

# Многоверсионность Заморозка



## Авторские права

© Postgres Professional, 2016–2025

Авторы: Егор Рогов, Павел Лузанов, Илья Баштанов, Игорь Гнатюк

Фото: Олег Бартунов (монастырь Пху и пик Бхрикути, Непал)

## Использование материалов курса

Некоммерческое использование материалов курса (презентации, демонстрации) разрешается без ограничений. Коммерческое использование возможно только с письменного разрешения компании Postgres Professional. Запрещается внесение изменений в материалы курса.

## Обратная связь

Отзывы, замечания и предложения направляйте по адресу:

[edu@postgrespro.ru](mailto:edu@postgrespro.ru)

## Отказ от ответственности

Компания Postgres Professional не несет никакой ответственности за любые повреждения и убытки, включая потерю дохода, нанесенные прямым или косвенным, специальным или случайным использованием материалов курса. Компания Postgres Professional не предоставляет каких-либо гарантий на материалы курса. Материалы курса предоставляются на основе принципа «как есть» и компания Postgres Professional не обязана предоставлять сопровождение, поддержку, обновления, расширения и изменения.

Проблема переполнения счетчика транзакций

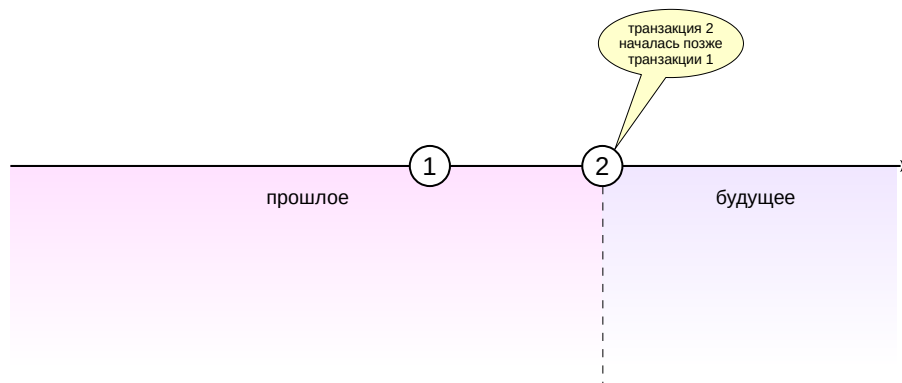
Заморозка версий строк и правила видимости

Настройка автоочистки для выполнения заморозки

Заморозка вручную

# Переполнение счетчика

меньшие номера — прошлое, бóльшие — будущее  
разрядность счетчика — 32 бита, что делать при переполнении?



3

Кроме освобождения места в страницах, очистка выполняет также задачу по предотвращению проблем, связанных с переполнением счетчика транзакций.

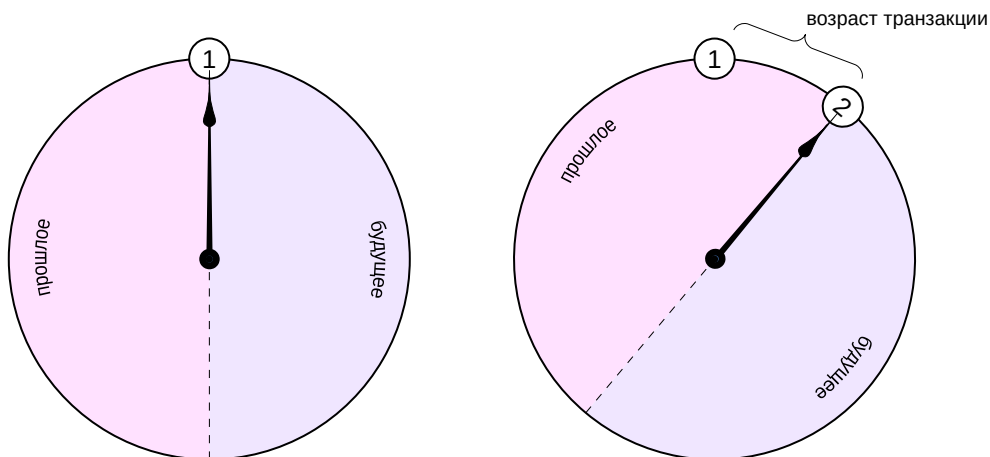
Под номер транзакции в PostgreSQL выделено 32 бита. Это довольно большое число (около 4 млрд номеров), но при активной работе сервера оно вполне может быть исчерпано. Например при нагрузке 1000 транзакций в секунду это произойдет всего через полтора месяца непрерывной работы.

Но мы говорили о том, что механизм многоверсионности полагается на последовательную нумерацию транзакций — из двух транзакций транзакция с меньшим номером считается начавшейся раньше. Понятно, что нельзя просто обнулить счетчик и продолжить нумерацию заново.

Почему под номер транзакции не выделено 64 бита — ведь это полностью исключило бы проблему? Дело в том, что (как рассматривалось в теме «Страницы и версии строк») в заголовке каждой версии строки хранятся два номера транзакций — xmin и xmax. Заголовок и так достаточно большой, а увеличение разрядности привело бы к его увеличению еще на 8 байт.

# Нумерация по кругу

пространство номеров транзакций закольцовано  
половина номеров — прошлое, половина — будущее

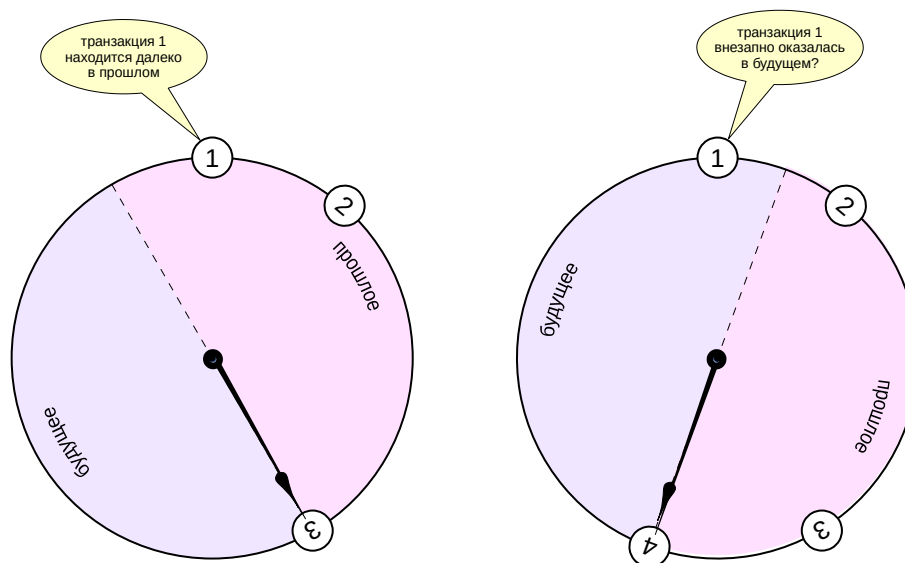


4

Поэтому вместо линейной схемы все номера транзакций закольцованы. Для любой транзакции половина номеров «против часовой стрелки» считается принадлежащей прошлому, а половина «по часовой стрелке» — будущему.

*Возрастом транзакции* называется число транзакций, прошедших с момента ее появления в системе (независимо от того, переходил ли счетчик через ноль или нет).

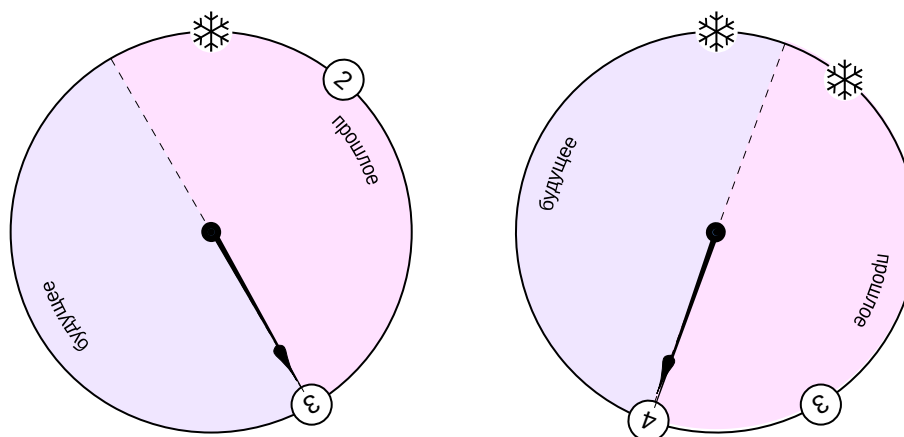
# Проблема видимости



В такой закольцованной схеме возникает неприятная ситуация. Транзакция, находившаяся в далеком прошлом (транзакция 1 на слайде), через некоторое время окажется в той половине круга, которая относится к будущему. Это, конечно, нарушит правила видимости и приведет к проблемам.

# Заморозка версий строк

замороженные версии строк считаются «бесконечно старыми»  
номер транзакции xmin может быть использован заново



6

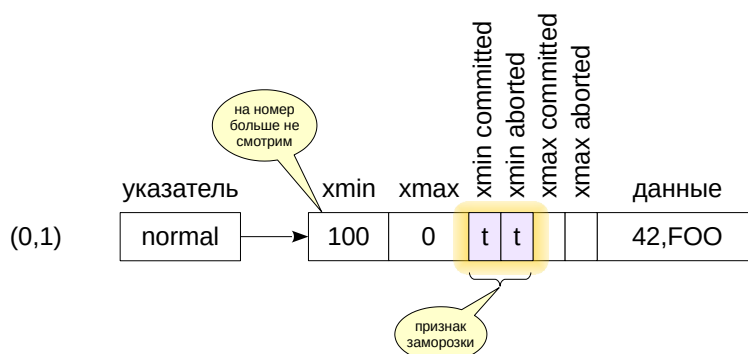
Чтобы не допустить путешествий из прошлого в будущее, процесс очистки выполняет еще одну задачу. Он находит достаточно старые и «холодные» версии строк (которые видны во всех снимках и изменение которых уже маловероятно) и специальным образом помечает — «замораживает» — их. Замороженная версия строки считается старше любых обычных данных и всегда видна во всех снимках данных. При этом уже не требуется смотреть на номер транзакции xmin, и этот номер может быть безопасно использован заново. Таким образом, замороженные версии строк всегда остаются в прошлом.

<https://postgrespro.ru/docs/postgresql/16/routine-vacuuming#VACUUM-FOR-WRAPAROUND>

# Заморозка версий строк

## Еще одна задача процесса очистки

если вовремя не заморозить версии строк, они окажутся в будущем и сервер остановится для предотвращения ошибки



7

Для того чтобы пометить версию строки как замороженную, для транзакции `xmin` выставляются одновременно оба бита-подсказки — бит фиксации и бит отмены.

Заметим, что транзакцию `xmax` замораживать не нужно. Ее наличие означает, что данная версия строки больше не актуальна. После того, как она перестанет быть видимой в снимках данных, такая версия строки будет очищена.

Многие источники (включая документацию) упоминают специальный номер `FrozenTransactionId = 2`, который записывается на место `xmin` в замороженных версиях. Такая система действовала до версии 9.4, но сейчас заменена на биты-подсказки — это позволяет сохранить в версии строки исходный номер транзакции, что удобно для целей поддержки и отладки. Однако транзакции с номером 2 еще могут встретиться в старых системах, даже обновленных до последних версий.

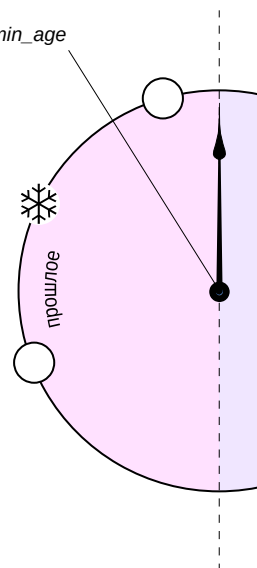
Важно, чтобы версии строк замораживались вовремя. Если возникнет ситуация, при которой еще не замороженный номер транзакции рискует попасть в будущее, PostgreSQL аварийно остановится. Это возможно в двух случаях: либо транзакция не завершена и, следовательно, не может быть заморожена, либо не сработала очистка.

При запуске сервера транзакция будет автоматически отменена; дальше администратор должен вручную выполнить очистку, и после этого система сможет продолжить работу.

***vacuum\_freeze\_min\_age***

минимальный возраст,  
с которого начинается заморозка

*vacuum\_freeze\_min\_age*



8

Заморозкой управляют четыре основных параметра.

Параметр ***vacuum\_freeze\_min\_age*** определяет минимальный возраст транзакции *xmin*, с которого начинается заморозка.

Чем меньше это значение, тем больше может быть накладных расходов. Если строка «горячая» и активно меняется, заморозка ее версий будет пропадать без пользы: уже замороженные версии будут вычищаться, а новые версии придется снова замораживать.

Поэтому более молодые версии строк замораживаются только в тех случаях, когда это точно не добавляет работы, например, если в странице уже требуется заморозка других (более старых) строк или при полной очистке таблицы.

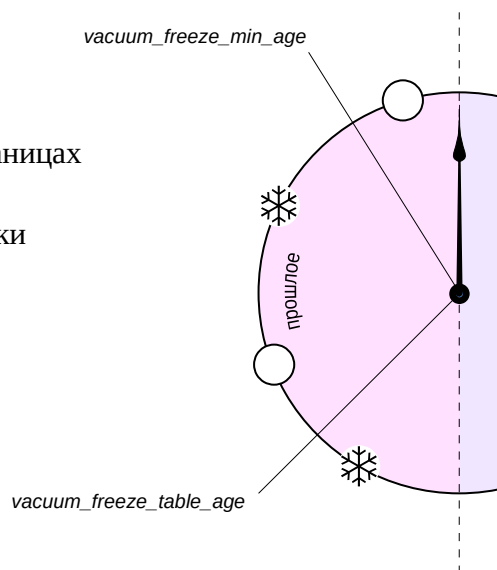
Заметим, что очистка просматривает только страницы, не отмеченные в карте видимости. Если на странице остались только актуальные версии, то очистка не придет в такую страницу и не заморозит их.

В заголовке табличной страницы также имеется признак видимости всех версий строк в ней; очистка использует его вместе с соответствующей отметкой в карте видимости.



## `vacuum_freeze_table_age`

при достижении такого возраста  
замораживаются версии строк на всех страницах  
(«агрессивная» заморозка)  
для ускорения используется карта заморозки



9

Параметр **`vacuum_freeze_table_age`** определяет возраст транзакции, при котором пора выполнять заморозку версий строк на всех страницах таблицы. Такая заморозка называется «агрессивной».

Для каждой таблицы хранится номер транзакции (`pg_class.relrozenxid`), для которого известно, что в версиях строк не осталось более старых незамороженных номеров транзакций. Возраст этой транзакции и сравнивается со значением параметра.

Чтобы не просматривать всю таблицу целиком, вместе с картой видимости ведется *карта заморозки*. В ней отмечены страницы, в которых заморожены все версии строк. Такие страницы при заморозке можно пропускать.

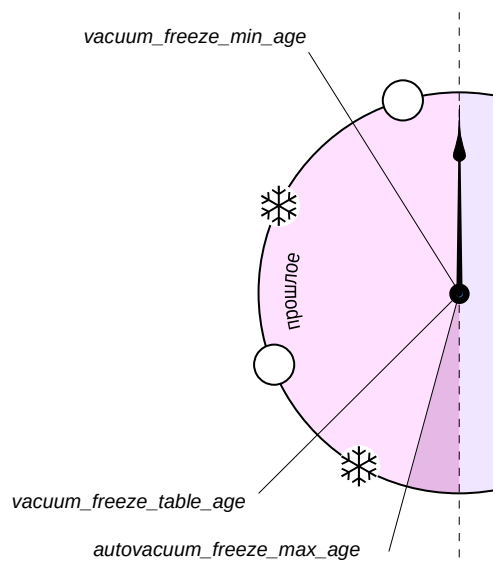
Даже в агрессивном режиме все версии строк с транзакциями младше `vacuum_freeze_min_age` не замораживаются, поэтому после заморозки новый возраст транзакции `relrozenxid` будет равен не нулю, а `vacuum_freeze_min_age`. Таким образом, заморозка всех страниц выполняется раз в  $(vacuum\_freeze\_table\_age - vacuum\_freeze\_min\_age)$  транзакций.

Мы уже говорили, что слишком маленькое значение параметра `vacuum_freeze_min_age` увеличивает накладные расходы на очистку. Но при больших значениях агрессивная заморозка будет выполняться слишком часто, что тоже плохо. Установка этого параметра требует компромисса.

## *autovacuum\_freeze\_max\_age*

при достижении такого возраста  
заморозка запускается принудительно  
определяет размер CLOG

VACUUM (index\_cleanup off)



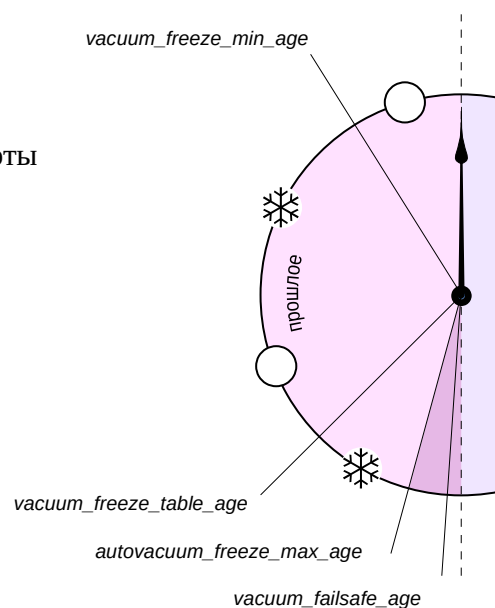
Параметр ***autovacuum\_freeze\_max\_age*** определяет возраст транзакции, при котором заморозка будет выполняться принудительно. Автоочистка для предотвращения последствий переполнения счетчика транзакций запустится, даже если она отключена параметрами.

Этот параметр также определяет размер структуры CLOG: данные о статусе более старых транзакций точно никогда не понадобятся, поэтому часть файлов из PGDATA/pg\_xact может быть удалена.

Если администратор понимает, что автоочистка не успеет заморозить версии строк до переполнения счетчика транзакций, можно воспользоваться ручной очисткой с параметром `index_cleanup off`. В этом случае индексы не будут очищаться, но за счет этого версии строк в таблицах будут заморожены быстрее.

*vacuum\_failsafe\_age*

при достижении такого возраста  
очистка переходит в защитный режим работы




11

Параметр ***vacuum\_failsafe\_age*** управляет включением защитного режима работы очистки, который служит для ускорения заморозки номеров транзакций.

В этом режиме будут отменены регламентные задержки `autovacuum_vacuum_cost_delay` и `vacuum_cost_delay`. Также не будут выполняться некоторые необязательные работы (например очистка индексов). Такие меры позволяют очистке быстрее заморозить старые транзакции и перейти в обычный режим работы.

## Конфигурационные параметры

<code>vacuum_freeze_min_age</code>	=	50 000 000	
<code>vacuum_freeze_table_age</code>	=	150 000 000	↑?
 <code>autovacuum_freeze_max_age</code>	=	200 000 000	
<code>vacuum_failsafe_age</code>	=	1 600 000 000	

## Параметры хранения таблиц

<code>autovacuum_freeze_min_age</code>
<code>toast.autovacuum_freeze_min_age</code>
<code>autovacuum_freeze_table_age</code>
<code>toast.autovacuum_freeze_table_age</code>
<code>autovacuum_freeze_max_age</code>
<code>toast.autovacuum_freeze_max_age</code>

12

Значения по умолчанию довольно консервативны. Предел для `autovacuum_freeze_max_age` – порядка 2 млрд транзакций, а используется значение в 10 раз меньшее. Можно увеличить значения `vacuum_freeze_table_age` и `autovacuum_freeze_max_age` для уменьшения накладных расходов, но важно понимать, что если по каким-то причинам (например, из-за незавершенной транзакции) автоочистка вовремя не справится с заморозкой, у администратора останется мало времени для принятия мер. Заметьте, что изменение параметра `autovacuum_freeze_max_age` требует перезапуска сервера.

Значение по умолчанию `vacuum_failsafe_age` значительно больше, чем `autovacuum_freeze_max_age`, и если до исчерпания номеров останется мало времени, защитный режим ускорит заморозку.

Отметим, что сервер может скорректировать установленные значения параметров `vacuum_freeze_min_age`, `vacuum_freeze_table_age` и `vacuum_failsafe_age` исходя из значения `autovacuum_freeze_max_age`.

Ряд параметров также можно устанавливать на уровне отдельных таблиц с помощью параметров хранения. Это имеет смысл делать только в особенных случаях, когда таблица действительно требует особого обхождения. Заметьте, что имена параметров на уровне таблиц немного отличаются от имен конфигурационных параметров.

В модуле «Блокировки» рассматриваются т. н. мультитранзакции и дополнительные параметры настройки заморозки для них.

<https://www.postgresql.org/docs/16/runtime-config-client.html>

<https://www.postgresql.org/docs/16/runtime-config-autovacuum.html>

## Заморозка

Установим для демонстрации параметры заморозки.

Небольшой возраст транзакции:

```
=> ALTER SYSTEM SET vacuum_freeze_min_age = 1;
```

ALTER SYSTEM

Возраст, после которого будет выполняться заморозка всех страниц:

```
=> ALTER SYSTEM SET vacuum_freeze_table_age = 3;
```

ALTER SYSTEM

И отключим автоматическую очистку, чтобы запускать ее вручную в нужный момент.

```
=> ALTER SYSTEM SET autovacuum = off;
```

ALTER SYSTEM

```
=> SELECT pg_reload_conf();
```

```
pg_reload_conf
-----
t
(1 row)
```

Создадим таблицу с данными. Установим минимальный fillfactor: на каждой странице будет всего две строки.

```
=> CREATE DATABASE mvcc_freeze;
```

CREATE DATABASE

```
=> \c mvcc_freeze
```

You are now connected to database "mvcc\_freeze" as user "student".

```
=> CREATE TABLE t(id integer, s char(300)) WITH (fillfactor = 10);
```

CREATE TABLE

Создадим представление для наблюдения за битами-подсказками на первых двух страницах таблицы.

Сейчас нас интересует только xmin и биты, которые относятся к нему, поскольку версии строк с ненулевым xmax будут очищены. Кроме того, выведем и возраст транзакции xmin.

```
=> CREATE EXTENSION pageinspect;
```

CREATE EXTENSION

```
=> CREATE VIEW t_v AS
SELECT '('||blkno||','||lp||')' as ctid,
       CASE lp_flags
         WHEN 0 THEN 'unused'
         WHEN 1 THEN 'normal'
         WHEN 2 THEN 'redirect to '||lp_off
         WHEN 3 THEN 'dead'
       END AS state,
       t_xmin AS xmin,
       age(t_xmin) AS xmin_age,
       CASE WHEN (t_infomask & 256) > 0 THEN 't' END AS xmin_c,
       CASE WHEN (t_infomask & 512) > 0 THEN 't' END AS xmin_a,
       t_xmax AS xmax
FROM (
  SELECT 0 blkno, * FROM heap_page_items(get_raw_page('t',0))
  UNION ALL
  SELECT 1 blkno, * FROM heap_page_items(get_raw_page('t',1))
) q
ORDER BY blkno, lp;
```

CREATE VIEW

Для того чтобы заглянуть в карту видимости и заморозки, воспользуемся еще одним расширением:

```
=> CREATE EXTENSION pg_visibility;
```

CREATE EXTENSION

Вставляем данные. Сразу выполним очистку, чтобы заполнить карту видимости.

```
=> INSERT INTO t(id, s) SELECT g.id, 'FOO' FROM generate_series(1,100) g(id);
```

```
INSERT 0 100
```

```
=> VACUUM t;
```

```
VACUUM
```

После очистки обе страницы отмечены в карте видимости (all\_visible):

```
=> SELECT * FROM generate_series(0,1) g(blkno), pg_visibility_map('t',g.blkno)
ORDER BY g.blkno;
```

blkno	all_visible	all_frozen
0	t	f
1	t	f

(2 rows)

Каков возраст транзакции, создавшей строки?

```
=> SELECT * FROM t_v;
```

ctid	state	xmin	xmin_age	xmin_c	xmin_a	xmax
(0,1)	normal	748	1	t		0
(0,2)	normal	748	1	t		0
(1,1)	normal	748	1	t		0
(1,2)	normal	748	1	t		0

(4 rows)

Возраст равен 1; версии строк с такой транзакцией еще не будут заморожены.

Обновим строку на нулевой странице. Новая версия попадет на ту же страницу благодаря небольшому значению fillfactor.

```
=> UPDATE t SET s = 'BAR' WHERE id = 1;
```

```
UPDATE 1
```

```
=> SELECT * FROM t_v;
```

ctid	state	xmin	xmin_age	xmin_c	xmin_a	xmax
(0,1)	normal	748	2	t		749
(0,2)	normal	748	2	t		0
(0,3)	normal	749	1			0
(1,1)	normal	748	2	t		0
(1,2)	normal	748	2	t		0

(5 rows)

Сейчас нулевая страница уже будет обработана заморозкой:

- возраст транзакции превышает значение, установленное в vacuum\_freeze\_min\_age;
- страница изменена и исключена из карты видимости.

```
=> SELECT * FROM generate_series(0,1) g(blkno), pg_visibility_map('t',g.blkno)
ORDER BY g.blkno;
```

blkno	all_visible	all_frozen
0	f	f
1	t	f

(2 rows)

Выполняем очистку.

```
=> VACUUM t;
```

```
VACUUM
```

Очистка обработала измененную страницу. У одной версии строки установлены оба бита — это признак заморозки. Другая версия строки слишком молода, однако тоже была заморожена при проходе страницы (это позволило отметить нулевую страницу в карте заморозки):

```
=> SELECT * FROM t_v;
```

ctid	state	xmin	xmin_age	xmin_c	xmin_a	xmax
(0,1)	redirect to 3					
(0,2)	normal	748	2	t	t	0
(0,3)	normal	749	1	t	t	0
(1,1)	normal	748	2	t		0
(1,2)	normal	748	2	t		0

(5 rows)

Теперь обе страницы отмечены в карте видимости (все версии строк на них актуальны). Очистка теперь не будет обрабатывать ни одну из этих страниц, и незамороженные версии строк на первой странице так и останутся незамороженными.

```
=> SELECT * FROM generate_series(0,1) g(blkno), pg_visibility_map('t',g.blkno)
ORDER BY g.blkno;
```

blkno	all_visible	all_frozen
0	t	t
1	t	f

(2 rows)

Именно для такого случая и требуется параметр `vacuum_freeze_table_age`, определяющий, в какой момент нужно просмотреть страницы, отмеченные в карте видимости, если они не отмечены в карте заморозки.

Для каждой таблицы сохраняется наибольший номер транзакции, для которого все версии строк с меньшими номерами `xmin` гарантированно заморожены. Ее возраст и сравнивается со значением параметра.

```
=> SELECT relfrozenxid, age(relfrozenxid) FROM pg_class WHERE relname = 't';
```

relfrozenxid	age
748	2

(1 row)

Сымитируем выполнение еще одной транзакции, чтобы возраст `relfrozenxid` таблицы достиг значения параметра `vacuum_freeze_table_age`.

```
=> SELECT pg_current_xact_id();
```

pg_current_xact_id
750

(1 row)

```
=> SELECT relfrozenxid, age(relfrozenxid) FROM pg_class WHERE relname = 't';
```

relfrozenxid	age
748	3

(1 row)

```
=> VACUUM t;
```

VACUUM

Теперь, поскольку гарантированно была проверена вся таблица, номер замороженной транзакции можно увеличить — мы уверены, что в страницах не осталось более старой незамороженной транзакции.

```
=> SELECT relfrozenxid, age(relfrozenxid) FROM pg_class WHERE relname = 't';
```

relfrozenxid	age
751	0

(1 row)

Вот что получилось в страницах:

```
=> SELECT * FROM t_v;
```

ctid	state	xmin	xmin_age	xmin_c	xmin_a	xmax
(0,1)	redirect to 3					
(0,2)	normal	748	3	t	t	0
(0,3)	normal	749	2	t	t	0
(1,1)	normal	748	3	t	t	0
(1,2)	normal	748	3	t	t	0

(5 rows)

Обе страницы теперь отмечены в карте заморозки.

```
=> SELECT * FROM generate_series(0,1) g(blkno), pg_visibility_map('t',g.blkno)
ORDER BY g.blkno;
```

blkno	all_visible	all_frozen
0	t	t
1	t	t

(2 rows)

Номер последней замороженной транзакции есть и на уровне всей БД:

```
=> SELECT datname, datfrozenxid, age(datfrozenxid)
FROM pg_database;
```

datname	datfrozenxid	age
postgres	722	29
student	722	29
template1	722	29
template0	722	29
mvcc_freeze	722	29

(5 rows)

Он устанавливается в минимальное значение из relfrozenxid всех таблиц этой БД. Если возраст datfrozenxid превысит значение параметра autovacuum\_freeze\_max\_age, автоочистка будет запущена принудительно.



## VACUUM

заморозка версий строк по возрасту в соответствии с настройками

## VACUUM FREEZE

принудительная заморозка версий строк с xmin любого возраста

тот же эффект и при VACUUM FULL, CLUSTER

## COPY ... WITH FREEZE

принудительная заморозка сразу после загрузки

таблица должна быть создана или опустошена в той же транзакции

могут нарушиться правила изоляции транзакции

Несмотря на то, что во время работы автоочистки при необходимости выполняется и заморозка, иногда бывает удобно управлять заморозкой вручную.

Команда VACUUM, как и автоочистка, выполнит заморозку в соответствии с настройками.

Если выполнить команду VACUUM FREEZE, будут заморожены все версии строк без оглядки на возраст транзакций (как будто параметры *vacuum\_freeze\_min\_age* и *vacuum\_freeze\_table\_age* равны нулю).

При перестройке таблицы командами VACUUM FULL или CLUSTER все строки также замораживаются.

<https://postgrespro.ru/docs/postgresql/16/sql-vacuum>

Данные можно заморозить и при начальной загрузке с помощью команды COPY, указав параметр FREEZE. Для этого таблица должна быть создана (или опустошена командой TRUNCATE) в той же транзакции, что и COPY. Поскольку для замороженных строк действуют отдельные правила видимости, такие строки будут видны в снимках данных других транзакций в нарушение обычных правил изоляции (для транзакций с уровнем Repeatable Read или Serializable), но обычно это не представляет проблемы. Подробнее такой случай рассматривается в практике.

<https://postgrespro.ru/docs/postgresql/16/sql-copy>

Текущая реализация такой заморозки для таблиц с TOAST полноценно обрабатывает только основную часть таблицы, а в карте видимости признак видимости всех строк не проставляется. А это означает дополнительный проход по всем страницам при последующей очистке.

Пространство номеров транзакций закольцовано  
Достаточно старые версии строк замораживаются  
процессом очистки  
Для оптимизации используется карта заморозки

1. Проверьте с помощью расширения `pageinspect`, что при использовании команды `COPY ... WITH FREEZE` версии строк действительно замораживаются.
2. Убедитесь, что даже на уровне изоляции `Repeatable Read` строки, загруженные командой `COPY ... WITH FREEZE`, оказываются видны в снимке данных.
3. Уменьшив значение параметра `autovacuum_freeze_max_age` и отключив автоочистку, воспроизведите ситуацию принудительного срабатывания автоочистки, выполнив соответствующее количество транзакций. Учтите, что срабатывание произойдет не сразу, а при выполнении ручной очистки какой-нибудь таблицы (или при перезапуске сервера).

16

3. Чтобы транзакциям выделялись настоящие (не виртуальные) номера, в транзакции нужно менять данные.

Можно организовать цикл в `bash`, в котором вызывать `psql` с командой обновления:

```
psql -c 'UPDATE ...'
```

Другой вариант — использовать для организации цикла `PL/pgSQL`.

Для этого можно создать процедуру, выполняющую фиксацию транзакции или использовать блок с обработкой исключений.

При перехвате исключения транзакция будет откатываться к неявной точке сохранения: фактически начнется новая вложенная транзакция (см. тему «Страницы и версии строк»).

Третий вариант — использовать утилиту `pgbench`:

<https://postgrespro.ru/docs/postgresql/16/pgbench>

## 1. Заморозка при COPY WITH FREEZE

Создаем таблицу и загружаем несколько строк в одной и той же транзакции:

```
=> CREATE DATABASE mvcc_freeze;
```

```
CREATE DATABASE
```

```
=> \c mvcc_freeze
```

```
You are now connected to database "mvcc_freeze" as user "student".
```

```
=> BEGIN;
```

```
BEGIN
```

```
=> CREATE TABLE t(n integer);
```

```
CREATE TABLE
```

```
=> COPY t FROM stdin WITH FREEZE;
```

```
=> 1
```

```
=> 2
```

```
=> 3
```

```
=> \.
```

```
COPY 3
```

```
=> COMMIT;
```

```
COMMIT
```

Проверяем версии строк:

```
=> CREATE EXTENSION pageinspect;
```

```
CREATE EXTENSION
```

```
=> CREATE VIEW t_v AS
```

```
SELECT '(0, ' || lp || ' )' as ctid,
       CASE lp_flags
         WHEN 0 THEN 'unused'
         WHEN 1 THEN 'normal'
         WHEN 2 THEN 'redirect to ' || lp_off
         WHEN 3 THEN 'dead'
       END AS state,
       t_xmin AS xmin,
       age(t_xmin) AS xmin_age,
       CASE WHEN (t_infomask & 256) > 0 THEN 't' END AS xmin_c,
       CASE WHEN (t_infomask & 512) > 0 THEN 't' END AS xmin_a,
       t_xmax AS xmax,
       t_ctid
```

```
FROM heap_page_items(get_raw_page('t',0))
```

```
ORDER BY lp;
```

```
CREATE VIEW
```

```
=> SELECT * FROM t_v;
```

ctid	state	xmin	xmin_age	xmin_c	xmin_a	xmax	t_ctid
(0,1)	normal	744	3	t	t	0	(0,1)
(0,2)	normal	744	3	t	t	0	(0,2)
(0,3)	normal	744	3	t	t	0	(0,3)

(3 rows)

## 2. COPY WITH FREEZE и изоляция

В другом сеансе начнем транзакцию с уровнем изоляции Repeatable Read.

```
| => \c mvcc_freeze
```

```
| You are now connected to database "mvcc_freeze" as user "student".
```

```
| => BEGIN ISOLATION LEVEL REPEATABLE READ;
```

```

| BEGIN
|
| => SELECT pg_current_xact_id();
|
|      pg_current_xact_id
|      -----
|                  747
|
| (1 row)

```

Обратите внимание, что эта транзакция не должна обращаться к таблице t.

Теперь опустошим таблицу и загрузим в нее новые строки в одной транзакции. Если бы параллельная транзакция прочитала содержимое t, команда TRUNCATE ожидала бы ее завершения.

```

=> BEGIN;

BEGIN

=> TRUNCATE t;

TRUNCATE TABLE

=> COPY t FROM stdin WITH FREEZE;

=> 10
=> 20
=> 30
=> \.

COPY 3

=> COMMIT;

COMMIT

```

Теперь параллельная транзакция видит новые данные, хотя это и нарушает изоляцию:

```

| => SELECT * FROM t;
|
|      n
|      ----
|      10
|      20
|      30
| (3 rows)
|
| => COMMIT;
|
| COMMIT
|
| => \q

```

### 3. Аварийное срабатывание автоочистки

Предварительно заморозим все транзакции во всех базах. Для этого удобно воспользоваться командой vacuumdb:

```

student$ vacuumdb --all --freeze

vacuumdb: vacuuming database "mvcc_freeze"
vacuumdb: vacuuming database "postgres"
vacuumdb: vacuuming database "student"
vacuumdb: vacuuming database "template1"

```

Максимальный возраст незамороженных транзакций по всем БД:

```

=> SELECT datname, datfrozenxid, age(datfrozenxid) FROM pg_database;

   datname   | datfrozenxid | age
-----+-----+----
 postgres   |          749 |  0
 student    |          749 |  0
 template1  |          749 |  0
 template0  |          722 | 27
 mvcc_freeze |          749 |  0
(5 rows)

```

Отключаем автоочистку.

```

=> ALTER SYSTEM SET autovacuum = off;

```

```
ALTER SYSTEM
```

Уменьшаем значения параметров:

```
=> ALTER SYSTEM SET vacuum_freeze_min_age = 1_000;
```

```
ALTER SYSTEM
```

```
=> ALTER SYSTEM SET vacuum_freeze_table_age = 10_000;
```

```
ALTER SYSTEM
```

```
=> ALTER SYSTEM SET autovacuum_freeze_max_age = 100_000; # минимальное значение
```

```
ALTER SYSTEM
```

Требуется перезагрузка сервера.

```
student$ sudo pg_ctlcluster 16 main restart
```

```
student$ psql mvcc_freeze
```

Получить большое количество транзакций можно разными способами; например, можно воспользоваться утилитой `pgbench`. Попросим ее инициализировать свои таблицы и выполнить 100000 транзакций.

```
student$ pgbench -i mvcc_freeze
```

```
dropping old tables...
NOTICE: table "pgbench_accounts" does not exist, skipping
NOTICE: table "pgbench_branches" does not exist, skipping
NOTICE: table "pgbench_history" does not exist, skipping
NOTICE: table "pgbench_tellers" does not exist, skipping
creating tables...
generating data (client-side)...
100000 of 100000 tuples (100%) done (elapsed 0.07 s, remaining 0.00 s)
vacuuming...
creating primary keys...
done in 0.41 s (drop tables 0.00 s, create tables 0.03 s, client-side generate 0.17 s,
vacuum 0.07 s, primary keys 0.13 s).
```

Выполнение ста тысяч транзакций может занять заметное время, ключ `--protocol=prepared` немного ускоряет работу за счет использования подготовленных операторов:

```
student$ pgbench -t 100000 -P 5 --protocol=prepared mvcc_freeze
```

```
pgbench (16.3 (Ubuntu 16.3-1.pgdg22.04+1))
starting vacuum...end.
progress: 5.0 s, 157.4 tps, lat 6.341 ms stddev 1.503, 0 failed
progress: 10.0 s, 145.2 tps, lat 6.886 ms stddev 1.808, 0 failed
progress: 15.0 s, 170.2 tps, lat 5.879 ms stddev 0.981, 0 failed
progress: 20.0 s, 171.6 tps, lat 5.820 ms stddev 0.781, 0 failed
progress: 25.0 s, 156.2 tps, lat 6.406 ms stddev 1.667, 0 failed
progress: 30.0 s, 170.0 tps, lat 5.879 ms stddev 1.002, 0 failed
progress: 35.0 s, 171.2 tps, lat 5.835 ms stddev 0.639, 0 failed
progress: 40.0 s, 167.8 tps, lat 5.962 ms stddev 0.756, 0 failed
progress: 45.0 s, 167.6 tps, lat 5.963 ms stddev 0.682, 0 failed
progress: 50.0 s, 152.8 tps, lat 6.546 ms stddev 1.727, 0 failed
progress: 55.0 s, 167.2 tps, lat 5.976 ms stddev 0.715, 0 failed
progress: 60.0 s, 165.4 tps, lat 6.042 ms stddev 0.841, 0 failed
progress: 65.0 s, 160.4 tps, lat 6.234 ms stddev 0.754, 0 failed
progress: 70.0 s, 146.6 tps, lat 6.822 ms stddev 1.538, 0 failed
progress: 75.0 s, 147.6 tps, lat 6.769 ms stddev 1.910, 0 failed
progress: 80.0 s, 163.0 tps, lat 6.131 ms stddev 0.943, 0 failed
progress: 85.0 s, 158.2 tps, lat 6.323 ms stddev 0.819, 0 failed
progress: 90.0 s, 159.4 tps, lat 6.275 ms stddev 0.905, 0 failed
progress: 95.0 s, 163.4 tps, lat 6.114 ms stddev 0.882, 0 failed
progress: 100.0 s, 160.6 tps, lat 6.226 ms stddev 1.049, 0 failed
progress: 105.0 s, 129.8 tps, lat 7.703 ms stddev 2.231, 0 failed
progress: 110.0 s, 160.8 tps, lat 6.212 ms stddev 1.219, 0 failed
progress: 115.0 s, 157.2 tps, lat 6.365 ms stddev 0.954, 0 failed
progress: 120.0 s, 156.8 tps, lat 6.368 ms stddev 0.995, 0 failed
progress: 125.0 s, 161.2 tps, lat 6.209 ms stddev 0.826, 0 failed
progress: 130.0 s, 145.2 tps, lat 6.883 ms stddev 1.639, 0 failed
progress: 135.0 s, 141.8 tps, lat 7.046 ms stddev 2.234, 0 failed
progress: 140.0 s, 157.8 tps, lat 6.335 ms stddev 0.889, 0 failed
progress: 145.0 s, 161.0 tps, lat 6.212 ms stddev 0.795, 0 failed
progress: 150.0 s, 160.2 tps, lat 6.241 ms stddev 0.880, 0 failed
progress: 155.0 s, 156.0 tps, lat 6.405 ms stddev 0.820, 0 failed
progress: 160.0 s, 159.2 tps, lat 6.282 ms stddev 0.956, 0 failed
progress: 165.0 s, 159.8 tps, lat 6.250 ms stddev 0.920, 0 failed
progress: 170.0 s, 160.6 tps, lat 6.230 ms stddev 0.849, 0 failed
progress: 175.0 s, 154.2 tps, lat 6.481 ms stddev 1.018, 0 failed
progress: 180.0 s, 157.6 tps, lat 6.338 ms stddev 0.901, 0 failed
progress: 185.0 s, 160.8 tps, lat 6.223 ms stddev 0.903, 0 failed
```

progress: 190.0 s, 142.8 tps, lat 6.988 ms stddev 1.606, 0 failed  
progress: 195.0 s, 154.4 tps, lat 6.481 ms stddev 1.036, 0 failed  
progress: 200.0 s, 158.6 tps, lat 6.302 ms stddev 0.852, 0 failed  
progress: 205.0 s, 159.8 tps, lat 6.260 ms stddev 1.021, 0 failed  
progress: 210.0 s, 161.4 tps, lat 6.195 ms stddev 0.970, 0 failed  
progress: 215.0 s, 163.0 tps, lat 6.127 ms stddev 0.941, 0 failed  
progress: 220.0 s, 158.4 tps, lat 6.312 ms stddev 0.853, 0 failed  
progress: 225.0 s, 159.4 tps, lat 6.272 ms stddev 0.975, 0 failed  
progress: 230.0 s, 157.4 tps, lat 6.353 ms stddev 0.817, 0 failed  
progress: 235.0 s, 157.8 tps, lat 6.333 ms stddev 0.949, 0 failed  
progress: 240.0 s, 164.2 tps, lat 6.085 ms stddev 0.772, 0 failed  
progress: 245.0 s, 161.0 tps, lat 6.210 ms stddev 0.850, 0 failed  
progress: 250.0 s, 145.0 tps, lat 6.893 ms stddev 1.422, 0 failed  
progress: 255.0 s, 151.0 tps, lat 6.621 ms stddev 1.020, 0 failed  
progress: 260.0 s, 174.0 tps, lat 5.751 ms stddev 0.873, 0 failed  
progress: 265.0 s, 182.8 tps, lat 5.468 ms stddev 0.519, 0 failed  
progress: 270.0 s, 181.8 tps, lat 5.498 ms stddev 0.537, 0 failed  
progress: 275.0 s, 185.2 tps, lat 5.400 ms stddev 0.643, 0 failed  
progress: 280.0 s, 182.6 tps, lat 5.475 ms stddev 0.572, 0 failed  
progress: 285.0 s, 178.4 tps, lat 5.605 ms stddev 0.822, 0 failed  
progress: 290.0 s, 181.0 tps, lat 5.518 ms stddev 0.638, 0 failed  
progress: 295.0 s, 182.0 tps, lat 5.496 ms stddev 0.612, 0 failed  
progress: 300.0 s, 178.4 tps, lat 5.604 ms stddev 3.013, 0 failed  
progress: 305.0 s, 184.0 tps, lat 5.434 ms stddev 0.521, 0 failed  
progress: 310.0 s, 159.2 tps, lat 6.275 ms stddev 1.671, 0 failed  
progress: 315.0 s, 179.2 tps, lat 5.583 ms stddev 0.905, 0 failed  
progress: 320.0 s, 182.8 tps, lat 5.471 ms stddev 0.578, 0 failed  
progress: 325.0 s, 184.0 tps, lat 5.430 ms stddev 0.620, 0 failed  
progress: 330.0 s, 183.8 tps, lat 5.438 ms stddev 0.489, 0 failed  
progress: 335.0 s, 173.8 tps, lat 5.755 ms stddev 4.975, 0 failed  
progress: 340.0 s, 181.2 tps, lat 5.519 ms stddev 0.634, 0 failed  
progress: 345.0 s, 183.4 tps, lat 5.450 ms stddev 0.514, 0 failed  
progress: 350.0 s, 183.2 tps, lat 5.459 ms stddev 0.586, 0 failed  
progress: 355.0 s, 183.4 tps, lat 5.447 ms stddev 0.542, 0 failed  
progress: 360.0 s, 175.6 tps, lat 5.695 ms stddev 0.992, 0 failed  
progress: 365.0 s, 181.8 tps, lat 5.500 ms stddev 0.686, 0 failed  
progress: 370.0 s, 164.4 tps, lat 6.077 ms stddev 1.429, 0 failed  
progress: 375.0 s, 180.2 tps, lat 5.552 ms stddev 0.819, 0 failed  
progress: 380.0 s, 182.4 tps, lat 5.479 ms stddev 0.472, 0 failed  
progress: 385.0 s, 172.0 tps, lat 5.815 ms stddev 1.480, 0 failed  
progress: 390.0 s, 182.4 tps, lat 5.482 ms stddev 0.574, 0 failed  
progress: 395.0 s, 181.8 tps, lat 5.497 ms stddev 0.672, 0 failed  
progress: 400.0 s, 182.0 tps, lat 5.494 ms stddev 0.667, 0 failed  
progress: 405.0 s, 176.2 tps, lat 5.669 ms stddev 0.656, 0 failed  
progress: 410.0 s, 170.2 tps, lat 5.873 ms stddev 0.803, 0 failed  
progress: 415.0 s, 164.8 tps, lat 6.066 ms stddev 0.701, 0 failed  
progress: 420.0 s, 163.0 tps, lat 6.135 ms stddev 0.877, 0 failed  
progress: 425.0 s, 158.4 tps, lat 6.315 ms stddev 0.892, 0 failed  
progress: 430.0 s, 149.2 tps, lat 6.692 ms stddev 1.619, 0 failed  
progress: 435.0 s, 156.0 tps, lat 6.411 ms stddev 1.229, 0 failed  
progress: 440.0 s, 161.8 tps, lat 6.182 ms stddev 1.076, 0 failed  
progress: 445.0 s, 161.8 tps, lat 6.179 ms stddev 0.805, 0 failed  
progress: 450.0 s, 161.2 tps, lat 6.199 ms stddev 0.922, 0 failed  
progress: 455.0 s, 162.8 tps, lat 6.138 ms stddev 0.870, 0 failed  
progress: 460.0 s, 161.8 tps, lat 6.177 ms stddev 0.975, 0 failed  
progress: 465.0 s, 146.2 tps, lat 6.842 ms stddev 1.710, 0 failed  
progress: 470.0 s, 159.6 tps, lat 6.263 ms stddev 0.928, 0 failed  
progress: 475.0 s, 160.0 tps, lat 6.247 ms stddev 0.755, 0 failed  
progress: 480.0 s, 161.4 tps, lat 6.192 ms stddev 1.048, 0 failed  
progress: 485.0 s, 158.6 tps, lat 6.307 ms stddev 0.897, 0 failed  
progress: 490.0 s, 147.0 tps, lat 6.799 ms stddev 1.780, 0 failed  
progress: 495.0 s, 156.8 tps, lat 6.372 ms stddev 1.260, 0 failed  
progress: 500.0 s, 161.0 tps, lat 6.208 ms stddev 0.910, 0 failed  
progress: 505.0 s, 156.8 tps, lat 6.377 ms stddev 1.018, 0 failed  
progress: 510.0 s, 157.4 tps, lat 6.353 ms stddev 1.231, 0 failed  
progress: 515.0 s, 153.2 tps, lat 6.523 ms stddev 0.793, 0 failed  
progress: 520.0 s, 156.6 tps, lat 6.384 ms stddev 0.848, 0 failed  
progress: 525.0 s, 156.6 tps, lat 6.387 ms stddev 0.778, 0 failed  
progress: 530.0 s, 153.4 tps, lat 6.513 ms stddev 2.562, 0 failed  
progress: 535.0 s, 150.2 tps, lat 6.657 ms stddev 1.663, 0 failed  
progress: 540.0 s, 158.0 tps, lat 6.329 ms stddev 1.116, 0 failed  
progress: 545.0 s, 159.4 tps, lat 6.267 ms stddev 0.800, 0 failed  
progress: 550.0 s, 144.2 tps, lat 6.929 ms stddev 1.490, 0 failed  
progress: 555.0 s, 158.2 tps, lat 6.326 ms stddev 1.059, 0 failed  
progress: 560.0 s, 167.8 tps, lat 5.957 ms stddev 0.818, 0 failed  
progress: 565.0 s, 160.6 tps, lat 6.225 ms stddev 1.096, 0 failed  
progress: 570.0 s, 162.6 tps, lat 6.141 ms stddev 0.922, 0 failed  
progress: 575.0 s, 163.6 tps, lat 6.116 ms stddev 0.923, 0 failed  
progress: 580.0 s, 161.6 tps, lat 6.185 ms stddev 0.924, 0 failed  
progress: 585.0 s, 154.8 tps, lat 6.451 ms stddev 0.948, 0 failed

```

progress: 590.0 s, 158.2 tps, lat 6.326 ms stddev 0.953, 0 failed
progress: 595.0 s, 145.2 tps, lat 6.885 ms stddev 1.808, 0 failed
progress: 600.0 s, 156.6 tps, lat 6.383 ms stddev 1.214, 0 failed
progress: 605.0 s, 155.4 tps, lat 6.428 ms stddev 0.910, 0 failed
progress: 610.0 s, 145.2 tps, lat 6.884 ms stddev 1.668, 0 failed
transaction type: <builtin: TPC-B (sort of)>
scaling factor: 1
query mode: prepared
number of clients: 1
number of threads: 1
maximum number of tries: 1
number of transactions per client: 100000
number of transactions actually processed: 100000/100000
number of failed transactions: 0 (0.000%)
latency average = 6.132 ms
latency stddev = 1.264 ms
initial connection time = 2.521 ms
tps = 163.033986 (without initial connection time)

```

Видно, что возраст незамороженных транзакций превышает установленное пороговое значение (100000):

```
=> SELECT datname, datfrozenxid, age(datfrozenxid) FROM pg_database;
```

datname	datfrozenxid	age
postgres	749	100013
student	749	100013
template1	749	100013
template0	722	100040
mvcc_freeze	749	100013

(5 rows)

Теперь при выполнении команды VACUUM для любой таблицы будет запущен процесс автоочистки.

```
=> VACUUM t;
```

VACUUM

Среди процессов появился autovacuum worker:

```
student$ sudo head -n 1 /var/lib/postgresql/16/main/postmaster.pid
```

173507

```
student$ ps -o pid,command --ppid 173507
```

PID	COMMAND
173508	postgres: 16/main: checkpointer
173509	postgres: 16/main: background writer
173511	postgres: 16/main: walwriter
173512	postgres: 16/main: logical replication launcher
173554	postgres: 16/main: student mvcc_freeze [local] idle
174291	postgres: 16/main: autovacuum worker

И через некоторое время транзакции окажутся замороженными:

```
=> SELECT datname, datfrozenxid, age(datfrozenxid) FROM pg_database;
```

datname	datfrozenxid	age
postgres	100762	0
student	100762	0
template1	100762	0
template0	100762	0
mvcc_freeze	99829	933

(5 rows)