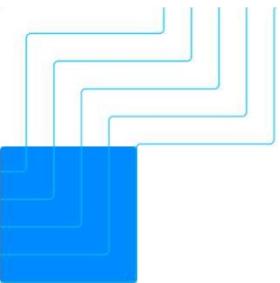




БЛОК 8. ИНДЕКСЫ

СТАТИСТИКА

begin



ЦЕЛЬ



01

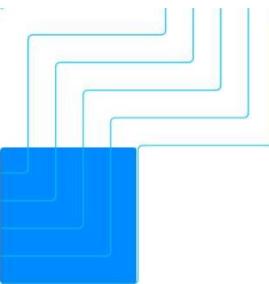
Понять задачи статистики

02

Узнать как использовать
статистику в Постгресе

03

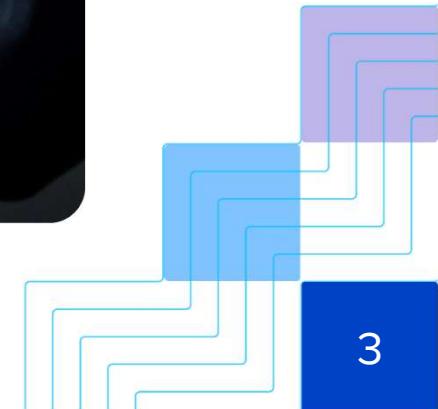
Научиться просматривать статистику



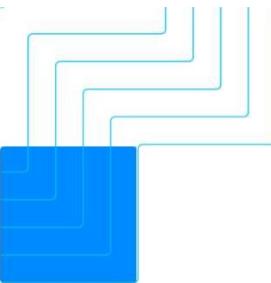
СОДЕРЖАНИЕ УРОКА



- 1 Статистика
- 2 Использование статистики
- 3 Практика



3



ДЛЯ ЧЕГО ИСПОЛЬЗУЕТСЯ СТАТИСТИКА

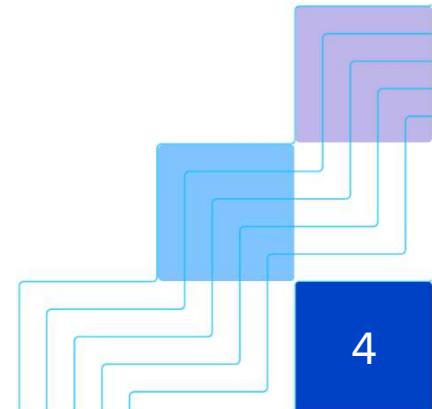
Начать урок я хотел бы с вопросов:

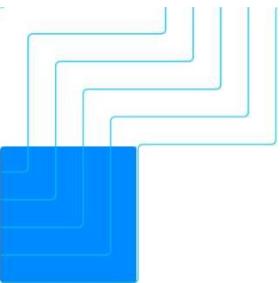
01 Как планировщик определяет что использовать: индекс или последовательное сканирование?

02 Сколько выделять памяти для результата запроса, ведь в одном случае может вернуться 2 записи, а при другом условии 200 000?

Если выделить **мало памяти**, результат не поместится и будет создана временная таблица на диске, а это медленно.

Если выделить **много памяти** - на небольшой результат бесполезно расходуется память...





ДЛЯ ЧЕГО ИСПОЛЬЗУЕТСЯ СТАТИСТИКА

Начать урок я хотел бы с вопросов:

01 Как планировщик определяет что использовать: индекс или последовательное сканирование?

02 Сколько выделять памяти для результата запроса, ведь в одном случае может вернуться 2 записи, а при другом условии 200 000?

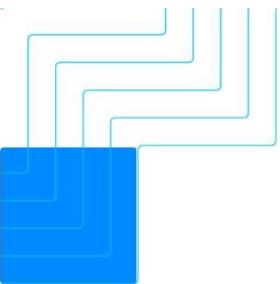
Если выделить **мало памяти**, результат не поместится и будет создана временная таблица на диске, а это медленно.

Если выделить **много памяти** - на небольшой результат бесполезно расходуется память...

Ответ - для этого всего используется статистика по данным в таблицах

Статистика содержит границы распределения данных в таблице, т.е., например, знает сколько городов у нас начинаются с букв А по В в поле город. И, следовательно, планировщик уже будет знать сколько выделять памяти





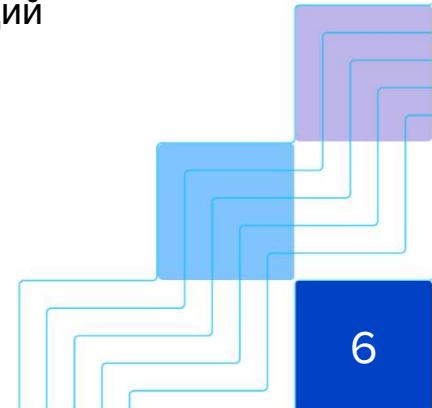
НАКЛАДНЫЕ РАСХОДЫ

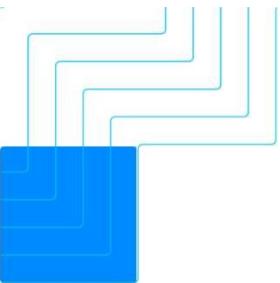


Поскольку сбор статистики несколько увеличивает накладные расходы (тратим больше ресурсов) при выполнении запроса, есть возможность настроить СУБД так, чтобы выполнять или не выполнять сбор статистической информации.

Список параметров Постгреса:

- **track_activities** включает мониторинг текущих команд, выполняемых любым серверным процессом
- **track_counts** определяет необходимость сбора статистики по обращениям к таблицам и индексам
- **track_functions** включает отслеживание использования пользовательских функций
- **track_io_timing** включает мониторинг времени чтения и записи блоков



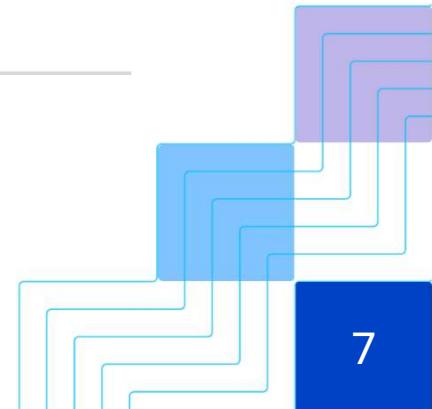


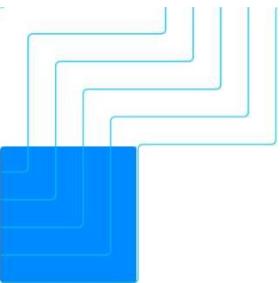
ОСНОВНЫЕ ТАБЛИЦЫ СТАТИСТИКИ



-
1. `pg_stat_database`
 2. `pg_stat_activity`
 3. `pg_stat_user_tables`
 4. `pg_stat_user_indexes`

Давайте разберем их более подробно





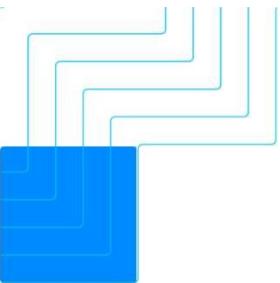
PG_STAT_DATABASE



01

Отсюда можем получить следующую информацию, которую мы анализируем для оптимизации наших запросов:

- Как много информации получаем из кэша
- Как часто бывают проблемы с транзакциями



PG_STAT_ACTIVITY

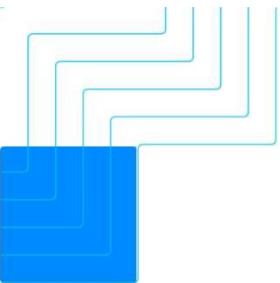


02

Содержит в себе по одной строке на каждый процесс сервера.



Напомню, что любое подключение к серверу порождает форк основного процесса и также отображается в этом представлении.

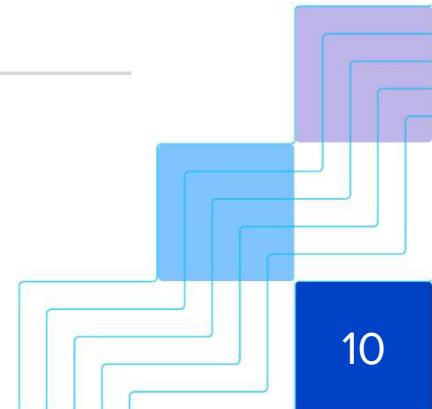


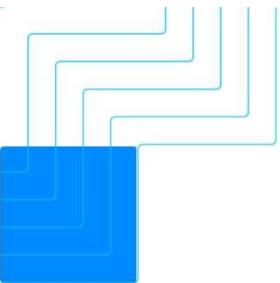
PG_STAT_USER_TABLES



03

Здесь собрана статистика, что вообще у нас происходит с таблицей.
Сколько строк, мёртвых строк и т.д.



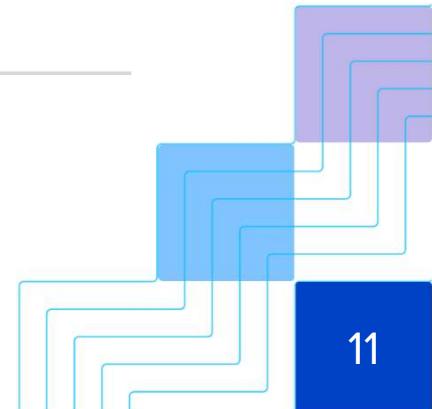


PG_STAT_USER_INDEXES



04

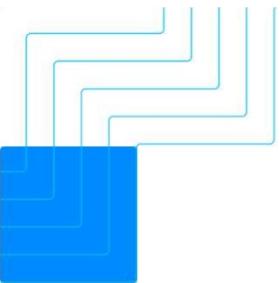
Здесь собирается статистика по используемым индексам.





ПОДВЕДЕНИЕ ИТОГОВ

&



Итоги занятия



01



Поняли задачи
статистики

02

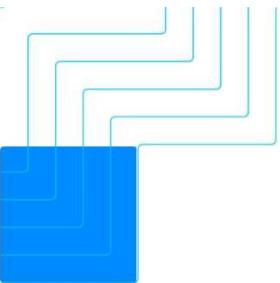


Узнали как использовать
статистику в Постгресе

03



Научились просматривать статистику



Задание для самопроверки



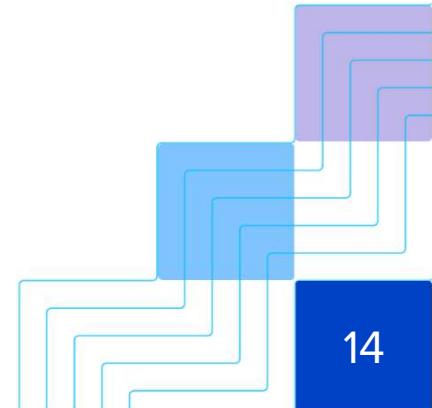
Цель задания:

Посмотреть статистику по вновь созданной таблице

Пошаговый план выполнения:

1. Создать таблицу из 100 000 строк
2. Сделать несколько запросов
3. Проанализировать таблицу
4. Посмотреть статистику по таблице
5. Выполните задание самостоятельно. Свое решение вы можете сравнить с эталонным (приложено к уроку)

Задание закончено



СПАСИБО!

На следующем занятии мы рассмотрим тему:

- Обслуживание индексов

end