

Исходные данные:

Есть excel файл с 6 листами, который скриптом разбивается на 5 CSV файлов (кроме листа ИТОГ).

Итого на входе программе нужно с определённой директории забрать 5 csv файлов и разложить их по таблицам БД (с предварительной обработкой)

Что необходимо реализовать:

Необходимо описать «архитектуру» решения:

- Составить ER-диаграмму с описанием таблиц и связей между ними для хранилища данных
- Описать, как и какими средствами забрать, обработать и загрузить данные в БД postgresSQL (ETL-процесс)

Пояснения:

- *Исходные данные на листах: «Отзывы на товары», «Отзывы на магазины», «Из внешних источников», «Опросы», «Другое».*
- *Итоговые данные будет использовать BI система, забирая данные из хранилища.*
- *Возможно лист ИТОГ не идеален и его можно изменить в БД, чтобы было удобнее с ним работать (предложите свой вариант).*
- *Главное, чтобы можно было получить в выборке для другой системы из БД точную копию листа ИТОГ (любым запросом).*
- *В идеале предложить «золотую середину» между скоростью обработки и потребляемыми ресурсами.*
- *Предлагаем выполнить решение на python (просто описание решения, сам код не нужен), но это не обязательно должен быть python.*
- *Для описания схемы БД предлагаем <https://dbdiagram.io/> (либо можете выполнить в привычном вам инструменте).*
- *В файле есть лист **уже есть в DWH** — в нём указаны уже существующие таблицы.*

В ответ ожидаем:

- Ссылку на ER-диаграмму или изображение/pdf файл с ER-диаграммой
- Описание ETL решения в текстовом виде