Инструкция по выполнению ДЗ к занятию «Внедрение и жизнь с Data Vault»

# Как развернуть проект у себя на компьютере?

1. Делаем Fork проекта <https://github.com/DWH-course-examples/dbt-datavault-postgres> и скачиваем к себе. В main лежит пустой проект без готовых моделей – вариант посложнее, в ветке practice-06.09 модели sql уже созданы и ими можно пользоваться - вариант попроще.
2. Устанавливаем Docker Desktop <https://www.docker.com/products/docker-desktop/>
3. В Терминале (например, Powershell в Windows) переходим в папку проекта, который вы скачали ранее.
4. В терминале выполняем команду docker-compose up – она развернет контейнеры, описанные в docker-compose файле. В этом файле для postgres по умолчанию настроены порты 5432:5432. Если вы хотите со своей локальной машины подключаться к БД с помощью через другой порт (например, 5433) – прописываете 5433:5432 (Так контейнер пробрасывает наружу порт 5433, а внутри себя и между контейнерами использует порт 5432).
5. Проверить, что контейнеры запущены: команда docker ps в терминале или в десктопном приложении Докера проверить, что 2 контейнера со статусом running. Должен быть отдельный контейнер для postgres и отдельный для dbt.
6. Убедимся , что dbt видит postgres и подключается к нему. Заходим в терминал dbt контейнера – либо через Docker desktop кликаем на CLI нужного контейнера, либо в терминале локального компьютера вводим команду : docker exec -it <имя\_или\_ID\_контейнера> /bin/bash – тогда вы попадете в терминал контейнера внутри терминала локальной машины.
7. Внутри контейнера нужно найти папку workspace – в ней будут файлы вашего проекта. Скорее всего, вам нужно будет попасть в корневую директорию с помощью cd .. – с помощью команды ls вы увидите директорию workspace – переходите в нее: cd workspace.
8. С помощью ls проверяете, что видите файлы вашего проекта, который скачали из гита.
9. Командой dbt –version проверяете, что установлен dbt-core и dbt-postgres
10. Dbt clean – для очистки зависимостей
11. Git config - -global http.sslVerify false – отключаем проверку сертификатов, иначе не получится скачать пакет automate\_dv
12. Dbt deps – скачивает пакеты из файла packages.yml
13. Dbt debug – проверяем что dbt успешно подключился к postgres

# Как выполнять dbt скрипты?

1. Dbt seed – заберет все данные из источников и положит в БД как есть.
2. Dbt run – запускает все преобразования из всех sql-моделей проекта
3. dbt run -s +hub\_customer – запусккает только модель hub\_customer
4. dbt run -s v\_stg\_orders – запускает только модель v\_stg\_orders

# Как подключиться к БД со своего компьютера с помощью Dbeaver (или любой другой IDE)?

1. Нужно перейти в терминал postgres. Либо через Docker Desktop CLI , либо через docker exec -it <имя\_или\_ID\_контейнера> /bin/bash
2. Переходим в директорию /var/lib/postgresql/data
3. Открываем файл postgresql.conf : nano postgresql.conf (Если nano не установлен – его надо устаноить). Должна быть строка listen\_addresses = '\*' и строка port = 5432 должна быть раскоментирована. Редактируем файл , если надо , сохраняем и закрываем.
4. В этой же директории открываем pg\_hba.conf. В конце файле нужно вставить строку:

host all all 0.0.0.0/0 md5

Она разрешит подключаться к БД с локальной машины. Сохраняем и закрываем. Если были изменения нужно перезапустить postgres контейнер.

1. В Dbeaver создаем подключение к postgres . Хост – localhost. Порт – тот , который указан в docker-compose.yml слева от двоеточия. БД , логин и пароль берем из файла profiles.yml.
2. Проверяем есть ли соединение.

# Домашнее задание

1. Добавить в любой csv файл новую строку с новым бизнес-ключом. Построить все модели, описать, что произошло с данными на всех этапах.
2. Добавить в любой csv файл новую строку с существующим бизнес-ключом, но другими свойствами. Построить все модели, описать , что произошло с данными на всех этапах.
3. Удалить строку из любого csv файла. Построить все модели, описать , что произошло с данными на всех этапах.
4. Создать витрину, которая покажет динамику количества всех заказов в разрезе недели. Описать словами, как вы это сделали и что получилось.
5. Создать витрину, в которой все клиенты будут отсортированы по количеству заказов со статусом «завершен» от наибольшего к наименьшему.
6. В кабинете Отуса в ДЗ выслать ссылку на форк и текстом в чате описать, что сделали (или создать отчет о ДЗ в виде документа и прикрепить его).