

Задание на практику 1

Подготовка

1. Создать пользователя и базу данных, установить права для пользователя на БД
2. Создать таблицу с полями id(integer) + integer, text и char(4)/varchar(4)
3. Включить отображение времени выполнения запроса \timing on

Блок 1. Размер таблицы, зависимость размера таблицы от данных

1. Посмотреть размер таблицы/данных/индекса на момент создания. Предположить, почему размер ненулевой.
2. Заполнить таблицу случайными значениями. Посмотреть размеры данных и индекса. Что изменилось?
3. Удалить все данные и посмотреть размер таблицы. Что изменилось/не изменилось, почему?
4. Запустить команду очистки данных (vacuum), посмотреть размеры. Что изменилось/не изменилось, почему?
5. Запустить команду полной очистки. Что изменилось/не изменилось, почему?
6. Дополнительно: посмотреть, как меняется размер таблицы при частичном удалении данных, затем добавлении

Функции размеров:

pg_total_relation_size()	Общий объём, который занимает на диске заданная таблица, включая все индексы и данные TOAST
pg_table_size()	Объём, который занимает на диске данная таблица, за исключением индексов (но включая TOAST, карту свободного места и карту видимости)
pg_indexes_size()	Общий объём индексов, связанных с указанной таблицей
pg_relation_size()	Объём, который занимает на диске таблица или индекс
pg_size_pretty()	Функция, которая переводит байты в человекочитаемый формат

Подробнее о функциях получения размера объектов БД можно прочитать тут:
<https://postgrespro.ru/docs/postgresql/9.6/functions-admin#functions-admin-dboobject>

Пример: общий объём таблицы test в человекочитаемом формате
`select pg_size_pretty(pg_total_relation_size('test')) as total_size;`

Как добавить случайные значения?
`insert into test (
id, rand_num, rand_text, rand_code`

```
)  
select  
  i,  
  random()*10000+1,  
  md5(random()::text),  
  left(md5(random()::text), 4)  
from generate_series(1, 100000) i;
```

- random() - случайное число в диапазоне $0.0 \leq x < 1.0$
- md5(string) - Вычисляет MD5-хеш строки string и возвращает результат в 16-ричном виде
- left(str text, n int) - Возвращает первые n символов в строке
- generate_series(start, stop) - Выдаёт ряд целых чисел от start до stop с шагом 1
- generate_series(start, stop, step) - Выдаёт ряд значений от start до stop с заданным шагом (step)

Блок 2. Индексы, влияние на время добавления

1. Снова добавить случайные данные. Замерить время добавления.
2. Поочерёдно создавать индексы на все столбцы, посмотреть размер индексов каждого столбца. Какой индекс больше (если больше) и почему?
3. Удалить данные, но оставить индексы (любое количество, кроме первичного ключа)
4. Добавить такое же количество случайных данных. Как изменилось время добавления? Почему?

Блок 3 (дополнительный). Explain (analyze)

1. Выполнить любой запрос к таблице с припиской explain
explain select * from test;
2. Выполнить запрос с припиской explain (analyze)
3. Как вы думаете, в чём между ними разница?