Департамент образования и науки города Москвы

Государственное автономное образовательное учреждение

высшего образования города Москвы

«Московский городской педагогический университет»

Институт цифрового образования

Департамент информатики управления и технологий

Ли Александр Андреевич БД-241м

**Лабораторная работа 2.1 Моделирование данных и SQL для Data Engineering**

Направление подготовки/специальность

38.04.05 - Бизнес-информатика

Бизнес-аналитика и большие данные

(очная форма обучения)

Вариант 12

Москва

2024

**Цель работы:**

освоить принципы проектирования баз данных, создания структуры таблиц и загрузки данных в PostgreSQL.

# Основная часть

# Перед работать убедимся, что нету запущенных контейнеров

# Sudo docker compose stop

**Изображение выглядит как текст, электроника, снимок экрана, программное обеспечение

Автоматически созданное описание**

# Рис.1 Останавливаем контейнера

Запускаем контейнер с pgadmin и postgres16

**Sudo docker compose start**

**Изображение выглядит как текст, электроника, снимок экрана, программное обеспечение

Автоматически созданное описание**

Рис.2 Запускаем posgress16 и pgadmin

Проверяем ip адресс для pgadmin

**Sudo docker inspect -f ‘{{range. NetworkSetting.Network}}{{.IpAdress}}{{end}}’ postgres16**

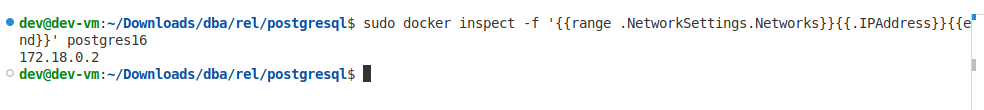
****

Рис.3 проверка ip адреса

Заходим в pgadmin через браузер

Используя адресс 192.168.31.249

Login: [admin@admin.com](mailto:admin@admin.com)

Pass: admin1617!

**В** lab\_02 создаем Staging layer(stg) и data warehouse(dw)

**Изображение выглядит как текст, снимок экрана, программное обеспечение, Значок на компьютере

Автоматически созданное описание**

**Рис.4 создание dw и stg**

Заполняем stg данными из github

Изображение выглядит как текст, программное обеспечение, Значок на компьютере, Мультимедийное программное обеспечение

Автоматически созданное описание

**Рис.5 заполняем данные**

**Изображение выглядит как текст, снимок экрана, программное обеспечение

Автоматически созданное описание**

**Рис.6 заполняем данные**

Проверяем данные

Изображение выглядит как текст, число, Шрифт, линия

Автоматически созданное описание

**Рис.7 проверка данных**

Делаем raw уровень

Заполняем данными таблицу pubic

Изображение выглядит как текст, снимок экрана, программное обеспечение, Значок на компьютере

Автоматически созданное описание

**Рис.8 заполнение таблицы public**

**Изображение выглядит как текст, снимок экрана, программное обеспечение, число

Автоматически созданное описание**

**Рис.9 Проверяем заполнение таблицы public**

Заполняем данными таблицу people

Изображение выглядит как текст, программное обеспечение, веб-страница, Веб-сайт

Автоматически созданное описание

**Рис.10 заполнение таблицы people**

**Изображение выглядит как текст, снимок экрана, программное обеспечение, Значок на компьютере

Автоматически созданное описание**

**Рис.11 Проверяем заполнение таблицы people**

Заполняем данными таблицу returns

Изображение выглядит как текст, снимок экрана, программное обеспечение, число

Автоматически созданное описание

**Рис.12 заполнение таблицы returns**

**Изображение выглядит как текст, снимок экрана, программное обеспечение, Значок на компьютере

Автоматически созданное описание**

**Рис.11 Проверяем заполнение таблицы returns**

Заполняем dw слой

Создаем первый справочник shipping\_dim

Изображение выглядит как текст, снимок экрана, программное обеспечение, веб-страница

Автоматически созданное описание

Изображение выглядит как текст, снимок экрана, программное обеспечение, веб-страница

Автоматически созданное описание

**Рис.12** shipping\_dim

Создания ship\_id

Изображение выглядит как текст, снимок экрана, Шрифт, программное обеспечение

Автоматически созданное описание

Рис.13 ship\_id

Изображение выглядит как текст, снимок экрана, программное обеспечение, число

Автоматически созданное описание

Рис.14 customer

Изображение выглядит как текст, снимок экрана, программное обеспечение, Шрифт

Автоматически созданное описание

Рис.15 geography

Заполняем пустые значения в таблицах dw и stg

Изображение выглядит как текст, снимок экрана, Шрифт, число

Автоматически созданное описание

Изображение выглядит как текст, снимок экрана, программное обеспечение, Шрифт

Автоматически созданное описание

Рис.16 заполнение пустых значений

Заполнение product

Изображение выглядит как текст, число, Шрифт, снимок экрана

Автоматически созданное описание

Рис.17 Заполнение product

Calendar

Изображение выглядит как текст, снимок экрана, число, Шрифт

Автоматически созданное описание

Рис18 calendar

Изображение выглядит как текст, программное обеспечение, веб-страница, Шрифт

Автоматически созданное описание

Рис.19 sales fact

Таблицы в dw

Изображение выглядит как текст, снимок экрана, программное обеспечение, веб-страница

Автоматически созданное описание

Рис 20 все таблицы в dw

Проверка качества данных

Изображение выглядит как текст, снимок экрана, программное обеспечение, дисплей

Автоматически созданное описание Изображение выглядит как текст, снимок экрана, программное обеспечение, дисплей

Автоматически созданное описание

Рис.21 проверка качества данных

Контрольные вопросы

1. В чем разница между схемами stg и dw?

В Staging мы храним сырые данные ,для возможности восстановления данных если слетит datawarehouse слой. В dw используется для хранения справочников.

Основное отличие заключается в их назначении.

2. Зачем нужны суррогатные ключи?

Суррогатные ключи нужны для того,чтобы у каждой строки в базе данных был свой собственный неизменяемый идентификатор.

3. Какие преимущества дает денормализация данных в DW?

Денормализация данных используется для улучшения производительности, позволяет выполнять запросы быстрее, упрощает написание запросов, уменьшает нагрузку на сервер, так как не нужно выполнять сложные операции join

# Заключение

В ходе практической работы, была создана база данных lab\_02 в pgadmin, 3 схемы (dw,stg,raw),внутри которых были созданы таблицы. Были написаны запросы к созданным таблицам и представлены ответы на вопросы.