

В дополнительных заданиях можно реализовывать логику только частично, тогда баллы будут снижены. На оценку за каждое задание также влияет корректность шагов и качество кода.

Задания под * не являются обязательными, но будут учитываться при отборе участников.

Рекомендации к оборудованию и ПО

- Для эффективной работы тебе потребуется ПК / ноутбук.
- Если с работой программы возникнут проблемы, пиши на shift_support@cft.ru
- Установи и подготовь к работе ПО: Docker, Jupyter Notebook, IDE PyCharm, IntelliJ IDEA

Задание

Необходимо реализовать ETL¹ пайплайн (В данном случае приложение на Scala, Java или Python), которое загружает исходные данные, трансформирует их, а затем выгружает для выгрузки данных прогнозов погоды.

Для получения данных будет использован <https://open-meteo.com/>

Пример запроса данных по API open-meteo:

```
https://api.open-meteo.com/v1/forecast?latitude=55.0344&longitude=82.9434&daily=sunrise,sunset,daylight_duration&hourly=temperature_2m,relative_humidity_2m,dew_point_2m,apparent_temperature,temperature_80m,temperature_120m,wind_speed_10m,wind_speed_80m,wind_direction_10m,wind_direction_80m,visibility,evapotranspiration,weather_code,soil_temperature_0cm,soil_temperature_6cm,rain,showers,snowfall&timezone=auto&timeformat=unixtime&wind_speed_unit=kn&temperature_unit=fahrenheit&precipitation_unit=inch&start_date=2025-05-16&end_date=2025-05-30
```

Описание данных JSON в приложении, таблица "Описание данных JSON"

1 – ETL (Extract, Transform, Load) - это процесс, используемый для извлечения данных из разных источников, преобразования их в нужный формат и загрузки в целевое хранилище данных.

Задание 1: Наивная реализация

Реализовать приложение, которое будет делать следующее:

1. считывать json в формате API open-meteo, json взять из запроса в

"Пример запроса данных по API open-meteo"

2. Выполнять трансформации данных, которые на Ваш взгляд необходимы для получения данных в формате для выгрузки - т.е в формате, как в таблице **"Итоговая таблица"** из приложения.
3. Сохранить данные в csv. Поля, необходимые в csv и их описание представлены в таблице **"Итоговая таблица"** из приложения.

Задание 2*: Вставка в базу данных

Реализовать возможность вставки в базу данных в формате выходной таблицы из задания 1 (базу можно выбрать любую).

1. Реализовать базу данных в виде **docker-compose** скрипта + скрипты для создания таблицы и базы данных.
2. Предусмотреть случай вставки дубликатов.

Задание 3*: Выгрузка по запросу

Модифицировать ETL таким образом, чтобы в качестве входных параметров была предусмотрена возможность задать временной интервал, по которому будет осуществлена выгрузка через API open-meteo

1. Предусмотреть выгрузку в csv.
2. Предусмотреть выгрузку в базу данных.

Задание 4*: Подготовка тестового окружения

Подготовить docker compose, чтобы приложение можно было протестировать

- Подготовить docker контейнер для ETL приложения.
- Подготовить docker compose файл для базы данных и ETL приложения .

Настроить docker compose и/или docker container для ETL таким образом, чтобы можно было передав запущенному контейнеру с ETL команду осуществить: вставить данные из json в базу, положить данные из json в csv, вставить данные из API в базу, положить данные из API в csv.

Инженерный подход

Наличие документации (Readme, комментарии и т.д). Что делает ваше приложение и тп.

1. Наличие тестов. В том формате и объеме, в котором считаете нужным.
2. Проект представлен в виде git репозитория на gitlab или github.
3. Общее качество кода и структура проекта.
4. Пояснение решений (например аргументация реализации подхода для борьбы с дубликатами).

Приложение

Описание данных JSON

Общая информация		
latitude	Широта запрашиваемого местоположения.	55.0 (широта 55 градусов).
longitude	Долгота запрашиваемого местоположения.	83.0 (долгота 83 градуса).
generationtime_ms	Время, затраченное на генерацию ответа API (в миллисекундах).	0.2586841583251953(около 0.26 мс).
utc_offset_seconds	Смещение времени относительно UTC в секундах.	25200 (соответствует GMT+7).
timezone	Часовой пояс запрашиваемого местоположения.	"Asia/Novosibirsk".
timezone_abbreviation	Аббревиатура часового пояса.	"GMT+7".
elevation	Высота над уровнем моря для запрашиваемого местоположения (в метрах).	135.0 (135 метров).
Единицы измерения для почасовых данных (hourly_units)		
time	Формат времени для почасовых данных.	"unixtime" (Unix-время в секундах).
temperature_2m	Единица измерения температуры на высоте 2 метров.	"°F" (градусы Фаренгейта).
relative_humidity_2m	Единица измерения относительной влажности на высоте 2 метров.	"%" (проценты).
dew_point_2m	Единица измерения точки росы на высоте 2 метров.	"°F" (градусы Фаренгейта).

apparent_temperature	Единица измерения кажущейся температуры (ощущаемой температуры).	"°F" (градусы Фаренгейта).
temperature_80m	Единица измерения температуры на высоте 80 метров.	"°F" (градусы Фаренгейта).
temperature_120m	Единица измерения температуры на высоте 120 метров.	"°F" (градусы Фаренгейта).
wind_speed_10m	Единица измерения скорости ветра на высоте 10 метров.	"kn" (узлы).
wind_speed_80m	Единица измерения скорости ветра на высоте 80 метров.	"kn" (узлы).
wind_direction_10m	Единица измерения направления ветра на высоте 10 метров.	"°" (градусы).
wind_direction_80m	Единица измерения направления ветра на высоте 80 метров.	"°" (градусы).
visibility	Единица измерения видимости.	"ft" (футы).
evapotranspiration	Единица измерения испаряемости.	"inch" (дюймы).
weather_code	Код погоды (стандарт WMO).	"wmo code".
soil_temperature_0cm	Единица измерения температуры почвы на глубине 0 см.	"°F" (градусы Фаренгейта).
soil_temperature_6cm	Единица измерения температуры почвы на глубине 6 см.	"°F" (градусы Фаренгейта).
rain	Единица измерения количества дождя.	"inch" (дюймы).
showers	Единица измерения количества ливней.	"inch" (дюймы).
snowfall	Единица измерения количества снегопада.	"inch" (дюймы).
Почасовые данные (hourly)		
time	Массив Unix-времени для каждого часа.	[1747328400, 1747332000, ...].

temperature_2m	Массив значений температуры на высоте 2 метров.	[53.1, 51.4, ...].
relative_humidity_2m	Массив значений относительной влажности на высоте 2 метров.	[39, 34, ...].
dew_point_2m	Массив значений точки росы на высоте 2 метров.	[28.7, 23.9, ...].
apparent_temperature	Массив значений кажущейся температуры.	[47.2, 44.4, ...].
temperature_80m	Массив значений температуры на высоте 80 метров.	[53.3, 51.2, ...].
temperature_120m	Массив значений температуры на высоте 120 метров.	[53.0, 50.6, ...].
wind_speed_10m	Массив значений скорости ветра на высоте 10 метров.	[3.1, 4.2, ...].
wind_speed_80m	Массив значений скорости ветра на высоте 80 метров.	[9.5, 11.3, ...].
wind_direction_10m	Массив значений направления ветра на высоте 10 метров.	[94, 101, ...].
wind_direction_80m	Массив значений направления ветра на высоте 80 метров.	[99, 105, ...].
visibility	Массив значений видимости.	[79199.477, 79199.477, ...].
evapotranspiration	Массив значений испаряемости.	[0.001, 0.001, ...].
weather_code	Массив кодов погоды.	[3, 3, ...].
soil_temperature_0cm	Массив значений температуры почвы на глубине 0 см.	[51.4, 50.5, ...].
soil_temperature_6cm	Массив значений температуры почвы на глубине 6 см.	[53.8, 52.6, ...].
rain	Массив значений количества дождя.	[0.000, 0.000, ...].
showers	Массив значений количества ливней.	[0.000, 0.000, ...].
snowfall	Массив значений количества снегопада.	[0.000, 0.000, ...].

Единицы измерения для ежедневных данных (daily_units)		
time	Формат времени для ежедневных данных.	"unixtime" (Unix-время в секундах).
sunrise	Единица измерения времени восхода солнца.	"unixtime" (Unix-время в секундах).
sunset	Единица измерения времени заката солнца.	"unixtime" (Unix-время в секундах).
daylight_duration	Единица измерения продолжительности светового дня.	"s" (секунды).
Ежедневные данные (daily)		
time	Массив Unix-времени для каждого дня.	[1747328400, 1747414800, ...].
sunrise	Массив времени восхода солнца.	[1747347493, 1747433791, ...].
sunset	Массив времени заката солнца.	[1747405835, 1747492341, ...].
daylight_duration	Массив значений продолжительности светового дня.	

Итоговая таблица

avg_temperature_2m_24h	Средняя температура на высоте 2 метров за 24 часа (в градусах Цельсия).
avg_relative_humidity_2m_24h	Средняя относительная влажность на высоте 2 метров за 24 часа (в процентах).
avg_dew_point_2m_24h	Средняя точка росы на высоте 2 метров за 24 часа (в градусах Цельсия).
avg_apparent_temperature_24h	Средняя кажущаяся температура (ощущаемая температура) за 24 часа (в градусах Цельсия).
avg_temperature_80m_24h	Средняя температура на высоте 80 метров за 24 часа (в градусах Цельсия).
avg_temperature_120m_24h	Средняя температура на высоте 120 метров за 24 часа (в градусах Цельсия).
avg_wind_speed_10m_24h	Средняя скорость ветра на высоте 10 метров за 24 часа (в метрах в секунду).
avg_wind_speed_80m_24h	Средняя скорость ветра на высоте 80 метров за 24 часа (в метрах в секунду).
avg_visibility_24h	Средняя видимость за 24 часа (в метрах).
total_rain_24h	Общее количество дождя за 24 часа (в миллиметрах).
total_showers_24h	Общее количество ливней за 24 часа (в миллиметрах).
total_snowfall_24h	Общее количество снегопада за 24 часа (в миллиметрах).
avg_temperature_2m_daylight	Средняя температура на высоте 2 метров за период светового дня (в градусах Цельсия).
avg_relative_humidity_2m_daylight	Средняя относительная влажность на высоте 2 метров за период светового дня (в процентах).
avg_dew_point_2m_daylight	Средняя точка росы на высоте 2 метров за период светового дня (в градусах Цельсия).
avg_apparent_temperature_daylight	Средняя кажущаяся температура (ощущаемая температура) за период светового дня (в градусах Цельсия).
avg_temperature_80m_daylight	Средняя температура на высоте 80 метров за период светового дня (в градусах Цельсия).
avg_temperature_120m_daylight	Средняя температура на высоте 120 метров за период светового дня (в градусах Цельсия).
avg_wind_speed_10m_daylight	Средняя скорость ветра на высоте 10 метров за период светового дня (в метрах в секунду).

avg_wind_speed_80m_daylight	Средняя скорость ветра на высоте 80 метров за период светового дня (в метрах в секунду).
avg_visibility_daylight	Средняя видимость за период светового дня (в метрах).
total_rain_daylight	Общее количество дождя за период светового дня (в миллиметрах).
total_showers_daylight	Общее количество ливней за период светового дня (в миллиметрах).
total_snowfall_daylight	Общее количество снегопада за период светового дня (в миллиметрах).
wind_speed_10m_m_per_s	Скорость ветра на высоте 10 метров, преобразованная в метры в секунду.
wind_speed_80m_m_per_s	Скорость ветра на высоте 80 метров, преобразованная в метры в секунду.
temperature_2m_celsius	Температура на высоте 2 метров, преобразованная в градусы Цельсия.
apparent_temperature_celsius	Кажущаяся температура (ощущаемая температура), преобразованная в градусы Цельсия.
temperature_80m_celsius	Температура на высоте 80 метров, преобразованная в градусы Цельсия.
temperature_120m_celsius	Температура на высоте 120 метров, преобразованная в градусы Цельсия.
soil_temperature_0cm_celsius	Температура почвы на глубине 0 см, преобразованная в градусы Цельсия.
soil_temperature_6cm_celsius	Температура почвы на глубине 6 см, преобразованная в градусы Цельсия.
rain_mm	Количество дождя, преобразованное в миллиметры.
showers_mm	Количество ливней, преобразованное в миллиметры.
snowfall_mm	Количество снегопада, преобразованное в миллиметры.
daylight_hours	Продолжительность светового дня в часах для каждой записи. Вычисляется как разница между временем заката (sunset) и времени восхода (sunrise), переведенная в часы.
sunset_iso	Время заката в формате ISO 8601 (например, 2025-05-16T21:06:29Z).
sunrise_iso	Время восхода в формате ISO 8601 (например, 2025-05-16T04:32:25Z).