

Для написания дипломной работы была использована база данных PostgreSQL, которая была развернута с помощью Docker.

## **1. Описание подготовки для работы с БД**

### **1.1**

Загрузка данных в контейнер

В рабочую папку мною были загружены 5 файлов в формате csv

### **1.2**

Переходим в рабочую папку и затем в папку с документами для запуска Docker и базы

Разворачиваем Docker и поднимаем контейнер

```
docker-compose --project-name postgres-client -f docker-compose.yml up --build -d
```

```
docker-compose --project-name postgres-client -f docker-compose.yml run --rm postgres-client
```

Проверяем данные и убеждаемся, что все четыре файла загружены:

```
ls -la data/
```

```
drwxr-xr-x 15 root  root    510 Aug 22 14:26 .
```

```
drwxr-xr-x  1 root  root   4096 Aug 22 14:29 ..
```

```
-rw-r--r--  1 root  root   60681 Aug 22 14:27 all_desktop.csv
```

```
-rw-r--r--  1 root  root   64994 Aug 22 14:27 id_info.csv
```

```
-rw-r--r--  1 root  root   54956 Aug 22 14:27 searching.csv
```

```
-rw-r--r--  1 root  root   66261 Aug 22 14:27 sharings.csv
```

### **1.3**

Переходим в БД для создания таблиц и запросов

```
psql --host $APP_POSTGRES_HOST -U postgres
```

## **2. Описание области бизнеса**

2.1 Для дипломной работы были взяты данные из аналитической системы для анализа статей на развлекательном портале

2.2. Данные были выгружены с помощью ClickHouse, куда ведутся записи событий

2.3 Использованы данные, записанные с 1 по 21 августа, которые содержат следующие сущности:

docUuid - id статьи

date - дата, когда на материале были посетители

visitors (visits) - общее количество посетителей с мобильной и десктопной версии

timeReadExpected - примерное время чтения в секундах (рассчитывается по количеству слов)

isLong - 0 в данной колонке означает, что статья короткая, 1 - что длинная (больше X слов)

MAX(shareTotal) - общее количество шар материала в социальных сетях (Facebook, VK, OK.ru)

visits\_social - общее количество переходов из социальных сетей (мобильный+десктоп)

desktop\_all\_V - всего визитов, десктоп (то есть не мобильная версия сайта)

desktop\_all\_R - сколько визитов закончилось тем, что пользователь находился на странице больше X секунд

desktop\_all\_S3 - сколько визитов закончились тем, что пользователь долистал до конца статьи

social\_V - всего визитов из социальных сетей, десктоп (то есть не мобильная версия сайта)

social\_R - сколько визитов закончилось тем, что пользователь, перешедший из соцсети, находился на странице больше X секунд. Только десктоп

social\_S3 - сколько просмотров закончились тем, что пользователь, перешедший из соцсети, долистал до конца статьи. Только десктоп

desktop\_search\_V - всего визитов из поисковиков (Яндекс, Гугл и т.п.)), десктоп (то есть не мобильная версия сайта). Только десктоп

desktop\_search\_R - сколько визитов из поисковиков закончилось тем, что пользователь находился на странице больше X секунд. Только десктоп

desktop\_search\_S3 - сколько визитов из поисковиков закончилось тем, что пользователь долистал до конца статьи. Только десктоп

Визиты - не уникальные переходы из разных источников трафика

Шары - пользователь поделился материалом в социальных сетях

Десктопная версия - версия сайта для десктопа (с мобильных устройств пользователь видит другую версию сайта)

### **3. Описание таблиц БД**

3.1. content\_info - таблица, которая описывает основные характеристики статьи: id, дату выхода, общее количество посетителей, тип статьи (короткая или длинная), предполагаемое время чтения

Столбцы:

docUuid - id статьи, visitors, date, isLong, timeReadExpected

3.2 sharings - таблица, которая описывает характеристики, связанные с социальными сетями, причем не только с трафиком, но и с шарами

Столбцы:

docUuid - id

статьи, visitors, MAX(shareTotal), visits\_social, social\_V, social\_R, social\_S3

3.3 searching - таблица, которая описывает поведение пользователей, которые перешли из поисковиков

Столбцы:

docUuid - id статьи, visitors,  
desktop\_search\_V, desktop\_search\_R, desktop\_search\_S3

3.4 all\_desktop - таблица, которая описывает поведение пользователей

Столбцы: docUuid - id статьи, visitors,  
desktop\_all\_V, desktop\_all\_R, desktop\_all\_S3