

Автоматическая рассылка выгрузок из БД на почту.

Мой последний рабочий проект. На данный завершён.

Решаемая проблема

Сотрудникам требовалось открыть несколько программ и тратить не менее 15 минут на выгрузку одного отчёта и его отправку почтой.

Задача

Сократить время создания отчётов.

Было принято решение создать веб-сервис для автоматической рассылки выгрузок из БД на почту получателей.

Используемые технологии: Python, Django, DRF, Docker, MS SQL Server

Дополнительные библиотеки: pandas, SQLAlchemy, SMTPlib

Как работает (основные этапы):

- 1) Пользователь указывает БД и выбирает SQL-view в форме UI
- 2) Создаётся подключение к выбранной БД
- 3) Данные считываются в файл внутри Docker-контейнера
- 4) Файл выгружается в сетевую папку или прикрепляется к письму
- 5) Письмо отправляется заказчику на почту.

Трудность 1

Некоторые отчёты содержали более 500к строк, из-за чего Docker-контейнер "падал". При этом в логах я не видел никаких ошибок. Позже выяснил, что происходит перезапуск всего приложения.

Решение

Я изменил логику, стал получать данные, разбивая на чанки.

Теперь все отчёты используют количество оперативной памяти не более фиксированного.

Трудность 2

В процессе работы над проектом выяснилось, что рабочий почтовый клиент имеет ряд ограничений на отправку файлов. Одно из них - вес файла до 20 Мб.

Пришлось думать, как производить замер объёма данных.

Столкнулся ещё с одной проблемой. Размер датафрейма, который хранит информацию о выгрузке

может быть как больше, так и меньше результирующего файла (из-за разных кодировок и способов хранения полей).

Решение

Сохранять данные в файлы, после проверять их действительный объём.

Те, которые весят менее 20 Мб отправлять почтой. Остальные - выкладывать в примонтированную сетевую папку.