PostgreSQL+1C. Нюансы

Смолкин Григорий
Петров Сергей

Тестирование PG+1C
Почему это важно

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

Hardware

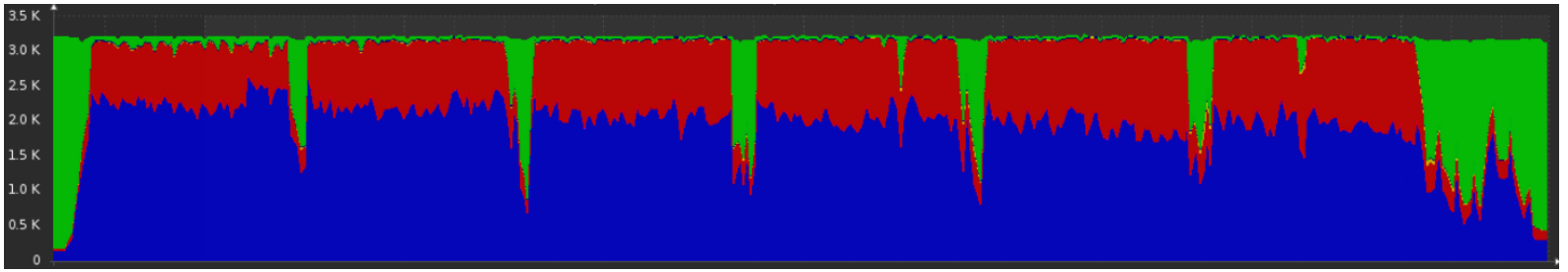
Intel(R) Xeon(R)
CPU E7- 8837 2.67GHz
64GB RAM

1С Платформа 8.3.8
АСКУ: 220GB(77 +7BPM)
АСБНУ: 47 GB(52 +5BPM)
АСЗУП: 77 GB(15 +2BPM)
Время: 10 часов

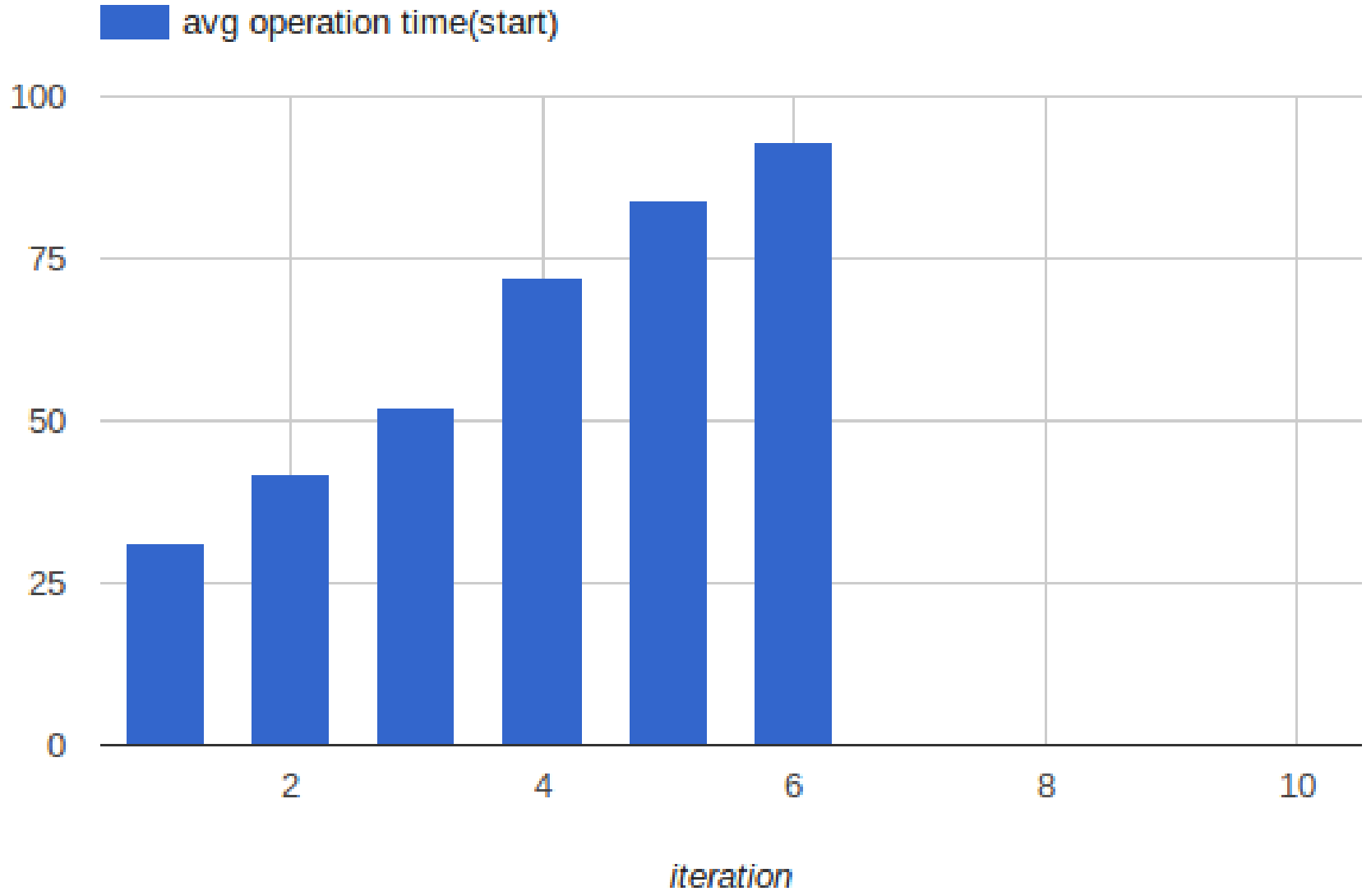
Нюансы 1С с точки зрения СУБД

1. Active/Idle connections
2. Idle in Transaction
3. Temp tables
4. Misc(uuid in bytea, plain storage)

Первые результаты



Первые результаты



Кто виноват?

Инструменты:

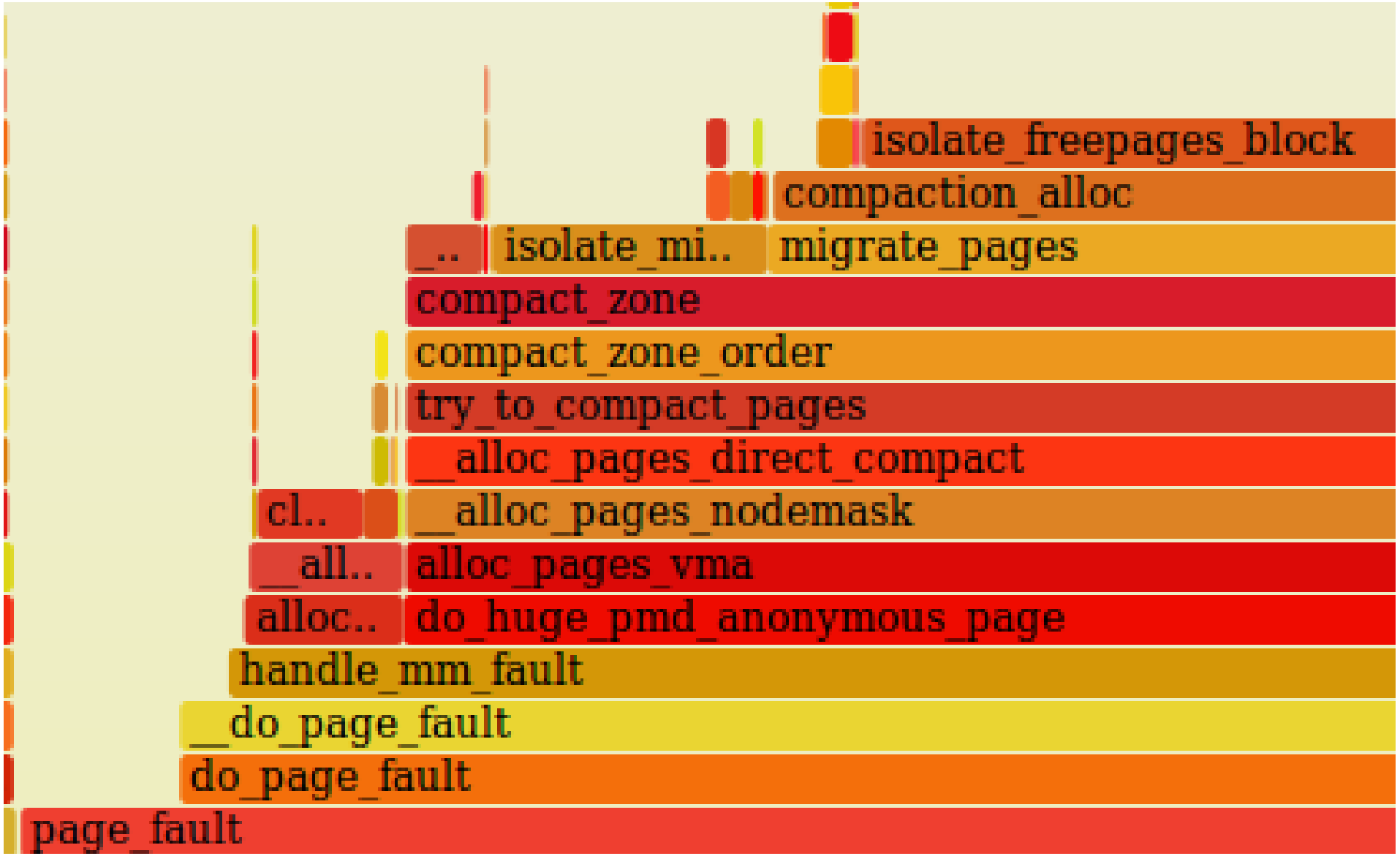
- mamonsu+zabbix
- atop+atopsar
- perf+flamegraph

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

Понятно, но не наглядно

Фрагментация памяти?



Postgres PROFESSIONAL

Direct compact

1. page_fault
2. alloc_pages_nodemask
3. alloc_pages_direct_compact
4. try_to_compact_pages(stall in atop)

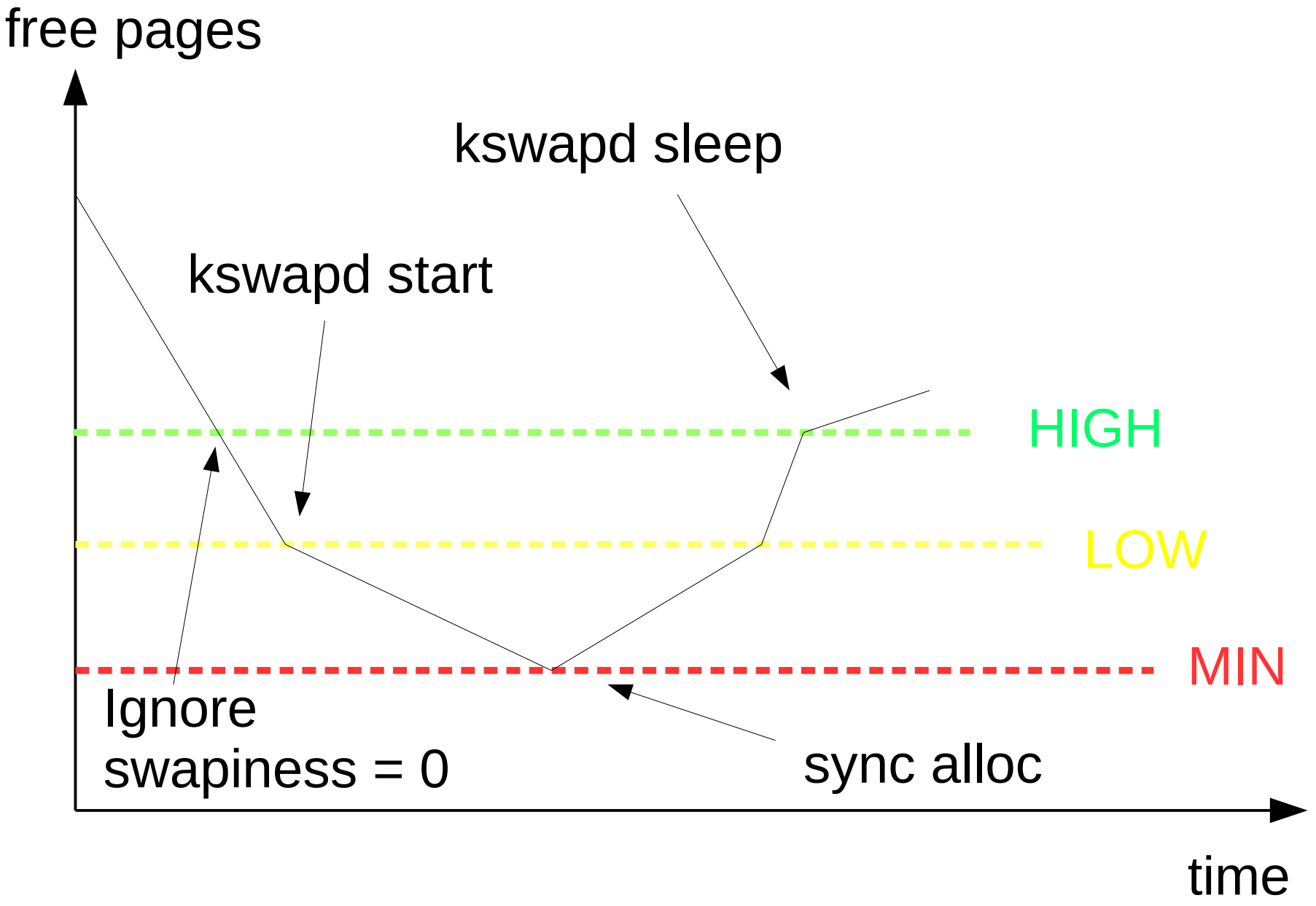
Buddy allocator

Проблемы:

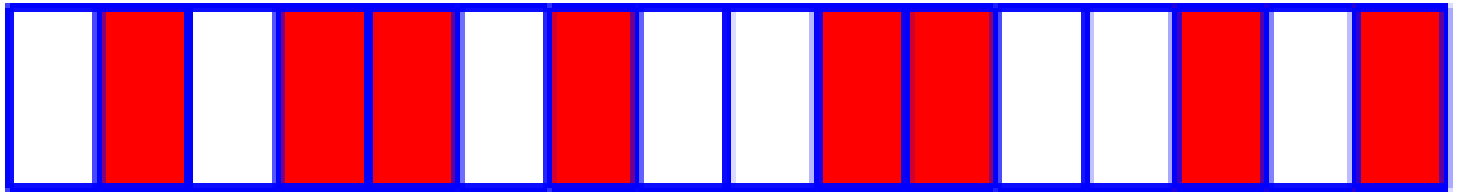
1. Внешняя фрагментация
2. Внутренняя фрагментация

Формула:
Диагностировать:
/sys/kernel/debug/extfrag_index
Рычаг:
extfrag_threshold = 500

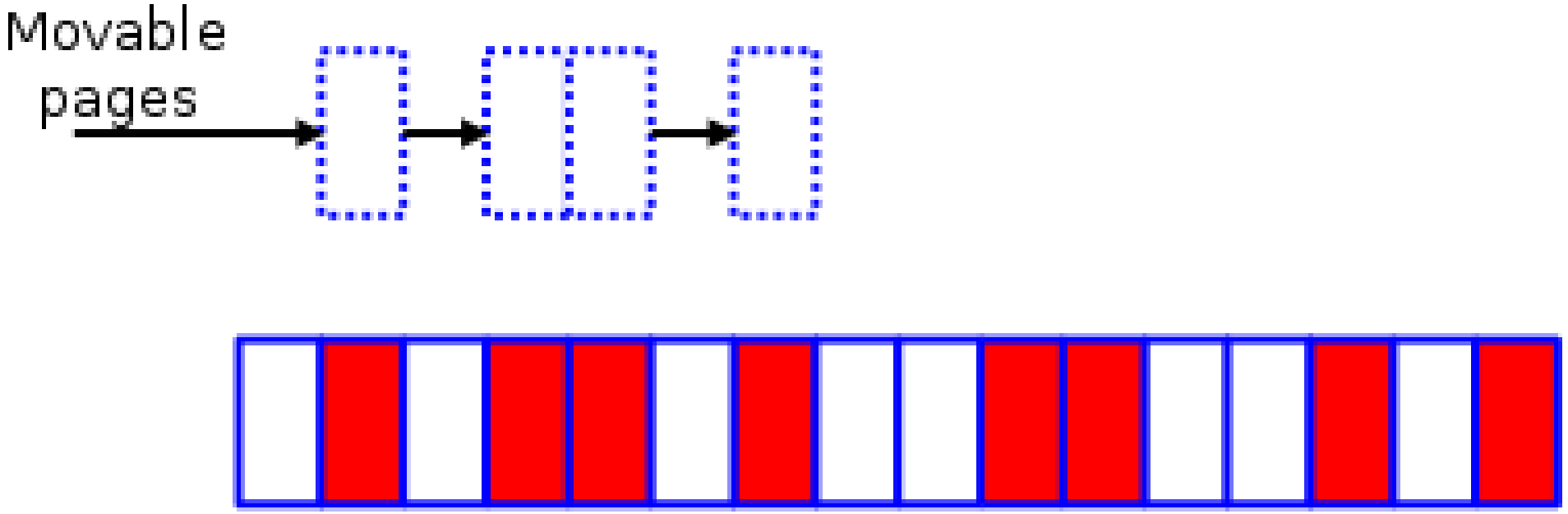
Memory Reclamation



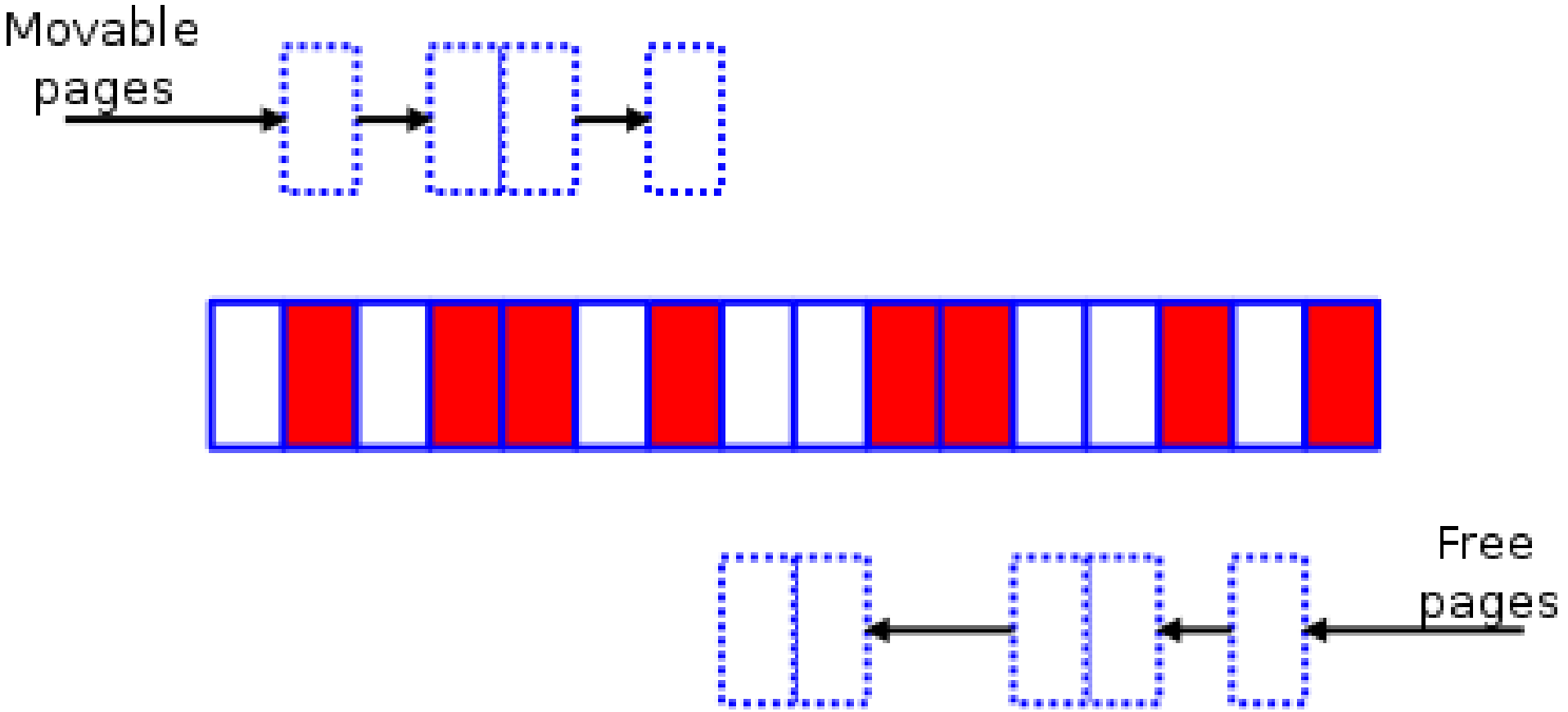
Compaction



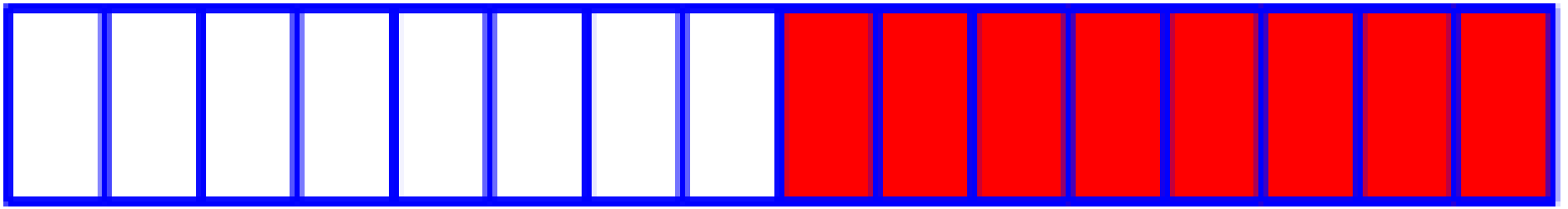
migrate_pages



Isolate_free_pages_block



Compaction



Решение:

vm.min_free_kbytes:

+kswapd раньше начнет работу

+больше свободной памяти

-меньше единовременно доступной памяти

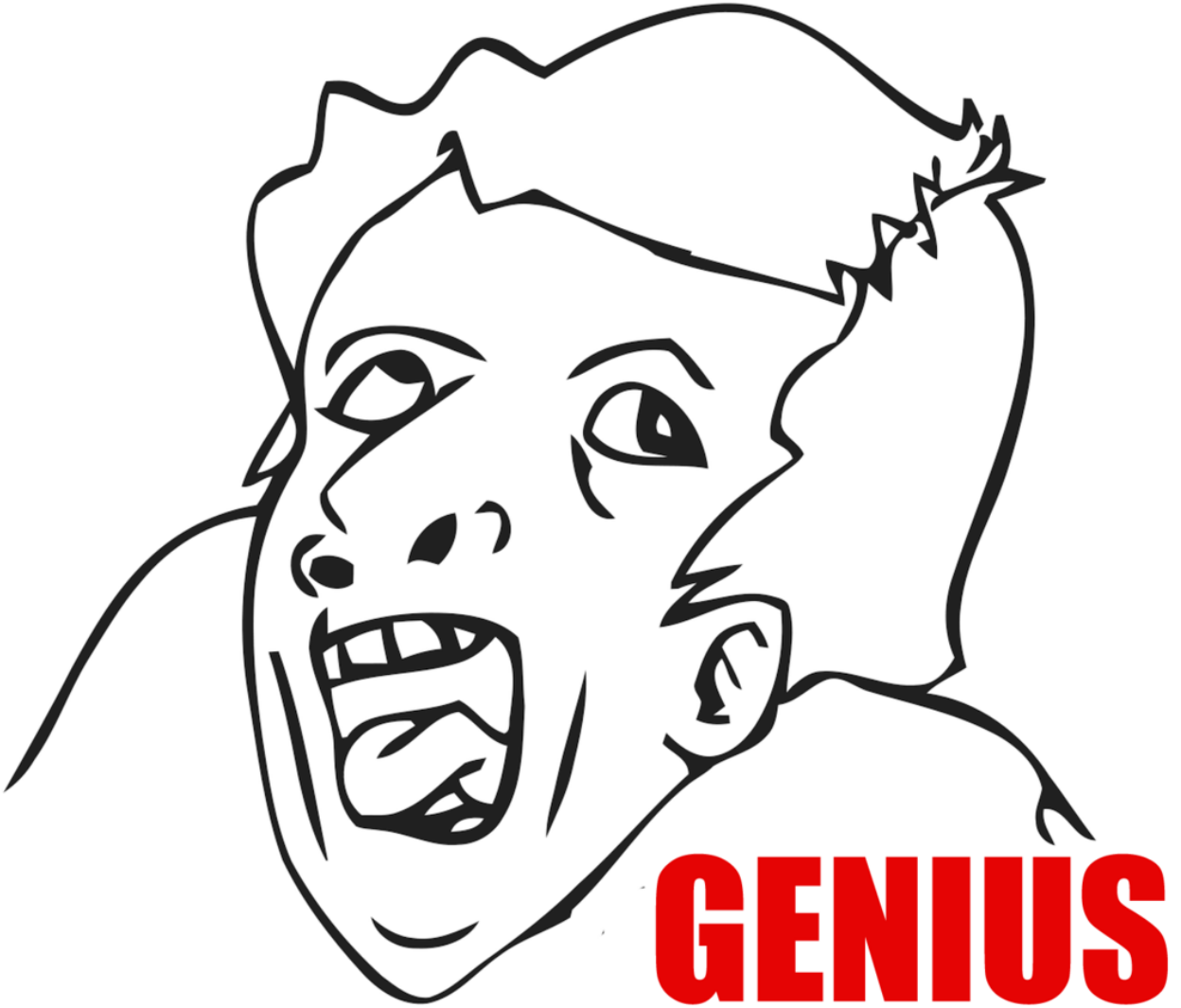
Временные таблицы

1. Привязаны к конкретному бэкенду
2. Не имеют механизма сбора статистики
3. Размещены в локальной памяти бэкенда(temp_buffers)
4. Удаляются после отключения клиента
5. НЕ ОТРАЖЕНО В ДОКУМЕНТАЦИИ: Постранично резервируют место на диске для страниц в temp_buffers

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
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



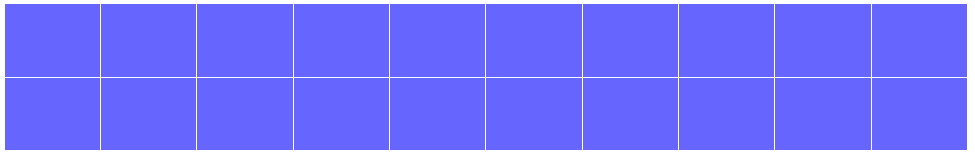
Postgres PROFESSIONAL

Почему это плохо

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

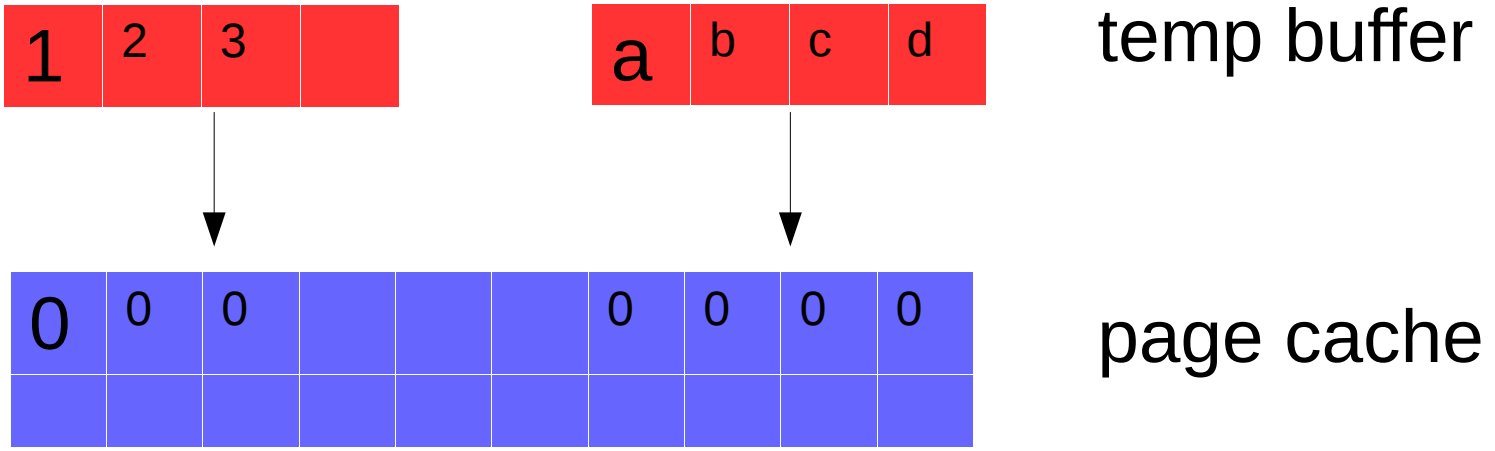
Буферизация

temp buffer

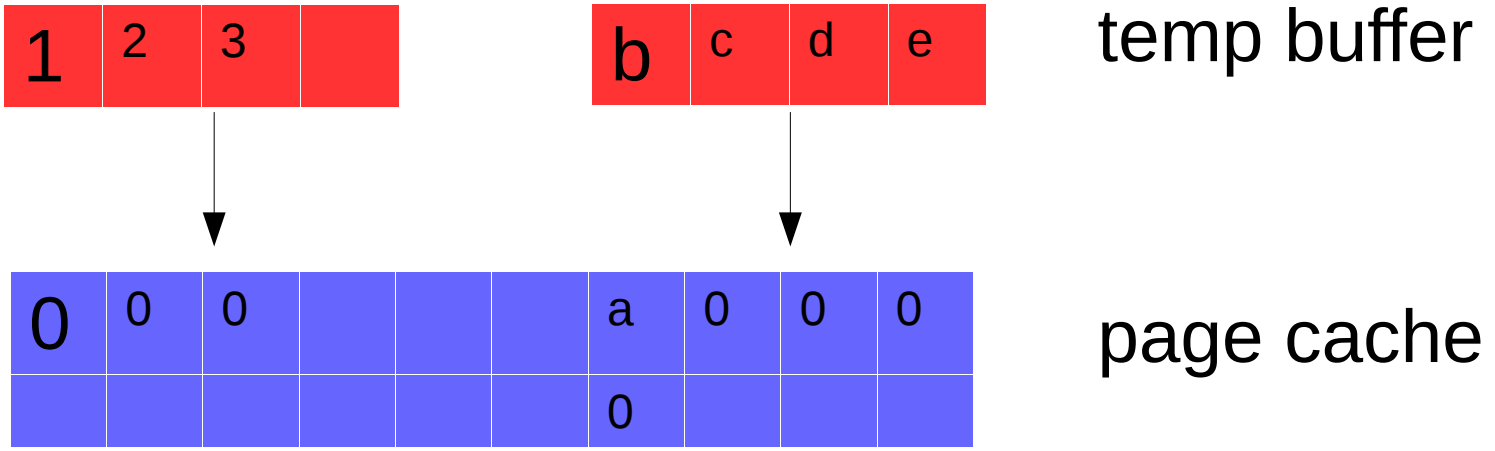


page cache

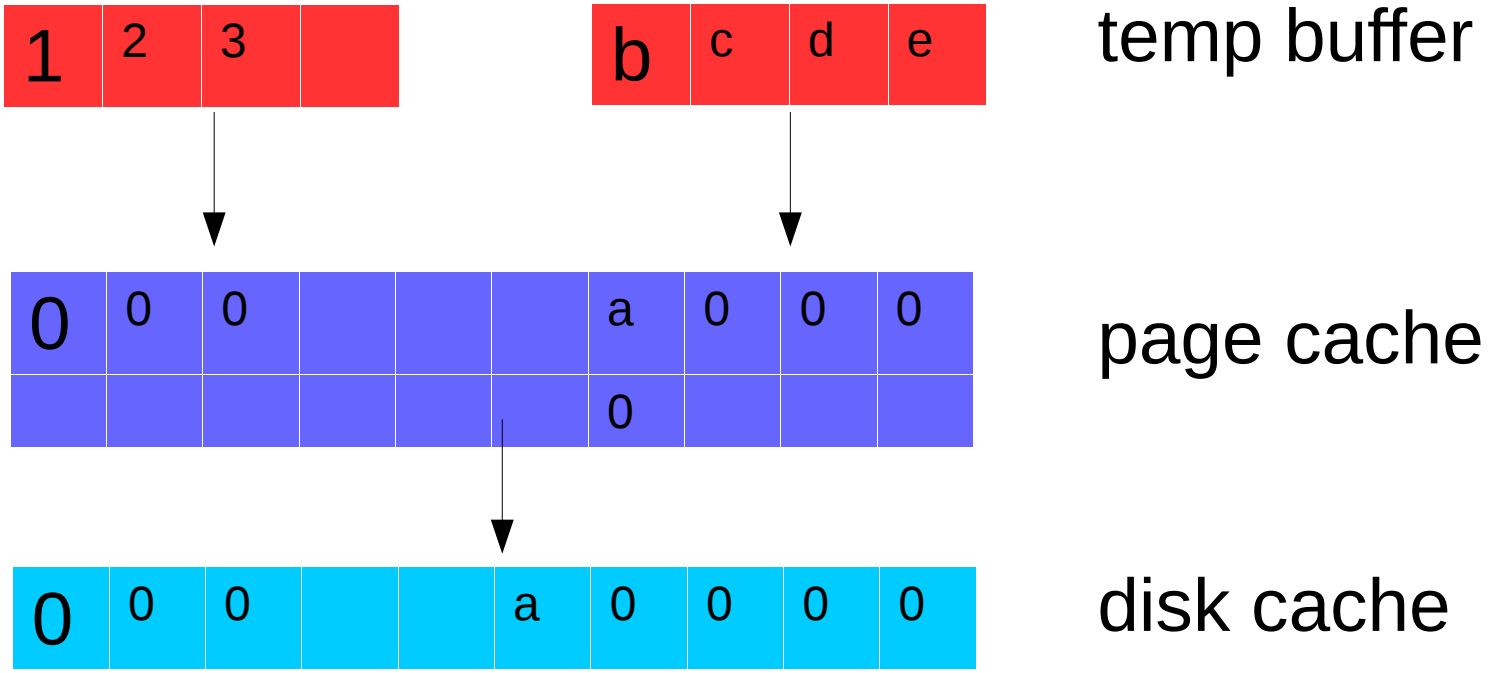
Буферизация



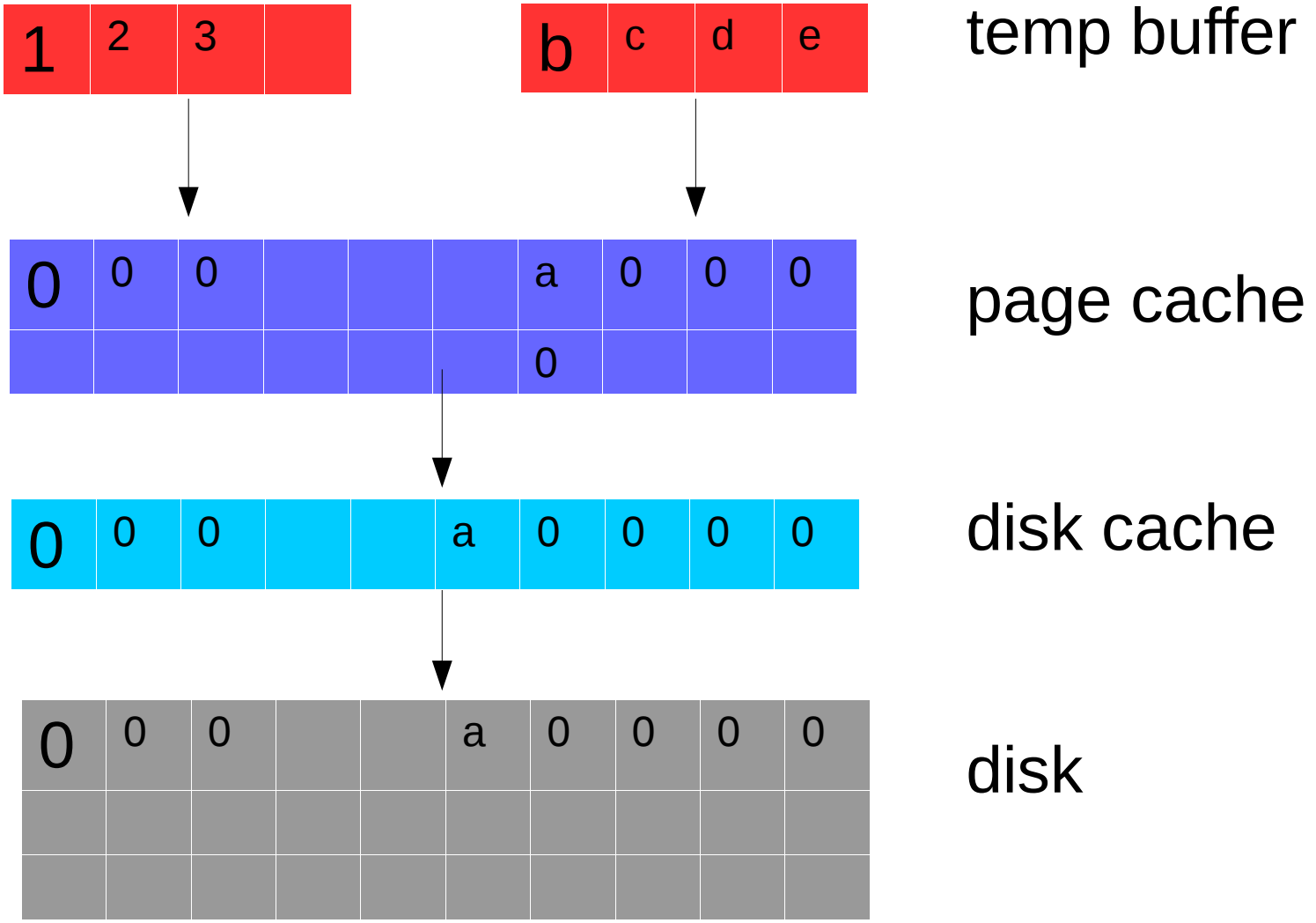
Буферизация



Буферизация



Буферизация



TMPFS

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

Мы отправились в отдел разработки

Патч на временные таблицы

Автор: Анастасия Лубенникова

- Решает все вышеуказанные проблемы
- Индексы,fsm, vm пишутся на диск
- Поведение временных таблиц стало соответствовать документации

Orphan temp tables

Условия появления:

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

Симптомы:

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

Orphan temp tables

Чем грозит?

1. Блоатинг каталога

2. Автовакуум не удаляет осиротевшие таблицы, но очень активно спамит в лог

Orphan temp tables

Что делать?

1. Увеличить max_locks_per_transactions
lock_table = max_locks_per_transaction * (max_conn + max_pred_locks_per_transaction)

2. DROP SCHEMA pg_temp_N CASCADE
DROP SCHEMA pg_toast_temp_N CASCADE

Мы отправились в отдел разработки

Orphan temp tables clean up patch

Автор: Константин Пан

- Бэкенд-процесс теперь корректно завершает свою работу
- Добавлена опция keep_orphan_tables, определяющая политику автовакуума относительно временных таблиц

temp_buffers

Постгрес не умеет возвращать системе память, аллоцированную под временные таблицы

Что делать?

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

online_analyze

Расширение, призванное решить проблему сбора статистики для временных таблиц:
1. В случае пишущего запроса принудительно делает analyze таблицы
2. Принимает решение на основе статистики, собранной stats collector`ом

online_analyze

Online_analyze

При большом кол-ве временных таблиц:

1. Чтение файла со статистикой становится дорогим

2. поиск по прочитанному хэшу становится дорогим

Online_analyze

Что делать?
1. Отключить :)

Postgres PROFESSIONAL

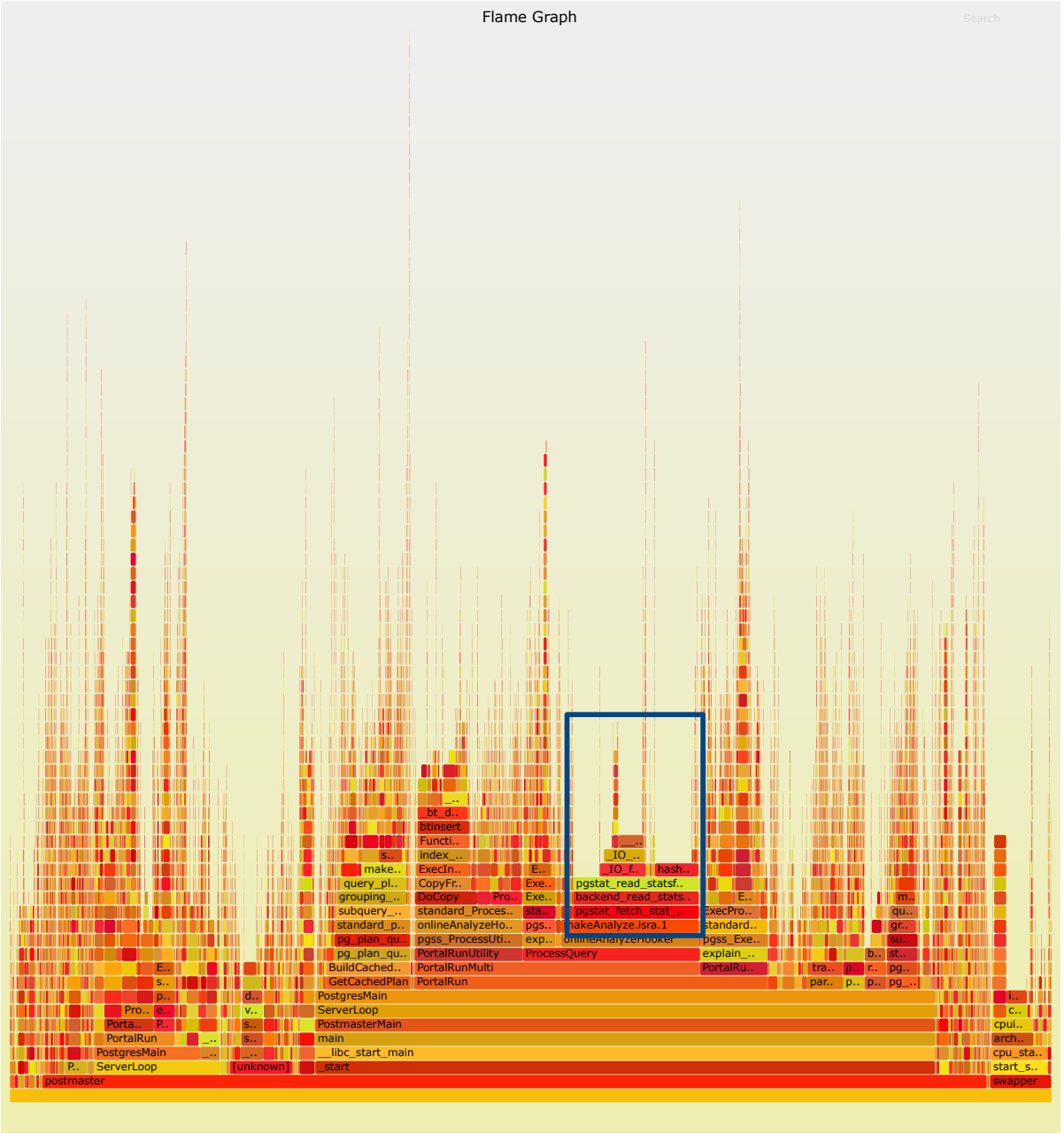
Патч для online_analyze

Автор: Фёдор Сигаев

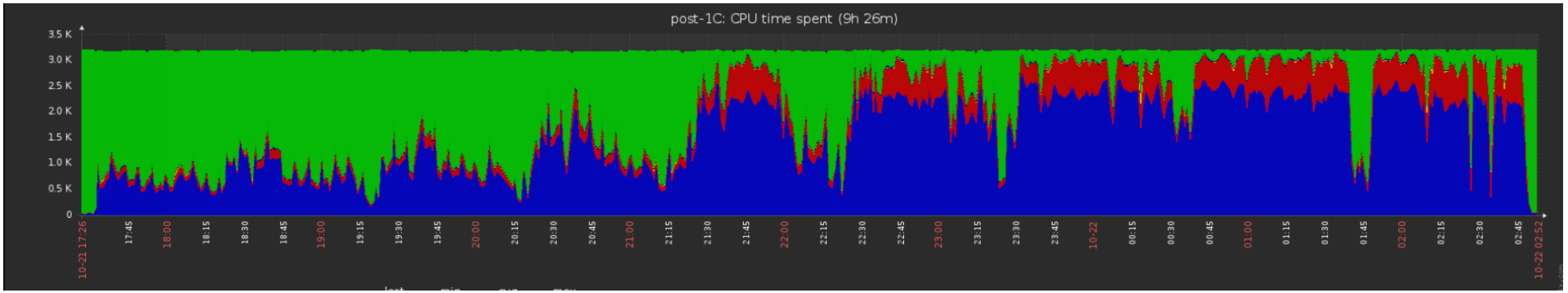
1. Изменена эвристика процедуры принятия решения о необходимости запуска analyze по таблице

2.Work in Progress

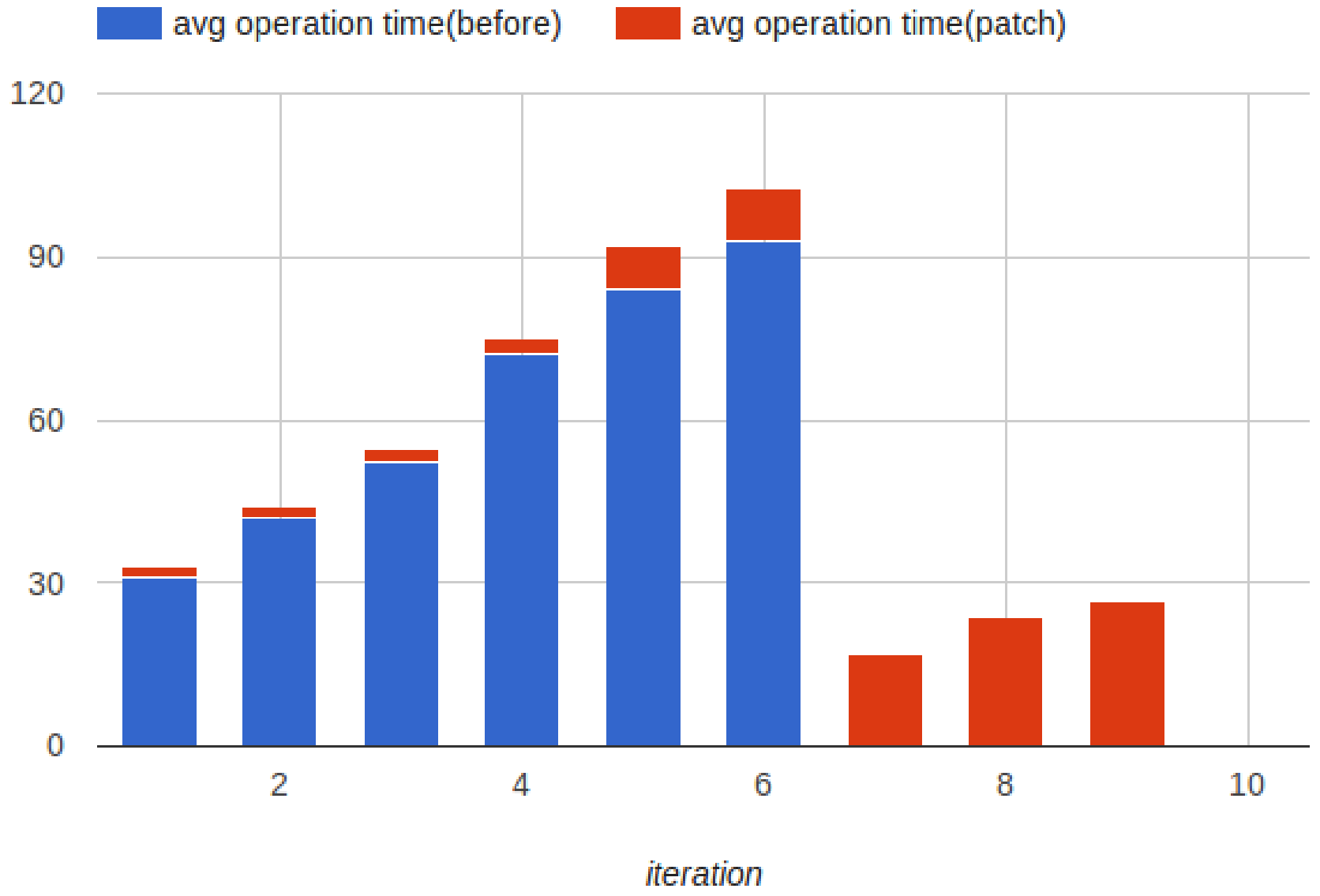
Патч на online_analyze



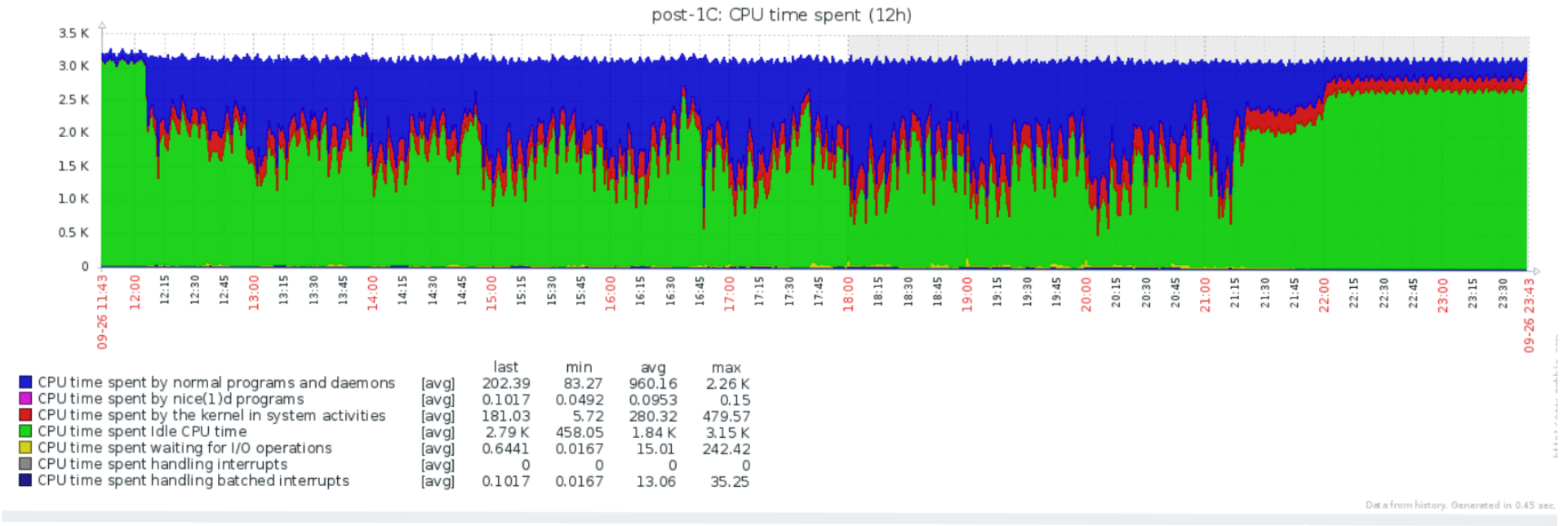
Online_analyze+fasttruncate патч



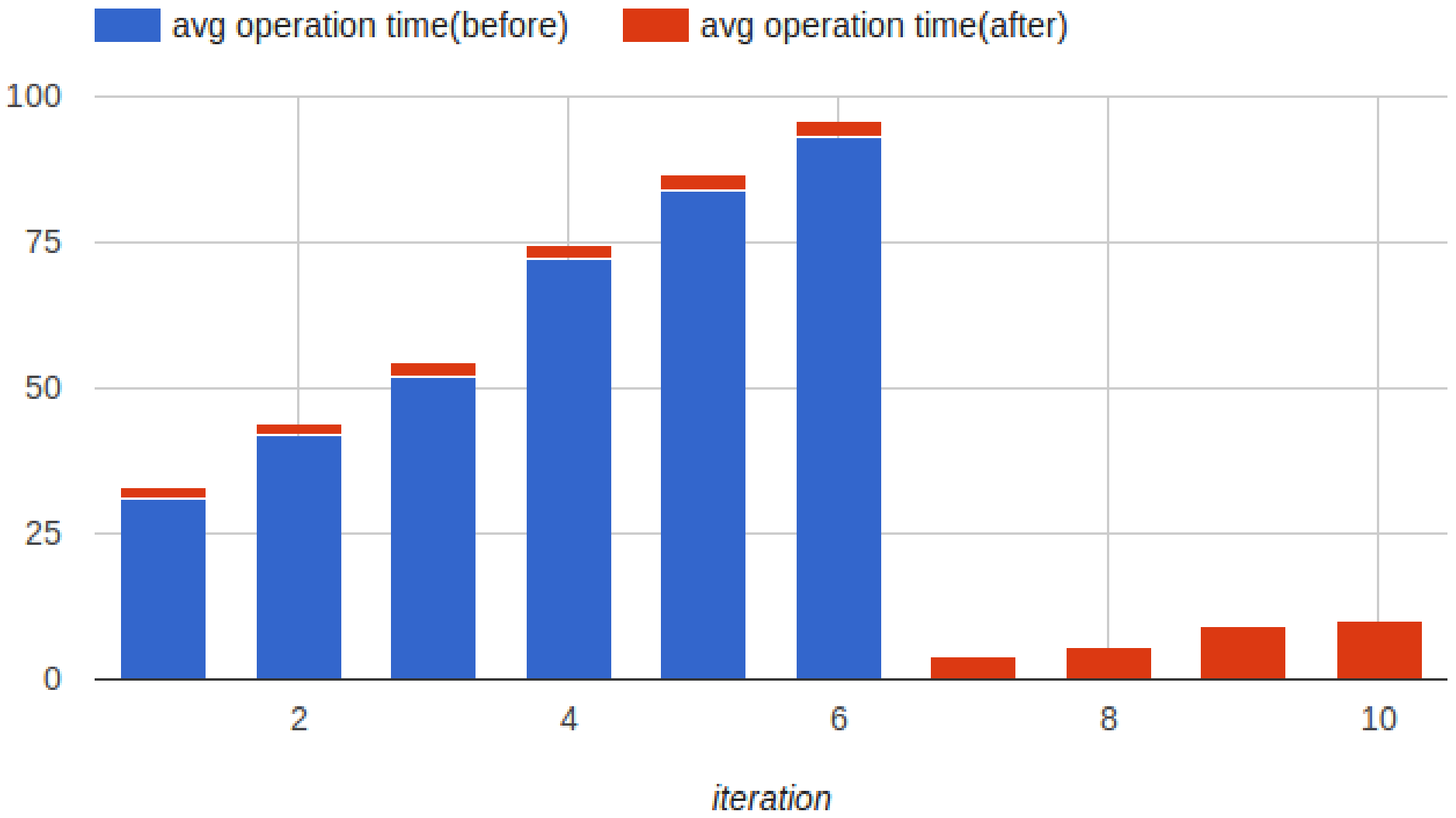
online_analyze+fasttruncate патч



online_analyze off



online_analyze off



Postgres PROFESSIONAL

Вопросы?

Postgres PROFESSIONAL

Спасибо за внимание!

Контакты:

g.smolkin@postgrespro.ru