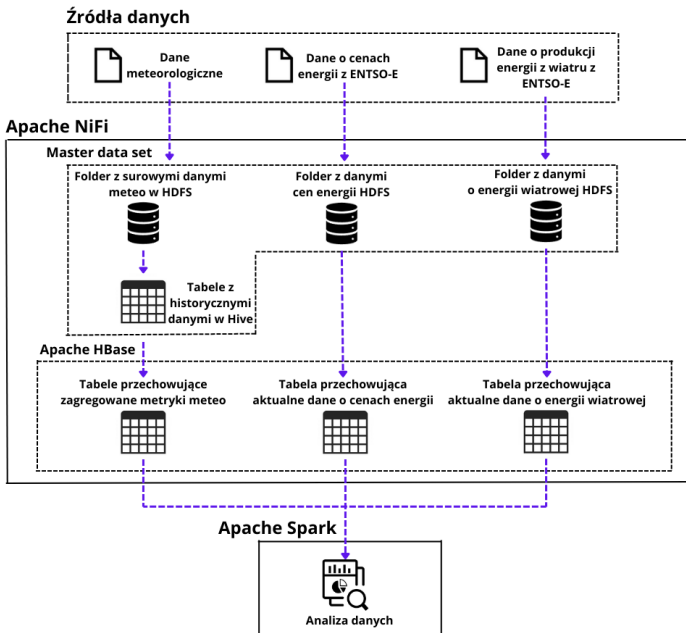
Źródła danych - podsumowanie

- **Źródło ENTSO-E:**
 - godzinowa częstotliwość danych
 - dzienna częstotliwość napływu
 - format XML
 - rynek energetyczny w Niemczech
 - ceny energii i produkcja energii wiatrowej
- **Źródło GEFS:**
 - dzienna częstotliwość danych
 - dzienna częstotliwość napływu
 - format GRIB2
 - długości geograficzne [5, 15] i szerokości geograficzne [45, 55]
 - w sumie 441 punktów na terenie Niemiec
 - tygodniowe prognozy składowych u i v prędkości wiatru
- zakres danych: 01.12.2023 r. - 08.01.2024 r.

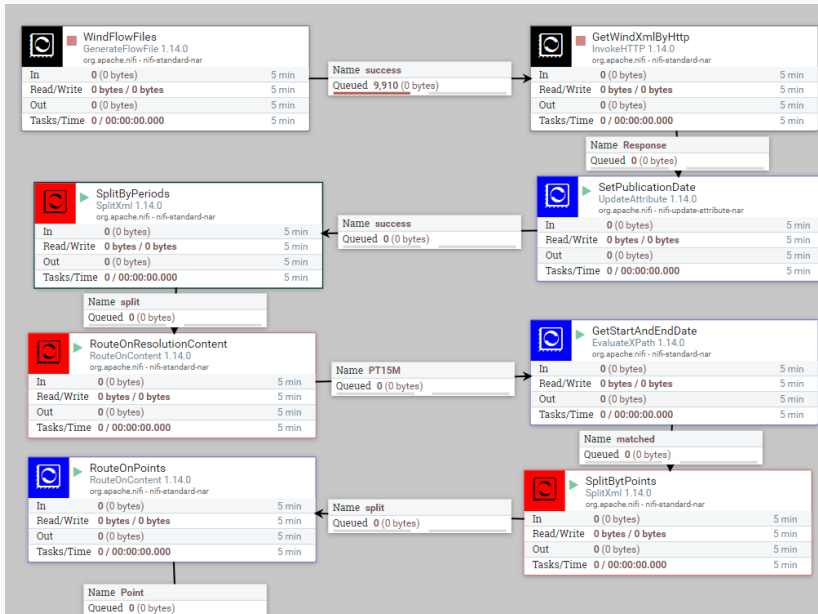
Zespół rozwiązań technologicznych z rodziny Apache:

- **Apache NiFi** - automatyzacja przepływu danych
- **Apache HDFS** - składowanie danych surowych
- **Apache Hive** - składowanie danych meteorologicznych
- **Apache HBase** - składowanie szeregów czasowych i zagregowanych danych meteorologicznych
- **Apache Spark** - wsadowa analiza danych

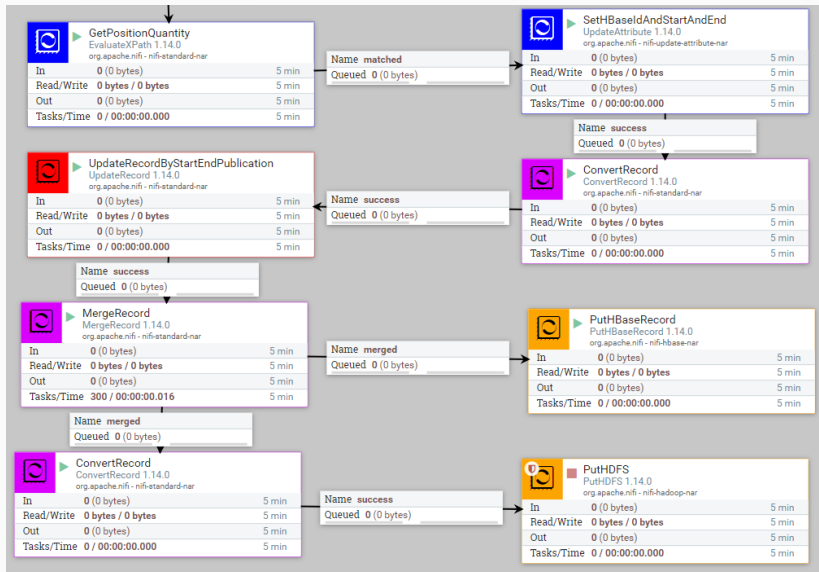
Architektura rozwiązania



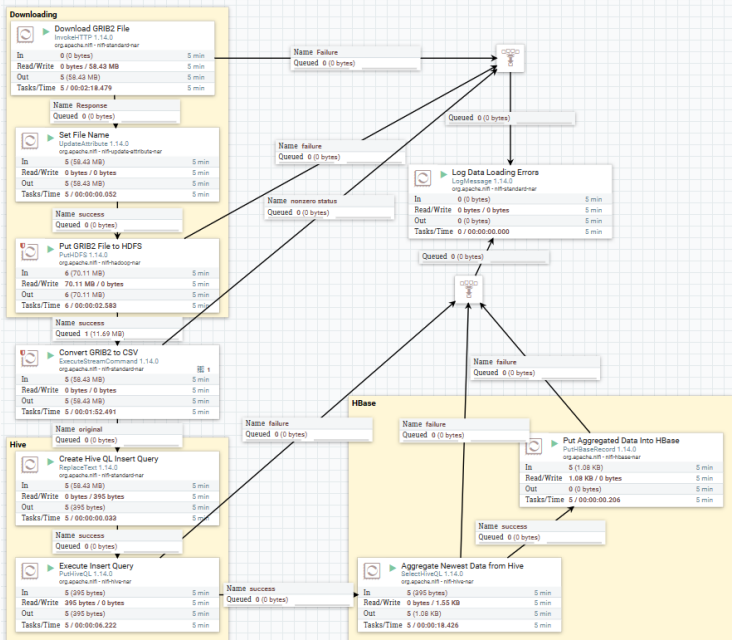
Zarządzanie przepływem danych - źródło ENTSO-E



Zarządzanie przepływem danych - źródło ENTSO-E



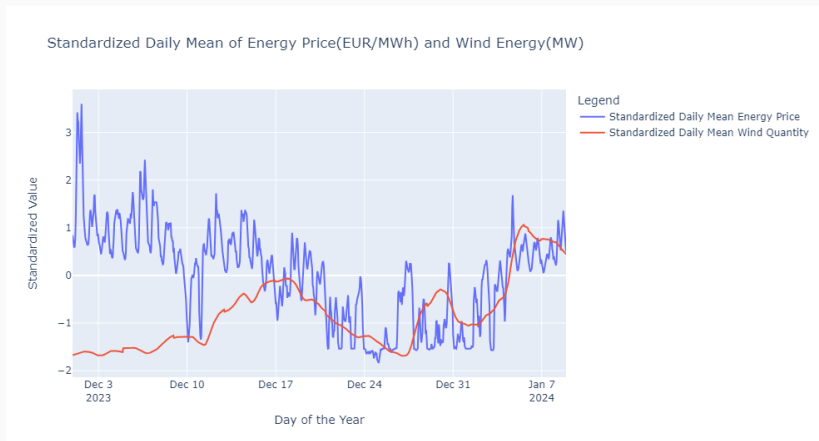
Zarządzanie przepływem danych - źródło GEFS



Warstwa analityczna - Apache Spark

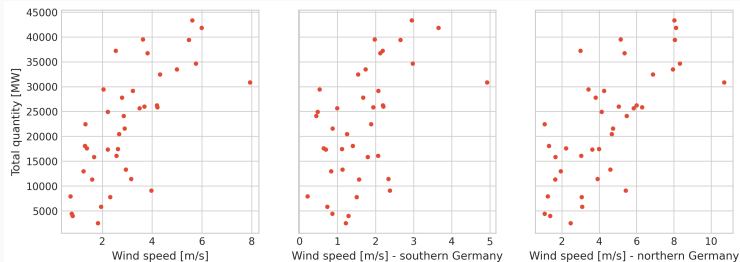


Rysunek 2: Minimalna, średnia i maksymalna cena energii w ciągu dnia



Rysunek 3: Ustandaryzowana średnia dzienna cena i produkcja energii

Warstwa analityczna - Apache Spark



Rysunek 4: Zależność produkcji energii wiatrowej od prędkości wiatru

	Wind speed	Wind speed - north	Wind speed - south	Mean price	Total quantity
Wind speed	1.00	0.97	0.81	-0.71	0.71
Wind speed - north	0.97	1.00	0.67	-0.70	0.70
Wind speed - south	0.81	0.67	1.00	-0.55	0.52
Mean price	-0.71	-0.70	-0.55	1.00	-0.88
Total quantity	0.71	0.70	0.52	-0.88	1.00

Rysunek 5: Macierz korelacji wszystkich zmiennych

Co się udało zrobić?

- zdobyć różnorodne dane energetyczne
- przygotować zaawansowany system Big Data zgodny z założeniami
- przygotować ciekawą analizę

Co można poprawić?

- rozszerzyć system na dane z całej Europy
- przygotować wizualizacje i szukać zależności, które adresują więcej pytań biznesowych