

Лабораторная работа №7

В текущей работе вам потребуется использовать продукт **RT.WareHouse**, который развернут в рамках нашего курса.

Цель задания: пересоздать модель данных в виде которую вы применяли для HBASE базы данных.

- Соединитесь с GreenPlum, используя, например Dbeaver
- Посмотрите на **реляционную модель** данных

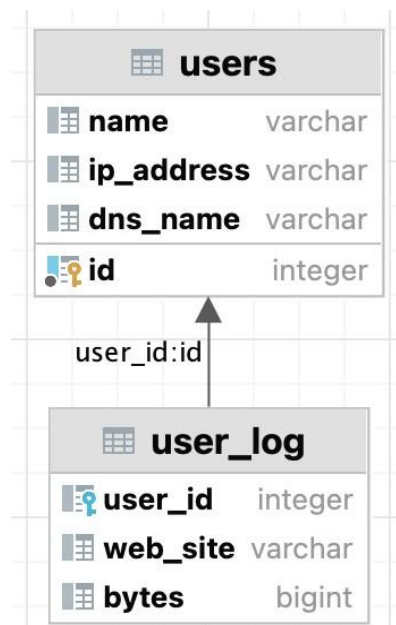


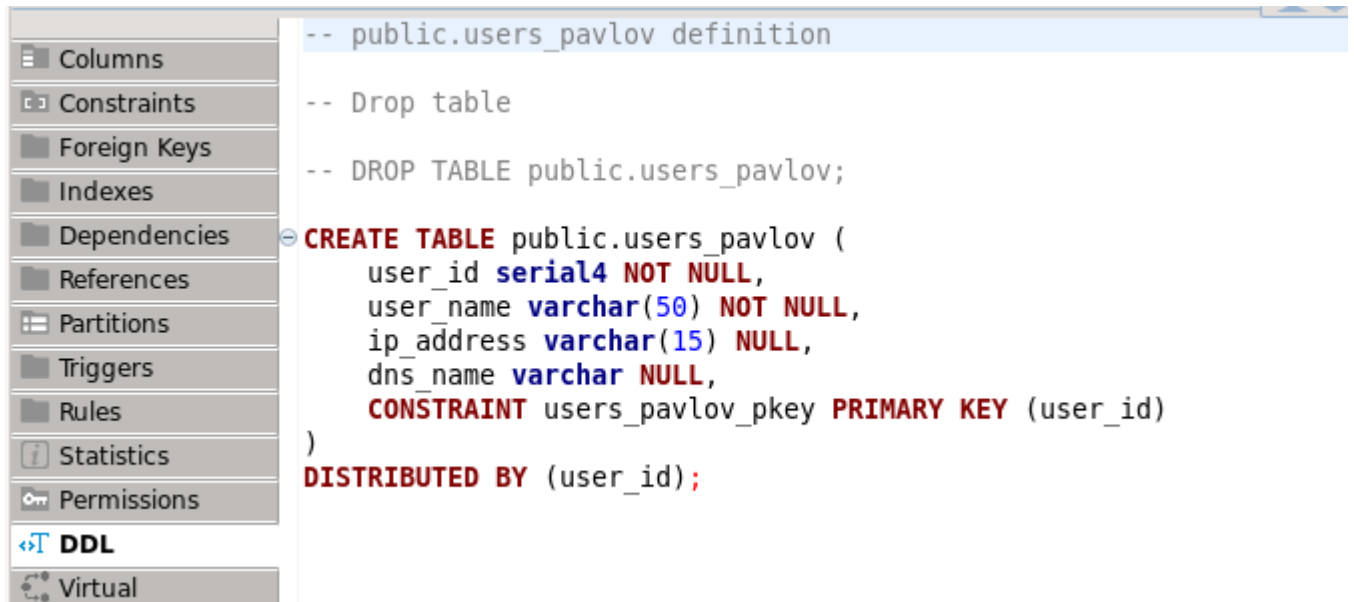
таблица *users*

Столбец	Описание
user_id	идентификатор пользователя
user_name	имя пользователя
ip_address	IP адрес пользователя
dns_name	Имя сервера пользователя

таблица *user_log*

Столбец	Описание
user_id	идентификатор пользователя
web_site	ссылка на веб сайт
bytes	количество загруженных байт

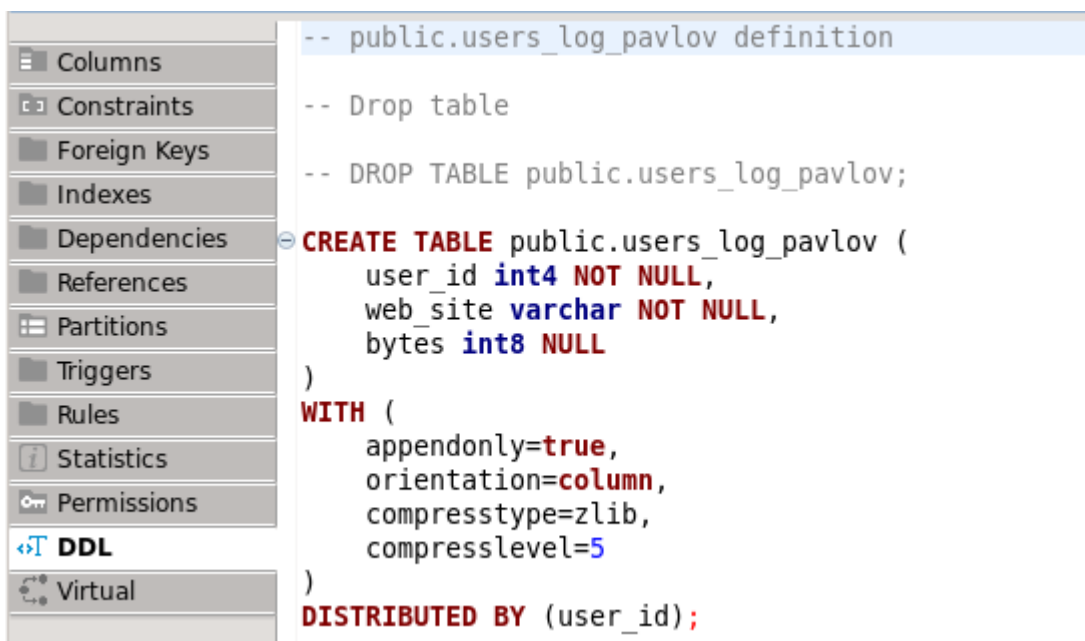
- с) Создайте Heap таблицу **users_pavlov** для хранения информации по пользователям. Определите уникальные ключи и CHECK CONSTRAINTS, которые по вашему мнению должны быть указаны на этой таблице.
- выберите столбец, по которому необходимо сделать distribution (возможно придется поменять модель данных)



The screenshot shows a database management tool interface. On the left is a sidebar with a tree view containing: Columns, Constraints, Foreign Keys, Indexes, Dependencies, References, Partitions, Triggers, Rules, Statistics, Permissions, DDL (selected), and Virtual. The main area displays SQL code for creating the `public.users_pavlov` table. The code includes comments for dropping the table and defining the table structure with columns `user_id` (serial4), `user_name` (varchar(50)), `ip_address` (varchar(15)), and `dns_name` (varchar), along with a primary key constraint and distribution by `user_id`.

```
-- public.users_pavlov definition
-- Drop table
-- DROP TABLE public.users_pavlov;
CREATE TABLE public.users_pavlov (
    user_id serial4 NOT NULL,
    user_name varchar(50) NOT NULL,
    ip_address varchar(15) NULL,
    dns_name varchar NULL,
    CONSTRAINT users_pavlov_pkey PRIMARY KEY (user_id)
)
DISTRIBUTED BY (user_id);
```

- d) Создайте Append-optimized таблицу `user_log_pavlov` для хранения посещения пользователем веб-ресурсов
- выберите любой уровень компрессии
 - выберите уровень хранения - колоночный
 - выберите столбец, по которому необходимо сделать distribution (возможно придется поменять модель данных)



The screenshot shows the same database management tool interface as above. The sidebar is identical, with 'DDL' selected. The main area displays SQL code for creating the `public.users_log_pavlov` table. The code includes comments for dropping the table and defining the table structure with columns `user_id` (int4), `web_site` (varchar), and `bytes` (int8). It also includes a `WITH` clause specifying `appendonly=true`, `orientation=column`, `compresstype=zlib`, and `compresslevel=5`, followed by a distribution clause by `user_id`.

```
-- public.users_log_pavlov definition
-- Drop table
-- DROP TABLE public.users_log_pavlov;
CREATE TABLE public.users_log_pavlov (
    user_id int4 NOT NULL,
    web_site varchar NOT NULL,
    bytes int8 NULL
)
WITH (
    appendonly=true,
    orientation=column,
    compresstype=zlib,
    compresslevel=5
)
DISTRIBUTED BY (user_id);
```

- е) Сгенерируйте данные в таблицу `users_pavlov` 10 произвольными пользователями, каждый из которых посетил около 10000 любых веб ресурсов (таблица `user_log_pavlov`).

Код генерации на github.com

- ф) Напишите SQL OLAP запрос, который вернет следующий срез данных (сделайте упорядочивание по 2 столбцу по убыванию и верните первые 10 строк)

адрес web_site	количество общих посещений	общее количество загруженных байт	количество уникальных пользователей
https://ekt.rt.ru/	100	32 767	57
...	

```
*<postgres> Script-1 x
SELECT WEB_SITE AS "АДРЕС WEB_SITE",
COUNT(*) AS "КОЛ-ВО ПОСЕЩЕНИЙ",
ROUND(SUM(BYTES)/1024/1024,1) AS "ОБЩЕЕ КОЛ-ВО ЗАГРУЖЕННЫХ МБ",
COUNT(DISTINCT USER_ID) AS "КОЛ-ВО УНИК.ПОЛЬЗ."
FROM USERS_LOG_PAVLOV
GROUP BY WEB_SITE
ORDER BY 2 DESC, 1
LIMIT 10
```

	ABC АДРЕС WEB_SITE ▼	123 КОЛ-ВО ПОСЕЩЕНИЙ ▼	123 ОБЩЕЕ КОЛ-ВО ЗАГРУЖЕННЫХ МВ ▼	123 КОЛ-ВО УНИК.ПОЛЬЗ. ▼
1	www.cnmazsudodtdbsnojhwjrxafnkhesnpqtrcfubjrquprqzvwmqsiggtezqfznoyle.mh	166	75,946.5	10
2	www.vcavrdvdb.qp	165	81,936.8	10
3	www.clcrrdvuyuhlgghlaefonwbfyzadbnvgjubifgjoyrmkvdehtmafkhkstmw.ib	163	74,008.7	10
4	www.cefbwdhnpqbpmhmczzhycjrrkuomjlmyafvdxird.uta	161	75,492.4	10
5	www.yjvwfhf.ki	161	71,235.2	10
6	www.ksgszqw.vj	160	72,608.4	10
7	www.oksewohmgrxpbdhxqyojabwspfizzmoeqkeilksfioxpqhilboyvtazqvg.zt	160	79,266.3	10
8	www.wzgajng.vu	160	73,217.6	10
9	www.gaoyido.zd	159	71,133.9	10
10	www.logtesxlvsemsgz.xk	159	72,104.3	10