

Прогноз погоды

Docker PostgreS Airflow
WEB DE ML DQ
и с любовью к животным

Что там с погодой ?

Температура с метеодатчиков

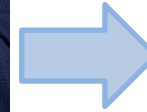
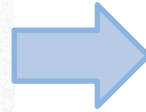
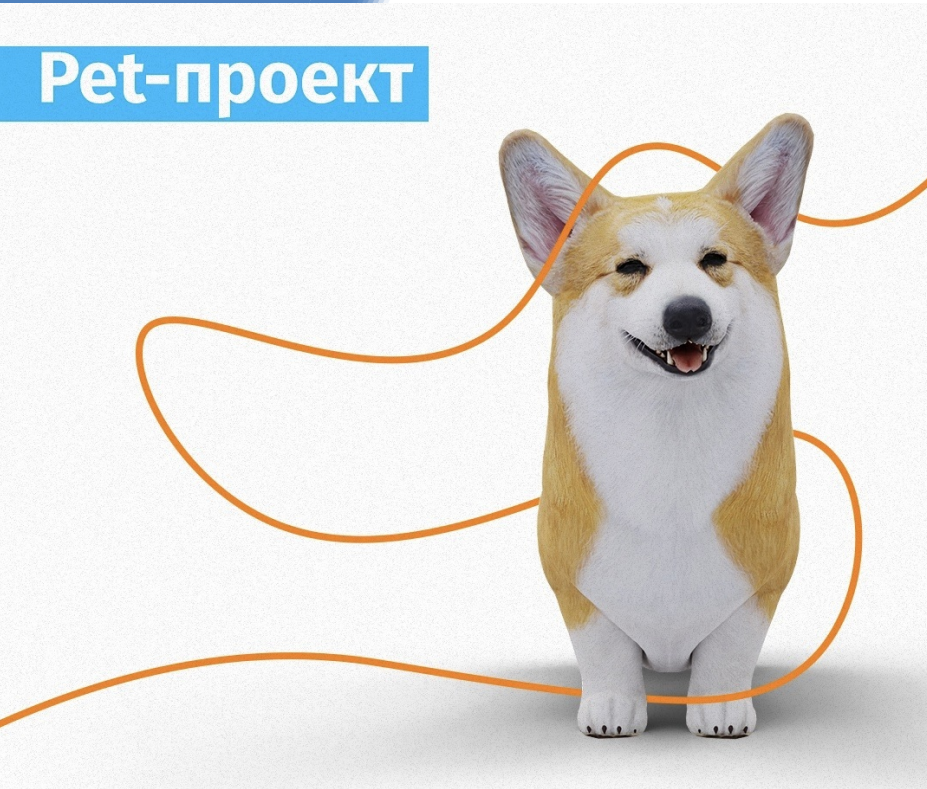
```
Погода в Москве (ВДНХ)
2024-05-12 21:00 3.8
2024-05-12 18:00 5.0
2024-05-12 15:00 6.0
2024-05-12 12:00 6.5
```

Прогноз температуры

```
Погода в Москве (ВДНХ)
2024-05-13 00:00 2.29
2024-05-13 03:00 0.88
2024-05-13 06:00 1.81
2024-05-13 09:00 3.88
```

Кратко :

Рет-проект



Программист чешет репу

О, всё работает!
Хороший мальчик!

Привет, Коллега! Есть время заняться новой задачкой от сантистов ?

Привет! Что за задача ? Надеюсь я справлюсь...
И, да, я в отпуске, так что время есть))

Руководство решило осчастливить пользователей новой функцией - прогноз погоды. Аналитики обещают рост посещаемости, но это нужно еще проверять... Поэтому пока только пилотный вариант «ВДНХ». Но! Базу проектируй так, чтобы добавить другие локации без доработок

подводные камни ?

Опытный ! Скоро Сеньором станешь))
Погода в источнике хранится в двух вариантах:
1. Замеры 1р / 3 часа - 0, 6, 9, 12, ... Можно выгрузить файлом за любой период. На этих данных я учу ML модели
2. Актуальные показания с метеодатчиков. Обновляются в произвольный момент, несколько раз в час, истории нет. Тебе брать здесь, усреднять за час и выбирать 4 трехчасовые замеры к данному часу, если 13ч - формируй температуру на 4, 7, 10, 13ч.

Пока задача не выглядит фантастически сложной.
Ничего не забыл мне сказать, чтоб потом не переделывать ?

Да, как раз собираюсь сказать: ты же Docker освоил уже ?

Ну как, освоил... С командами всё ясно и репозиторий завел, но ни одного контейнера пока не запустил

Уффф, отлично! Прямо выручил! А то и не знаю что делать.
DevOps, который должен этим заниматься, представляешь, тоже в отпуске. Только он телефон не берет, говорят на Мальдивах...

Вот ведь ж ! Я тоже так хочу

На Мальдивы ?

Нет, хотел бы телефон не брать! ну ок, потренируюсь с контейнерами.

Не злишь, зачтешь работу за Пет-проект в ЯндексПрактикуме, всё равно же будешь делать. И вот еще, нужно будет импортировать и использовать ML-объекты в Docker-контейнере. Ты сможешь, я в тебя верю)) Подробности почтой

Требования DS

Читаем из файла для обучения ML-модели

Формат данных, на которых обучалась ML-модель
В таком же формате нужно готовить данные

DT		T										
10.05.2024 15:00	12.0	m0	m1	m2	m3	hh	T-3	T-2	T-1	T0	T+1	
10.05.2024 12:00	10.3	сезон			час	температура в истории						
10.05.2024 09:00	7.3	0	1	0	0	3	3.0	2.6	7.3	10.3	12.0	
10.05.2024 06:00	2.6	0	1	0	0	0	2.9	3.0	2.6	7.3	10.3	
10.05.2024 03:00	3.0	0	1	0	0	21	3.4	2.9	3.0	2.6	7.3	
10.05.2024 00:00	2.9	0	1	0	0	18	5.9	3.4	2.9	3.0	2.6	

Формат поступающих актуальных данных

```
<id>http://rp5.ru</id>
<title xml:lang="ru">Погода в Москве (ВДНХ)</title>
<subtitle>Погода в Москве (ВДНХ). Погода на неделю в
<updated>2024-05-12T07:34:13+00:00</updated>
<link rel="self" href="http://rp5.ru/rss/5483/ru"/>
▼<author>
  <name>RP5</name>
```

на метеодатчиках

было

в среднем

```
<span class="t_0" style="display: inline;">+4.4</span>
(
  <span class="t_0" style="display: inline;">+3.2</span>
  ...
  <span class="t_0" style="display: inline;">+5.7</span>
)
<span class="t_0" style="display: inline;">°C</span>
```

. В Москве (ВДНХ) сегодня в 15:00 ожидается +7°C, без осадков, легкий ветер.
</summary>

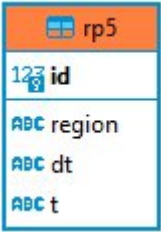
Детализация DWH

Source

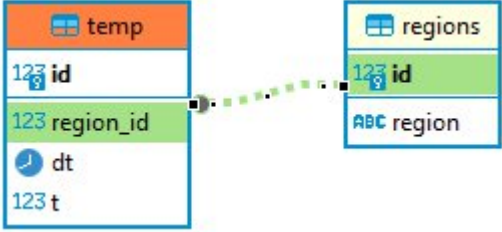
This XML file does not appear to have any style information associated with it. The document

```
<?xml version="1.0" encoding="UTF-8"?>
<feed xmlns="http://www.w3.org/2005/Atom" xmlns:rd="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:fh="http://purl.org/syndication/history/1.0" xmlns:fa="http://purl.org/atompub/facets/1.0" xmlns:rp="http://rp5.ru/2004/05/12/07:34:13+00:00">
  <id>http://rp5.ru/5483/ru</id>
  <title xml:lang="ru">Погода в Москве (ВДНХ)</title>
  <subtitle>Погода в Москве (ВДНХ). Погода на неделю в Москве (ВДНХ). Погода на де</subtitle>
  <updated>2024-05-12T07:34:13+00:00</updated>
  <link rel="self" href="http://rp5.ru/rss/5483/ru/">
  <author>
    <name>RP5</name>
    <email>support@rp5.ru</email>
    <uri>http://rp5.ru/docs/about/ru</uri>
  </author>
  <generator>http://rp5.ru</generator>
  <category xml:lang="ru" term="Погода"/>
  <logo>http://rp5.ru/images/ru/logo.png</logo>
  <rights xml:lang="ru">Copyright © ООО «Расписание Погоды», 2004-2024</rights>
  <fh:incremental>false</fh:incremental>
  <fa:max-age>10800000</fa:max-age>
  <entry>
    <id>http://rp5.ru/5483/ru#time_2024-05-12T07:34:13+00:00</id>
    <updated>2024-05-12T07:34:13+00:00</updated>
    <published>2024-05-12T07:34:13+00:00</published>
    <category xml:lang="ru" term="Погода"/>
    <title xml:lang="ru">8 10:34 на метеодатчиках</title>
    <link href="http://rp5.ru/5483/ru/">
    <summary xml:lang="ru">
      на метеодатчиках
      <span class="was_t">было</span>
      в среднем
      <span class="t_0" style="display: inline;">+4.4</span>
      (
      <span class="t_0" style="display: inline;">+3.2</span>
      ...
      <span class="t_0" style="display: inline;">+5.7</span>
      )
      <span class="t_0" style="display: inline;">°C</span>
      . В Москве (ВДНХ) сегодня в 15:00 ожидается +7°C, без осадков, легкий ветер.
    </summary>
  </entry>
</feed>
```

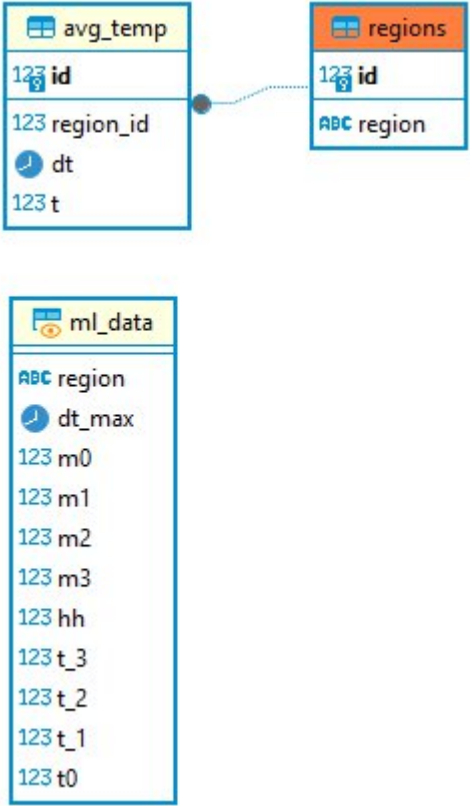
Stage



DDS



Mart



data_share_history.txt

Погода в Москве (ВДНХ)

2024-05-13	05:00	6.3
2024-05-13	02:00	3.6
2024-05-12	23:00	3.4
2024-05-12	20:00	4.0

data share future.txt

Погода в Москве (ВДНХ)

2024-05-13	08:00	6.65
2024-05-13	11:00	9.33
2024-05-13	14:00	9.24
2024-05-13	17:00	9.59

Общая структура решения




Суть решения



только с собакой

Форма решения

 **dockerhub**

Explore Repositories Organizations

ctrl+K ? ⌵ R

rus02

Search by repository name

All Content

Create repository

rus02 / weather_yandex_de_ml_server

Contains: Image • Last pushed: 4 minutes ago

Security unknown

☆ 0

↓ 4

Public

rus02 / weather_yandex_web_server

Contains: Image • Last pushed: 17 minutes ago

Security unknown

☆ 0

↓ 9

Public

Containers













[Give feedback](#)

Container CPU usage ⓘ
2.95% / 1000% (10 cores allocated)

Container memory usage ⓘ
2.64GB / 15.22GB

Show charts ▾

☰ ☑ Only show running containers

<input type="checkbox"/>	Name	Image	Status	CPU (%)	Port(s)	Last start... ▾	Actions
<input type="checkbox"/>	 weather_yandex_web_server caf83cb89b22 	rus02/weather_yandex_web_server:0.1	Running	0%	8889:80 	17 minutes ago	  
<input type="checkbox"/>	 weather_yandex_de_ml_server 03f5b7d10d2b 	rus02/weather_yandex_de_ml_server:0.1	Running	2.95%	3000:3000  Show all ports (2)	12 hours ago	  

Спасибо за внимание

