PostgreSQL+1C
Нюансы эксплуатации

Смолкин Григорий

www.postgrespro.ru

Postgres PROFESSIONAL

Тестирование приложением
Почему это важно для нас

Участники

- Смолкин Григорий
- Петров Сергей
- Попов Николай
- Лубенникова Анастасия
- Федор Сигаев
- Пан Константин

- Жданюк Александр
- Елисеев Андрей

Условия

Software

Centos 7.2
PostgreSQL 9.5.4
+1C patchset
+рекомендуемые 1C настройки

Hardware

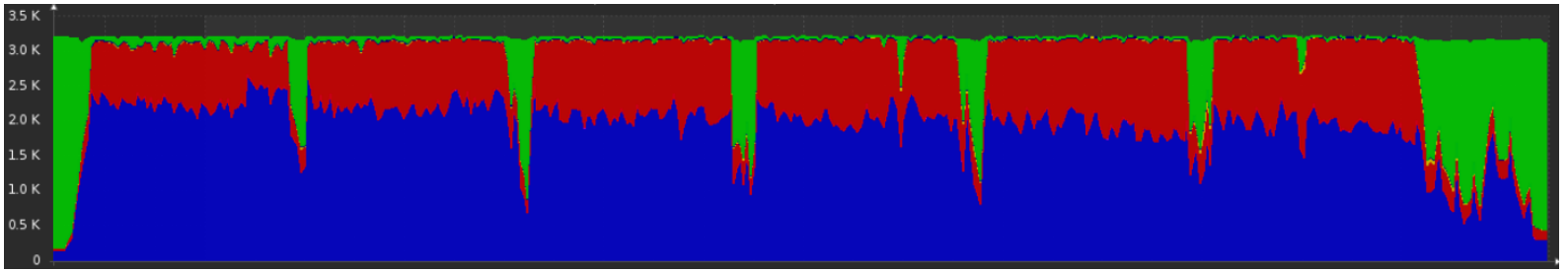
Intel(R) Xeon(R)
CPU E7- 8837 2.67GHz
64GB RAM
Fusion IODriver

1C Платформа 8.3.8
АСКУ: 220 GB(77 +7BPM/час)
АСБНУ: 47 GB(52 +5BPM/час)
АСЗУП: 77 GB(15 +2BPM/час)
Продолжительность: 10 часов

Временные таблицы

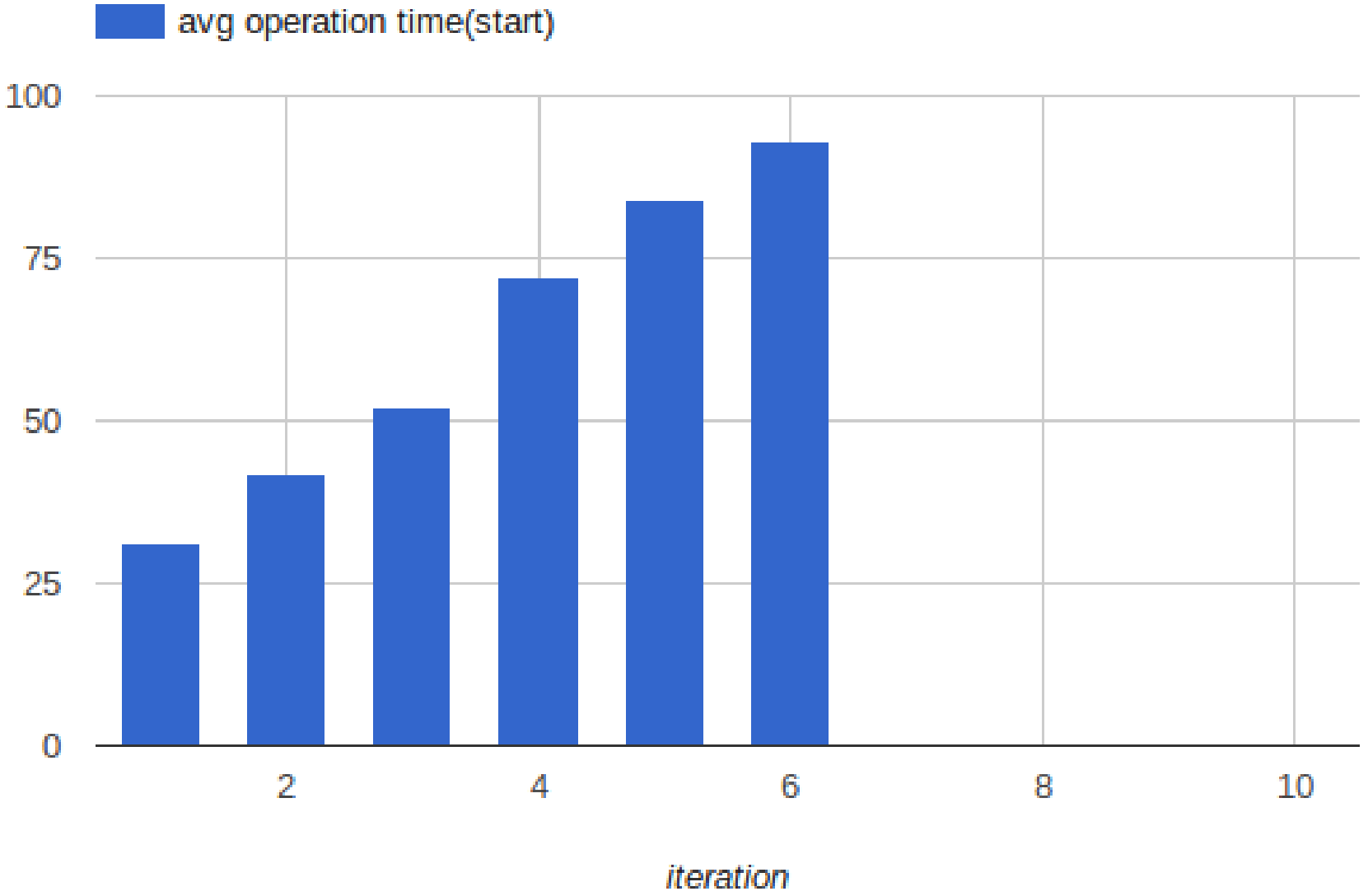
1. Привязаны к определенному соединению
2. Не имеют механизма сбора статистики
3. Размещены в локальной памяти воркера(temp_buffers)
4. Удаляются после отключения клиента

Первые попытки



Первые попытки

Time(sec)



Инструментарий

- mamonsu+zabbix
- atop+atopsar
- perf+flamegraph

Perf

+	30.86%	0.04%	postmaster	postgres	[.]	uint32_hash
+	23.21%	1.25%	postmaster	[kernel.kallsyms]	[k]	page_fault
+	21.95%	0.00%	postmaster	[kernel.kallsyms]	[k]	do_page_fault
+	21.85%	0.30%	postmaster	[kernel.kallsyms]	[k]	__do_page_fault
+	21.32%	0.13%	postmaster	[kernel.kallsyms]	[k]	handle_mm_fault
+	19.17%	15.12%	postmaster	postgres	[.]	hash_search_with_hash_value
+	15.76%	0.21%	postmaster	postgres	[.]	0x0000000003a2f79
+	15.12%	0.04%	postmaster	[kernel.kallsyms]	[k]	__alloc_pages_nodemask
+	14.81%	0.01%	postmaster	[kernel.kallsyms]	[k]	alloc_pages_vma
+	14.68%	0.03%	postmaster	[kernel.kallsyms]	[k]	migrate_pages
+	14.12%	0.00%	postmaster	[kernel.kallsyms]	[k]	do_huge_pmd_anonymous_page
+	13.93%	0.00%	postmaster	[kernel.kallsyms]	[k]	__alloc_pages_direct_compact
+	13.92%	0.00%	postmaster	[kernel.kallsyms]	[k]	try_to_compact_pages
+	13.92%	0.00%	postmaster	[kernel.kallsyms]	[k]	compact_zone_order
+	13.92%	0.08%	postmaster	[kernel.kallsyms]	[k]	compact_zone
+	8.95%	0.24%	postmaster	[kernel.kallsyms]	[k]	compaction_alloc
+	8.41%	7.95%	postmaster	[kernel.kallsyms]	[k]	isolate_freepages_block

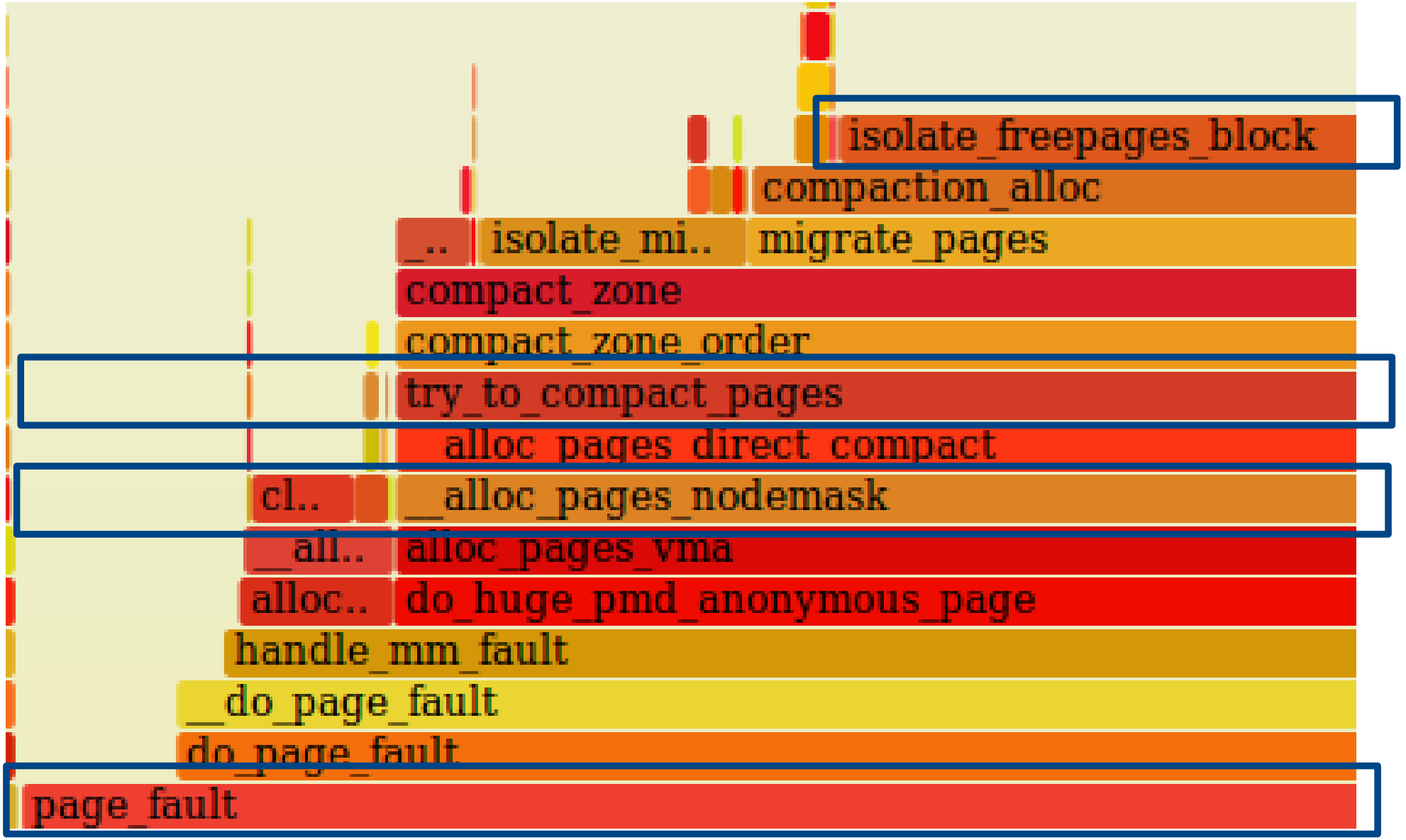
Занятно, но не понятно

Perf

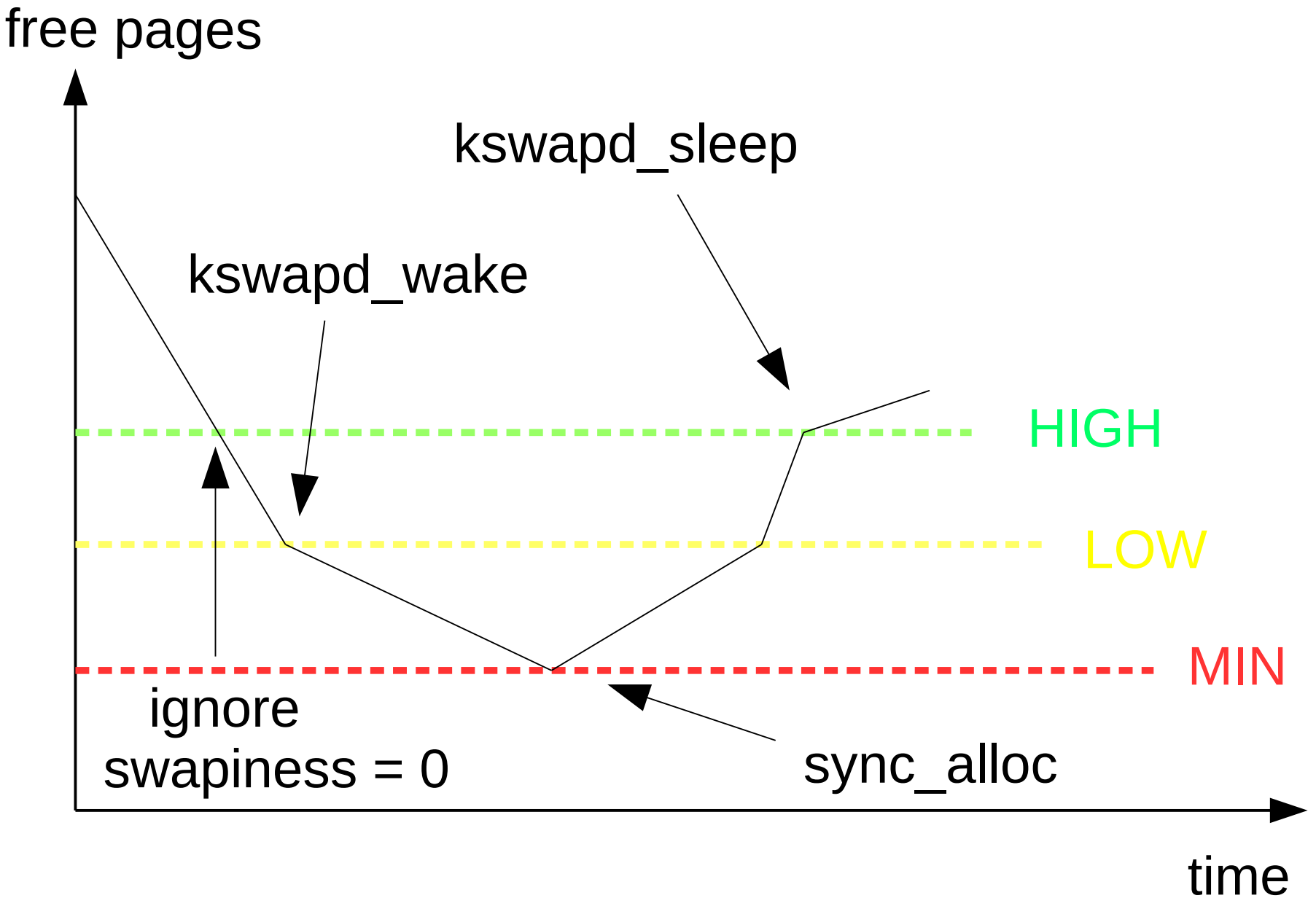
```
14.82%  postmaster      [kernel.kallsyms]      [k] isolate_freepages_block
11.07%  postmaster      postgres               [.] hash_search_with_hash_value
```

По-прежнему непонятно

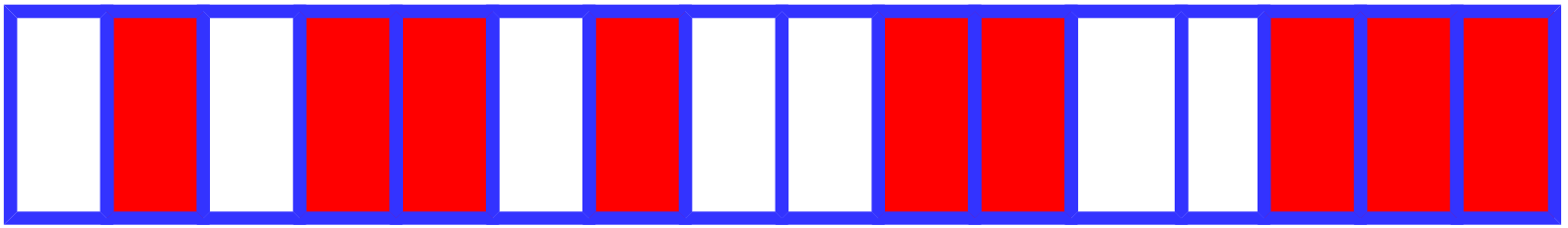
Дефрагментируем память?



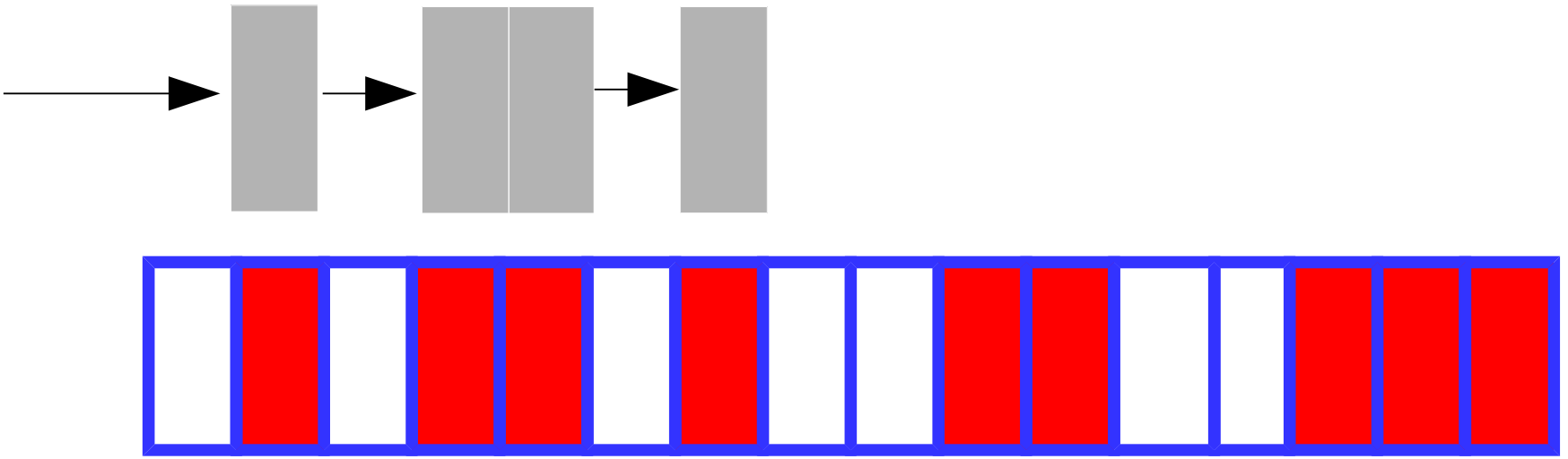
Memory Reclamation



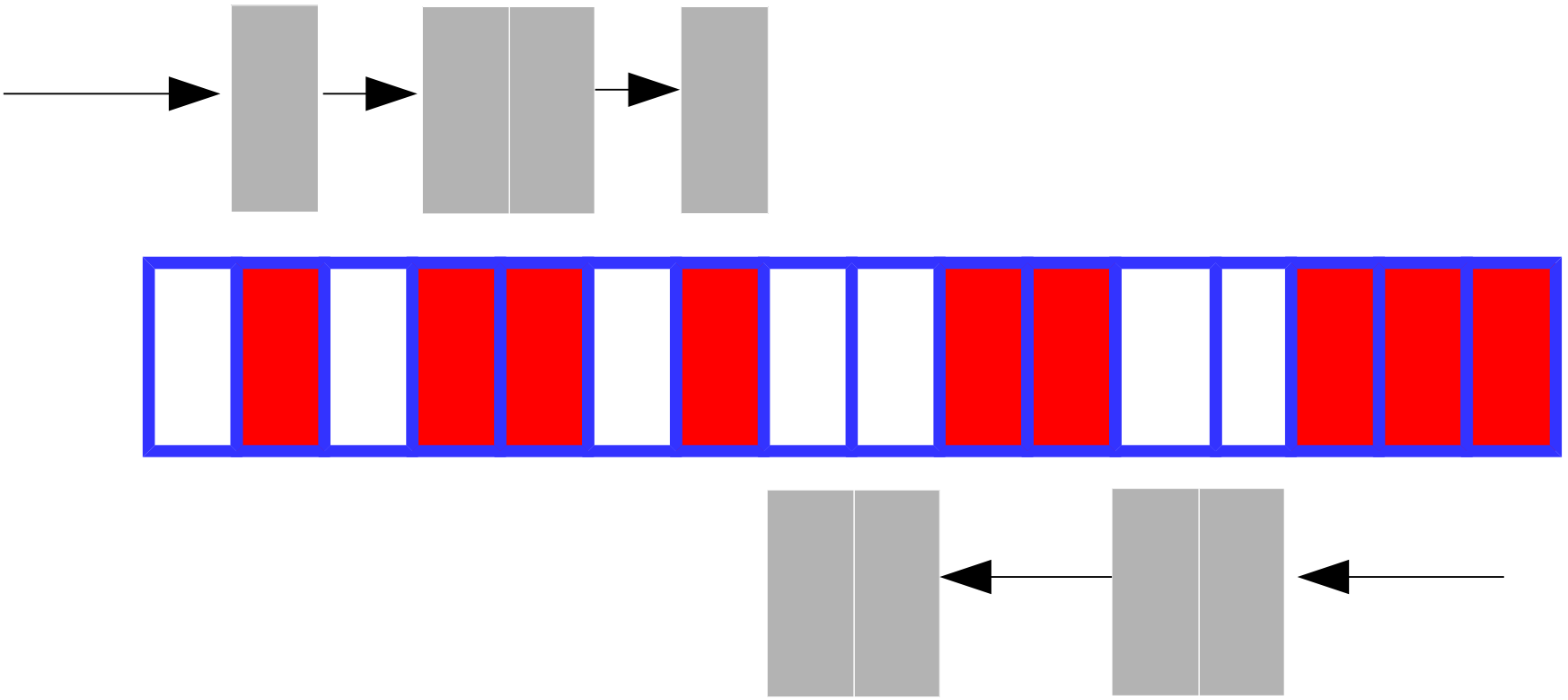
Дефрагментация



Дефрагментация

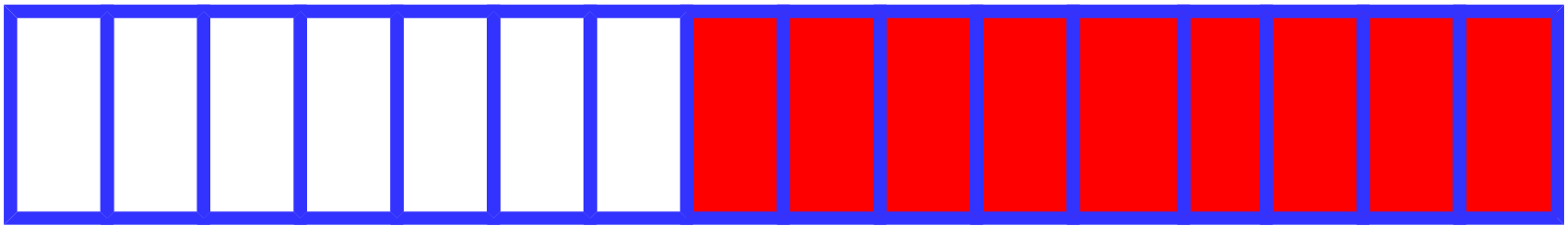


Дефрагментация



Postgres PROFESSIONAL

Дефрагментация



Дефрагментация

vm.min_free_kbytes:

+kswapd раньше начнет работу

+больше свободной памяти

-меньше единовременно доступной памяти

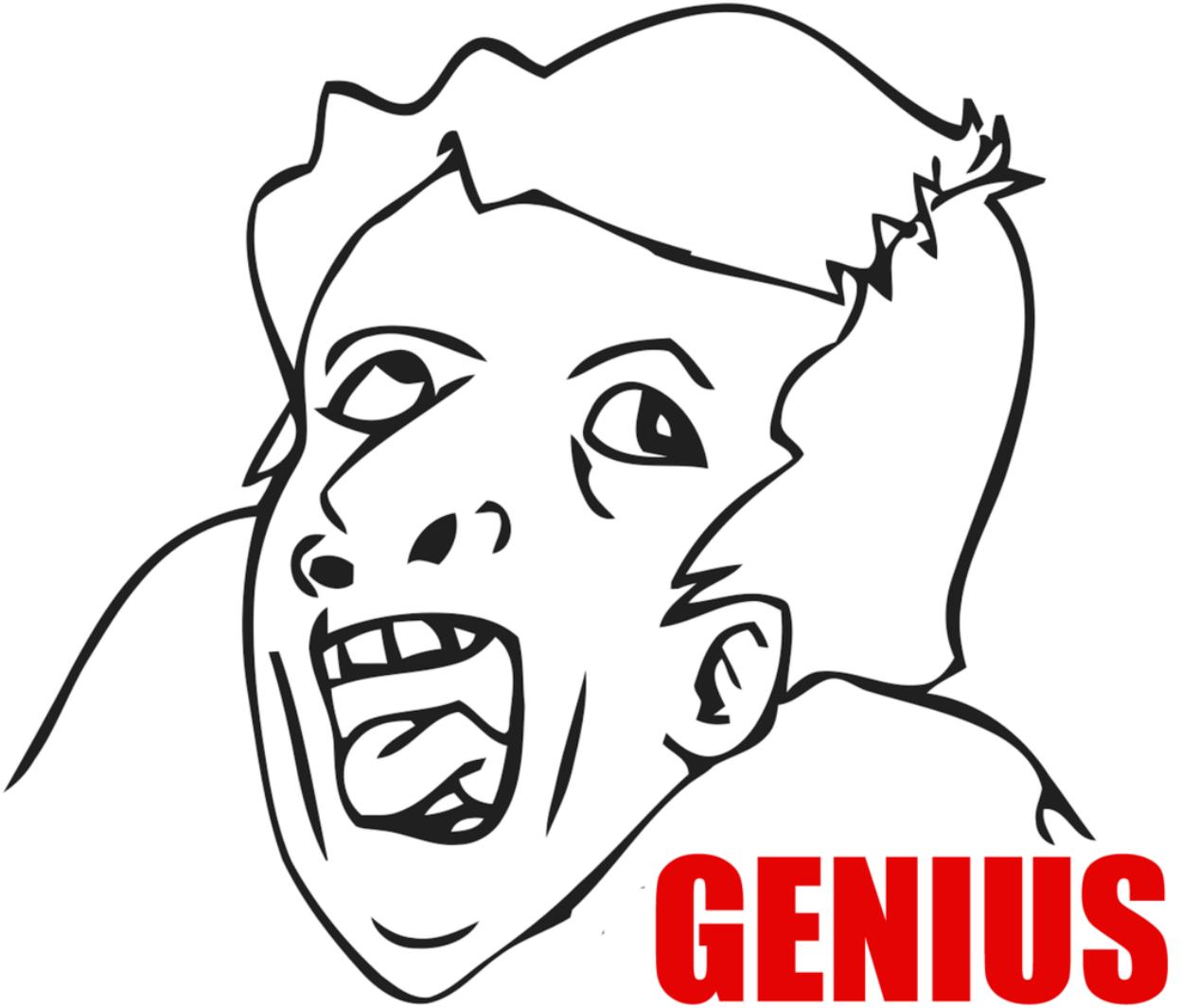
/sys/kernel/debug/extfrag_index
/proc/buddyinfo

extfrag_threshold = 500

Резервирование места для временных таблиц

```
open("base/13090/t2_1043562", O_RDWR)   = 40
write(40, "\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0"..., 8192) = 8192
write(40, "\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0"..., 8192) = 8192
write(40, "\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0"..., 8192) = 8192
write(40, "\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0"..., 8192) = 8192
write(40, "\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0"..., 8192) = 8192
write(40, "\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0"..., 8192) = 8192
write(40, "\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0"..., 8192) = 8192
write(40, "\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0"..., 8192) = 8192
write(40, "\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0"..., 8192) = 8192
write(40, "\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0"..., 8192) = 8192
```

Postgres PROFESSIONAL



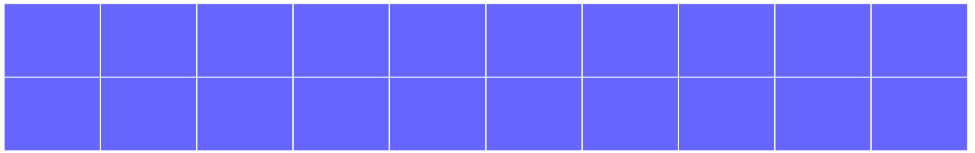
Postgres PROFESSIONAL

Почему это плохо

1. Вытеснение полезного файлого кэша ОС
2. Вытеснение полезного кэша накопителя
3. Ненужное I/O
4. Фрагментация памяти(8КБ = 2^1 * 4KB)
5. Тройная буферизация!!!

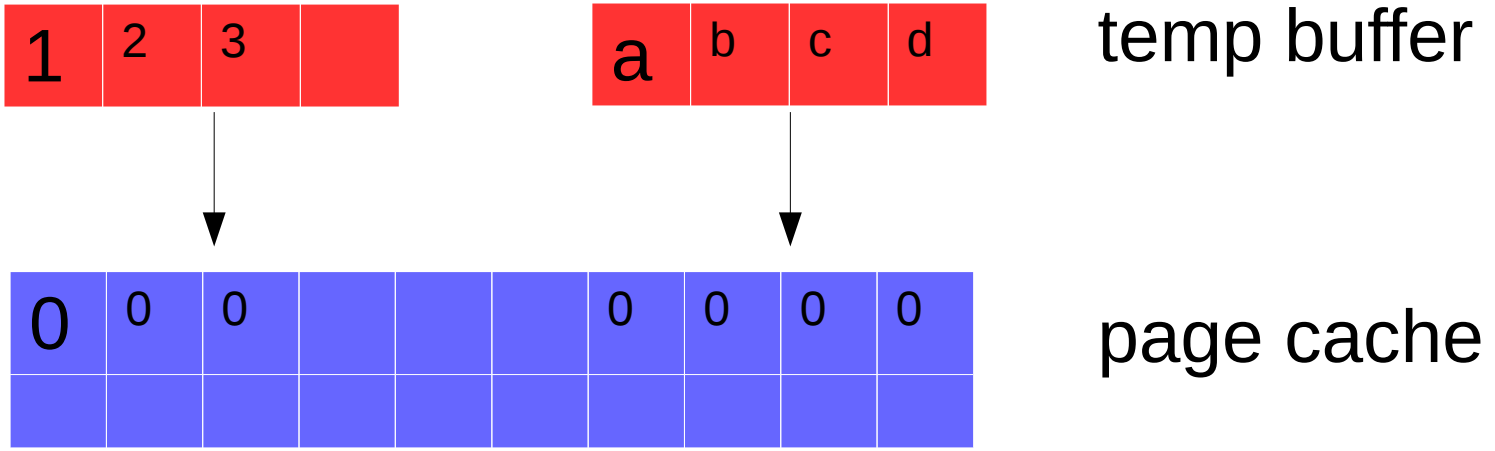
Буферизация

temp buffer

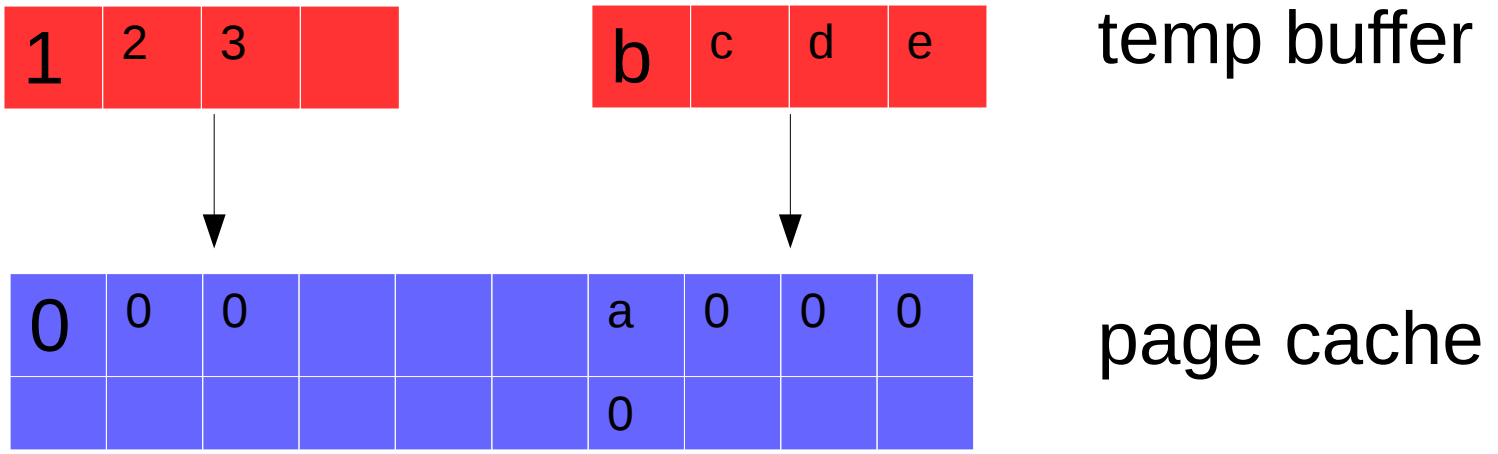


page cache

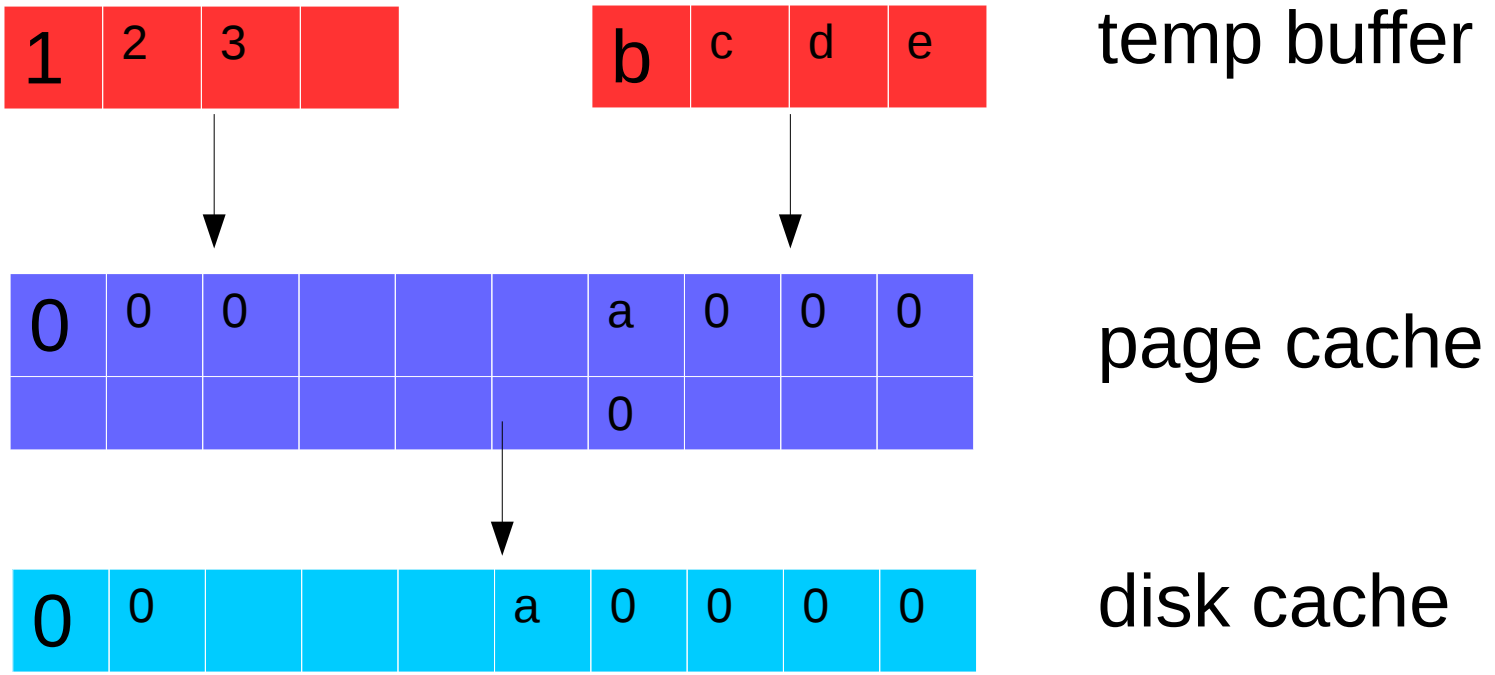
Буферизация



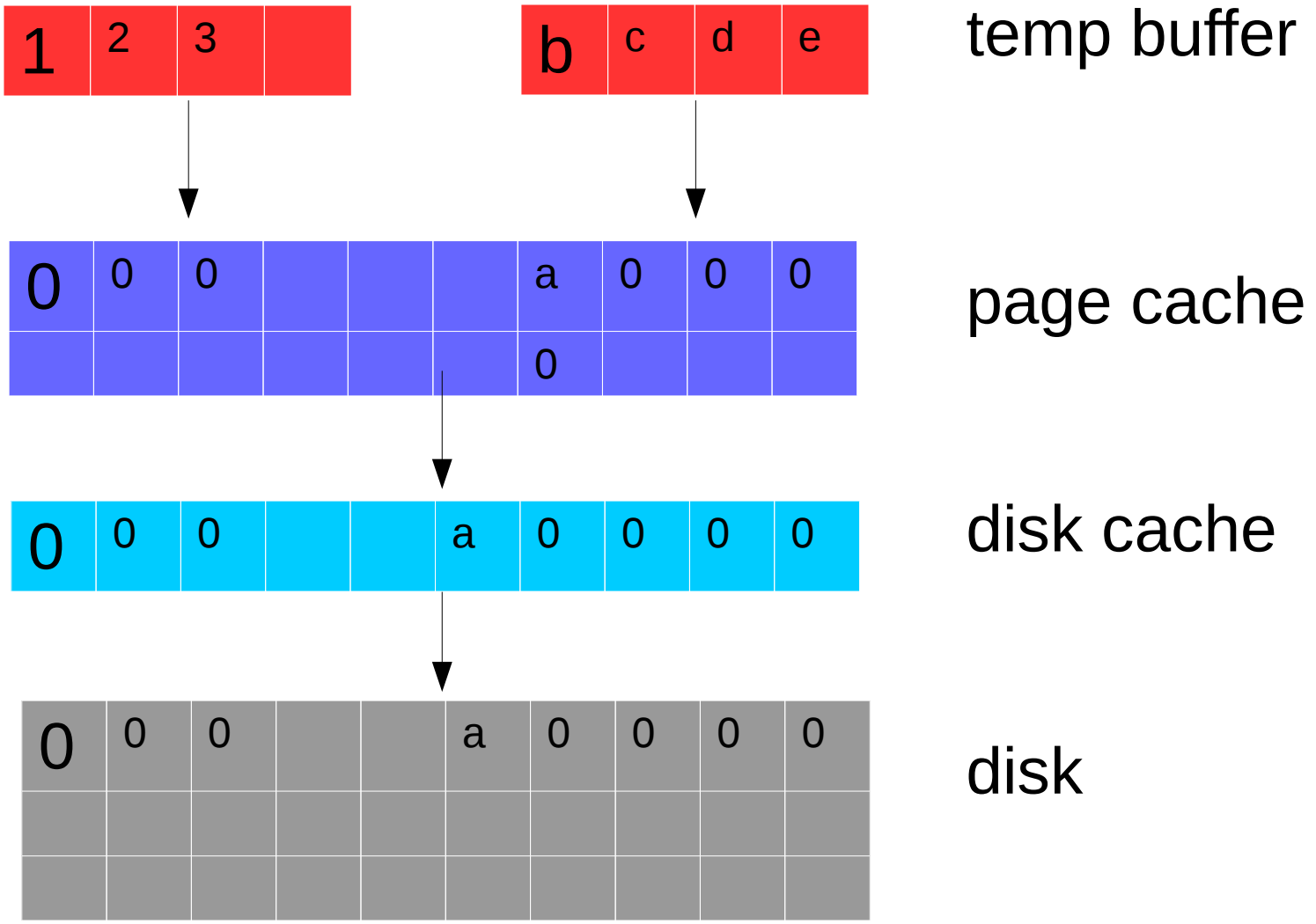
Буферизация



Буферизация



Буферизация



TMPFS

1. Вытеснение полезного файлого кэша ОС
2. ~~Вытеснение полезного кэша накопителя~~
3. ~~Ненужное I/O~~
4. Фрагментация памяти(8КБ = 2^1 * 4KB)
5. ~~Тройная буферизация!!!~~
6. Двойная буферизация

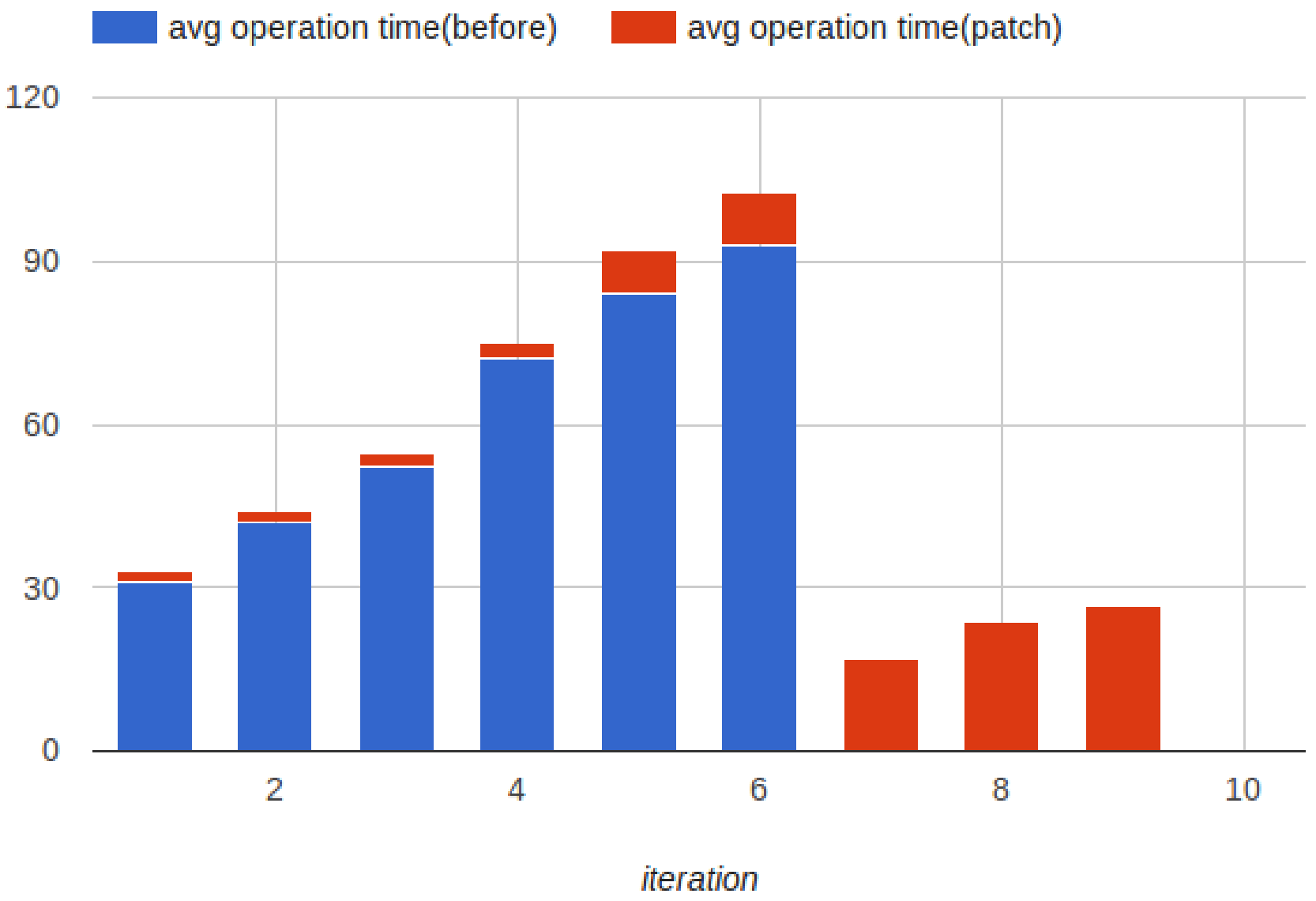
Мы отправились в отдел разработки

Патч для временных таблиц

Автор: Анастасия Лубенникова

- Решает все вышеуказанные проблемы
- Индексы,fsm, vm пишутся на диск
- В ближайшее время войдет в одну из наших сборок

Результат


avg operation time(before)
avg operation time(patch)
120
90
60
30
0
2
4
6
8
10
iteration

Нюансы 1С с точки зрения СУБД

1. Active/Idle connections
2. Idle in Transaction
3. Временные таблицы
4. Разное(uuid in bytea, plain storage)

Orphan temp tables

Условия появления:

1. Крэш постгреса(питание, oom killer, etc.)
2. Не хватило памяти на локи

Симптомы:

1. out of shared memory
2. autovacuum «found orphan table»

Orphan temp tables

Чем грозит?

- Распухание каталога
- Автовакуум не удаляет осиротевшие таблицы, но очень активно спамит в лог

orphan temp tables

Что делать?

- Увеличить max_locks_per_transactions
  lock_table = max_locks_per_transaction * (max_conn + max_pred_locks_per_transaction)
- DROP SCHEMA pg_temp_N CASCADE
  DROP SCHEMA pg_toast_temp_N CASCADE

Мы отправились в отдел разработки

Результат

Автор: Константин Пан

- Бэкенд-процесс теперь корректно завершает свою работу
- Добавлена опция keep_orphan_tables, определяющая политику автовакуума относительно временных таблиц
- https://commitfest.postgresql.org/11/831/

temp_buffers

Постгрес не умеет возвращать системе память, аллоцированную под временные таблицы

temp_buffers

Что делать?

- Рассчитывать temp_buffers исходя из max_connections.
- Переходить на 1С 8.3.9
- Мы рассматриваем возможность сделать патч, исправляющий данное поведение

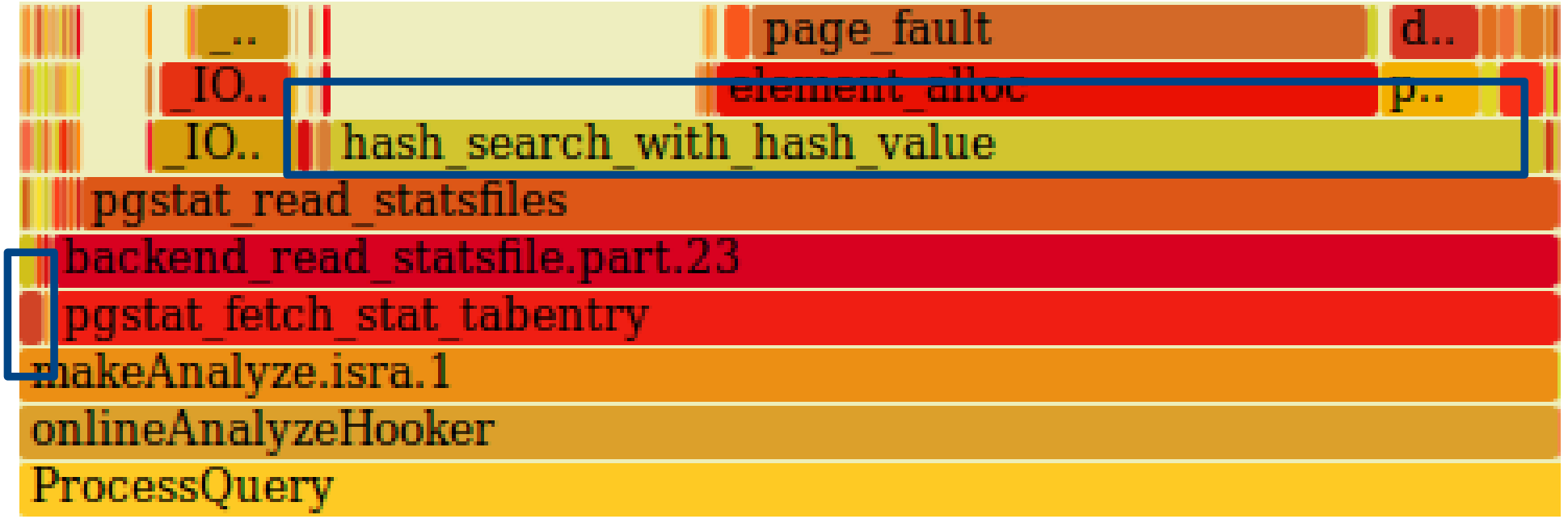
online_analyze

Расширение, призванное решить проблему сбора статистики для временных таблиц:

- В случае пишущего запроса принудительно делает ANALYZE таблицы
- Принимает решение на основе статистики, собранной stats collector`ом

online_analyze

online_analyze



online_analyze

При большом кол-ве временных таблиц:

- Чтение файла со статистикой становится дорогим
- Поиск по прочитанному хэшу становится дорогим
- Большой оверхед на получение статистики

online_analyze

Что делать? Отключать?

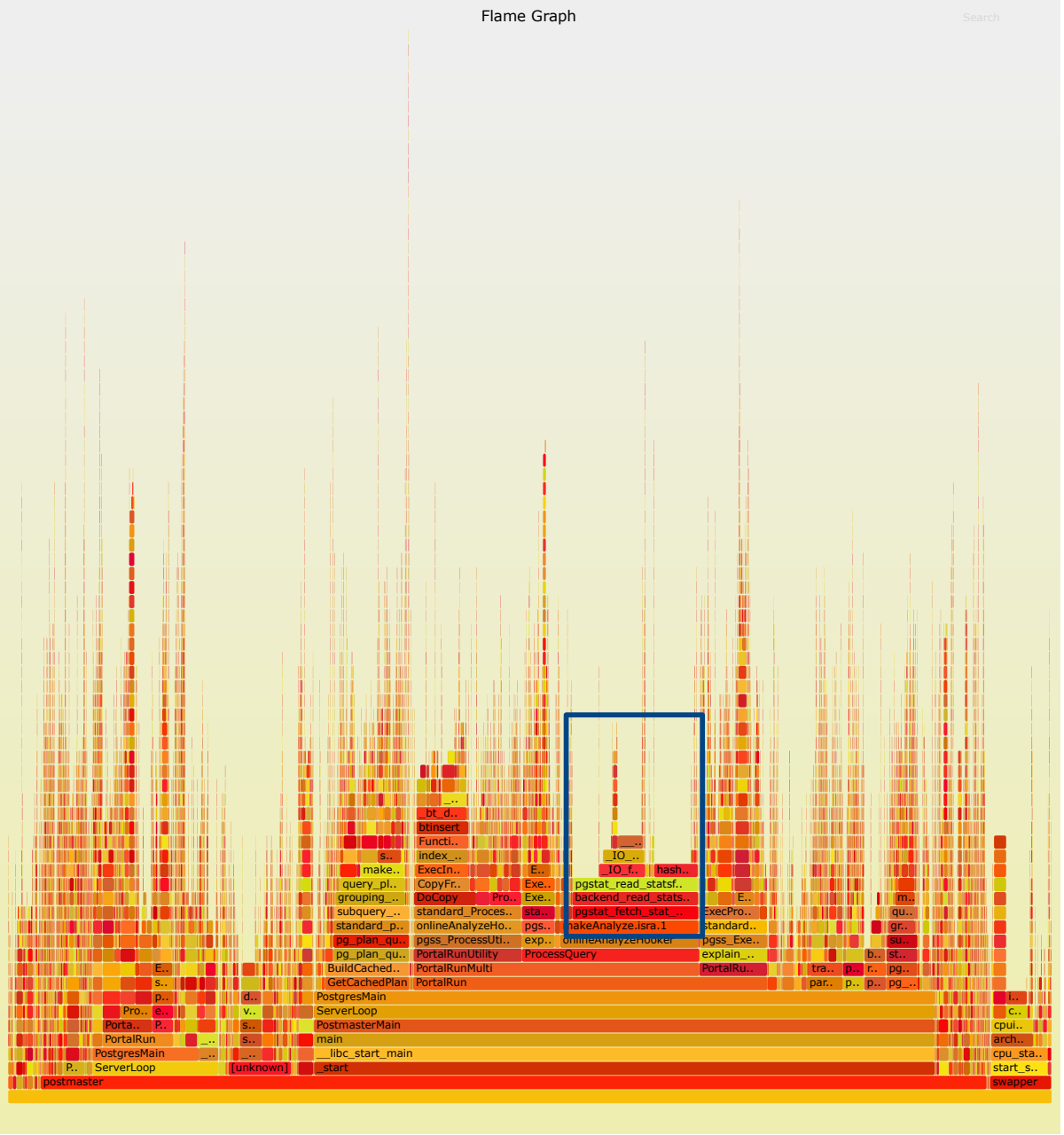
Postgres PROFESSIONAL

Патч для online_analyze

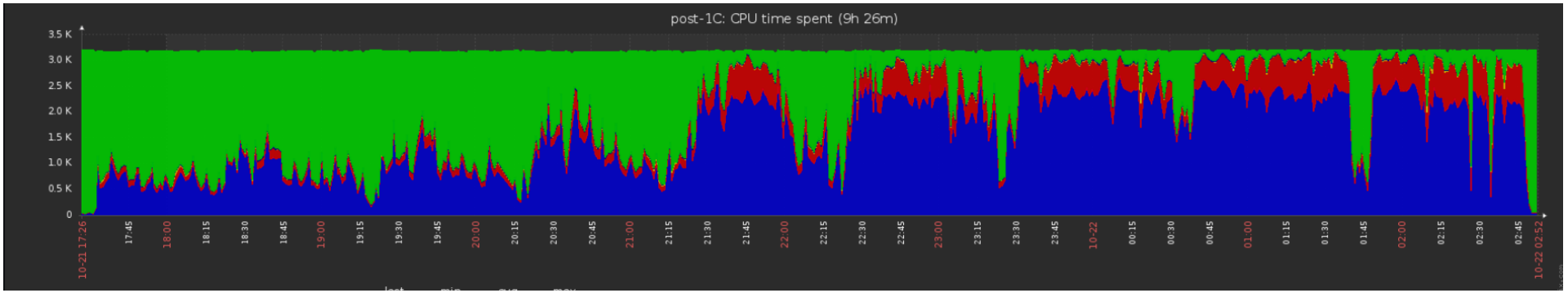
Автор: Фёдор Сигаев

1. Изменена эвристика процедуры принятия решения о запуске ANALYZE по таблице
2. Ожидаем, что войдет в 1С-сборку PostgresPro
3. Work in Progress

Патч для online_analyze

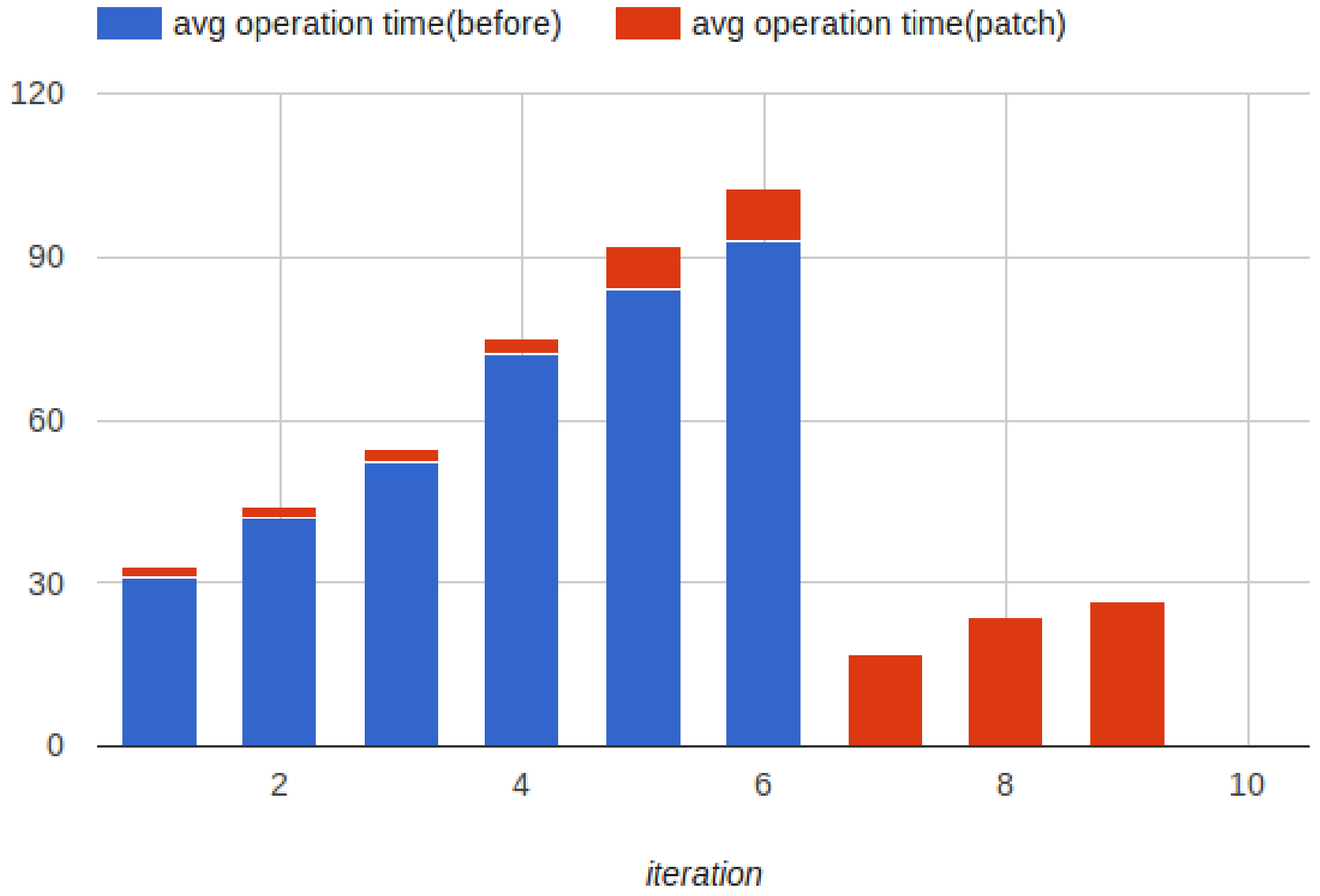


Online_analyze+fasttruncate патч

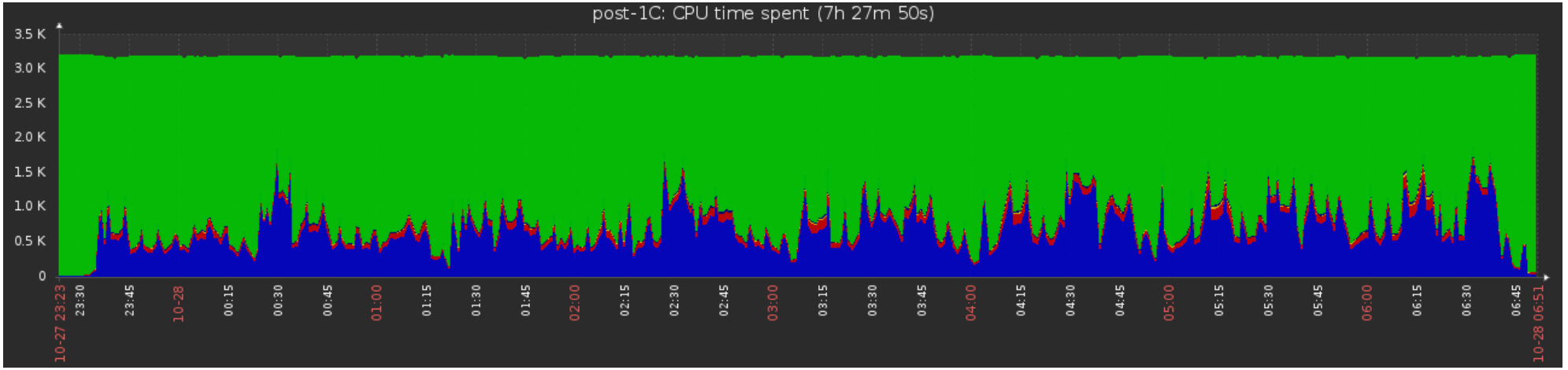


online_analyze+fasttruncate патч

Time(sec)

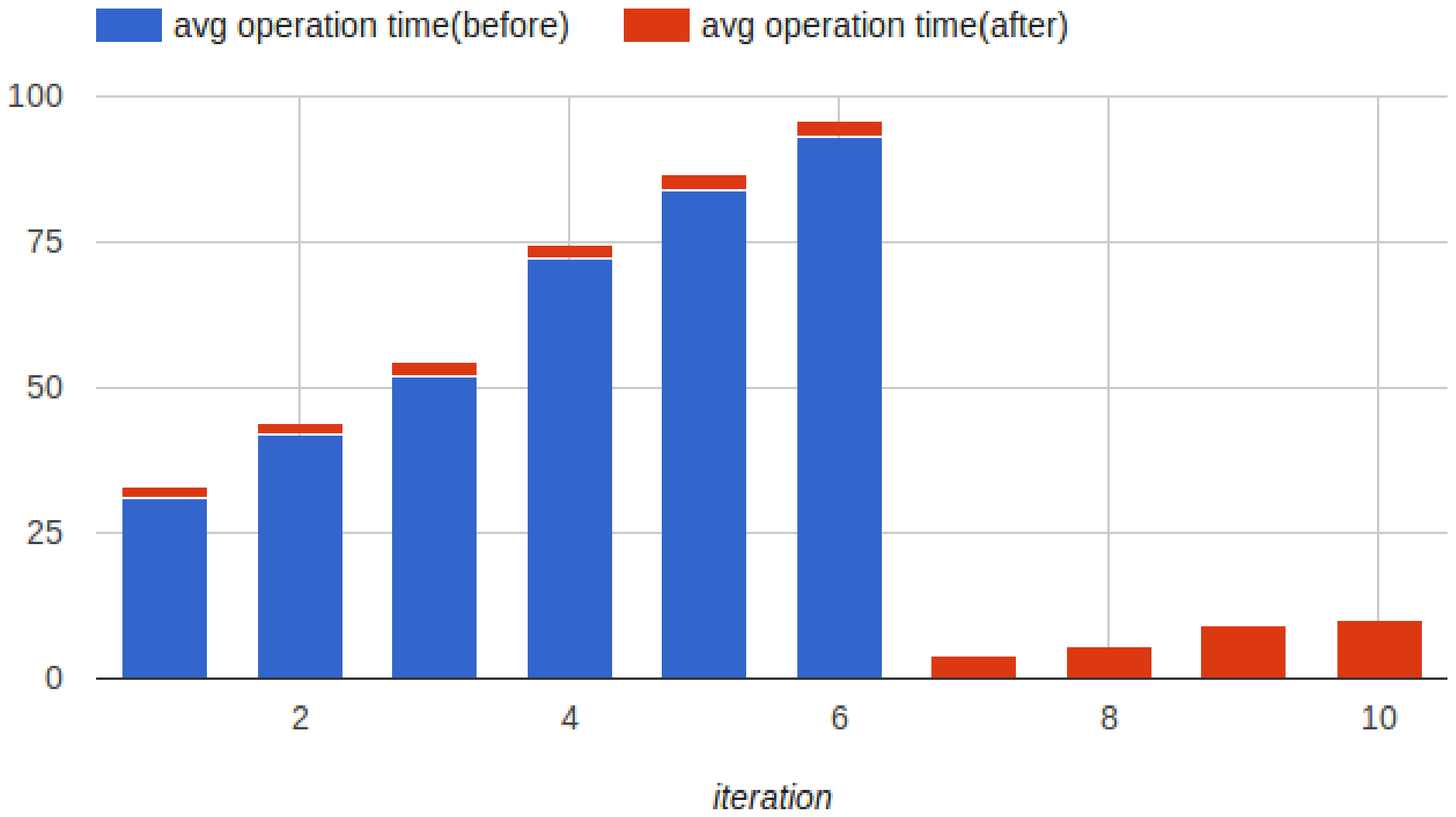


online_analyze off



online_analyze off

Time(sec)



Postgres PROFESSIONAL

Вопросы?

Postgres PROFESSIONAL

Спасибо за внимание!

Контакты:

g.smolkin@postgrespro.ru