



Концерн ГРАНИТ



ООО «КВАНТОМ» | 119019, г. Москва, Филипповский пер, 13 стр.1

ЗАЩИЩЕННОЕ ХРАНЕНИЕ ДАННЫХ БЕЗ ПОТЕРИ ПРОИЗВОДИТЕЛЬНОСТИ НА БАЗЕ СУБД КВАНТ-ГИБРИД

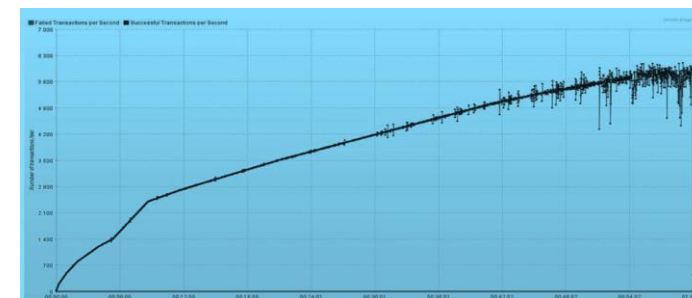
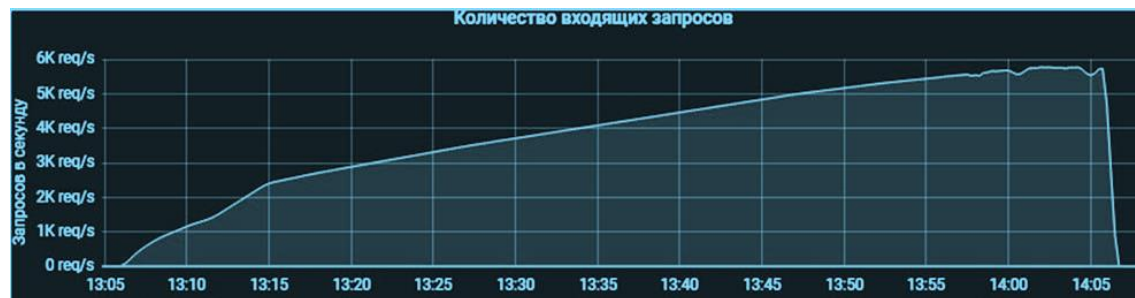
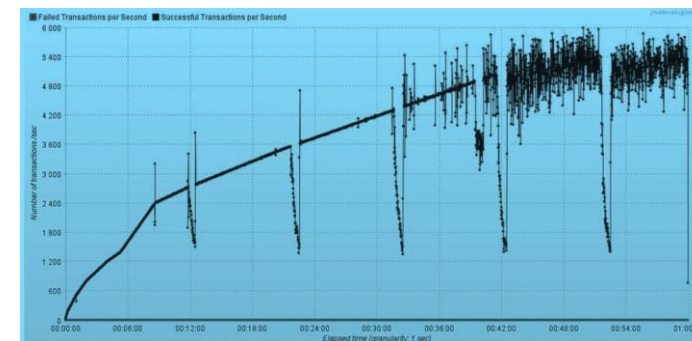
- Сертификат ФСТЭК № 4691 от 12 июля 2023 г.
- Сертификат ФСБ № СФ/124-4721 от 15 января 2024 г.

- I. Разрабатывается с 2019 г. АО «Концерн ГРАНИТ» на базе собственного Научного центра и дочерней компании системном разработчике ООО «КВАНТОМ»
- II. Основана на коде открытого ПО PostgreSQL с высокой степенью изменений и дополнений, сделанных на языке программирования Rust (20% кода ядра СУБД)
- III. Обеспечивает шифрование всех или части данных в фоновом режиме со стойкостью 10^{58} степени операций с использованием алгоритмов по ГОСТ РФ
- IV. Позволяет выдерживать высокие нагрузки на ограниченных технических средствах за счет встроенного балансировщика сетевой нагрузки с асинхронными очередями
- V. Содержит >10 функциональных расширений и программных модулей собственной разработки, включая встроенный расчет риск-коэффициентов на основе хранимых данных
- VI. Успешно протестирована рядом организаций и ведомств (РЖД, Новатэк, УВЗ, Фирма 1С, НРД, Гарда, КНС Групп, НЦИ, Гостех и др.)

- I. Обеспечивает шифрование хранимых и обрабатываемых данных по ГОСТ РФ с минимальным влиянием на производительность СУБД (до 5%)
- II. Предоставляет возможности расширенной диагностики СУБД за счет встроенной системы мониторинга с более чем 250 параметрами работы (при < 100 у конкурентных СУБД)
- III. Не содержит ряд мест, доставшихся от материнской технологии, блокирующих систему при резком увеличении нагрузки
- IV. Имеет функцию быстрой загрузки данных из внешних файлов со скоростью в 3 раза превышающую стандартную скорость конкурентных решений на базе PostgreSQL
- V. Поддерживает уникальный режим работы «только на запись», гарантируя невозможность подмены поступающих в систему данных на уровне ядра СУБД
- VI. Поддерживает битемпоральную модель данных, обеспечивая хранение данных, относящихся к прошлым, а также будущим периодам времени

- I. Балансировщик сетевой нагрузки для оптимального использования серверных подключений
- II. Самобалансирующийся менеджер кэша дисковых блоков с автоматической компенсацией нагрузки на дисковую систему
- III. Размещение таблиц для работы только в оперативной памяти (InMemory)
- IV. Работа с таблицами в режиме только "вставка" (AppendOnly)
- V. Встроенные алгоритмы блочного шифрования, включая ГОСТ РФ
- VI. Подсистема мониторинга и сбора метрик с минимальным влиянием на производительность
- VII. Инкрементальное резервное копирование и восстановление
- VIII. Прямая загрузка больших объёмов данных непосредственно в страницы данных
- IX. Поддержка битемпоральной модели хранения данных
- X. Трассировка долгих запросов
- XI. Внешнее хранение больших бинарных объектов с сохранением способа их обработки в прикладных системах и возможностью шифрования

За счет использования языка программирования Rust в разработке и рефакторинге кода ядра, Квант-Гибрид демонстрирует более стабильные нагрузочные характеристики в сравнении с материнской технологией

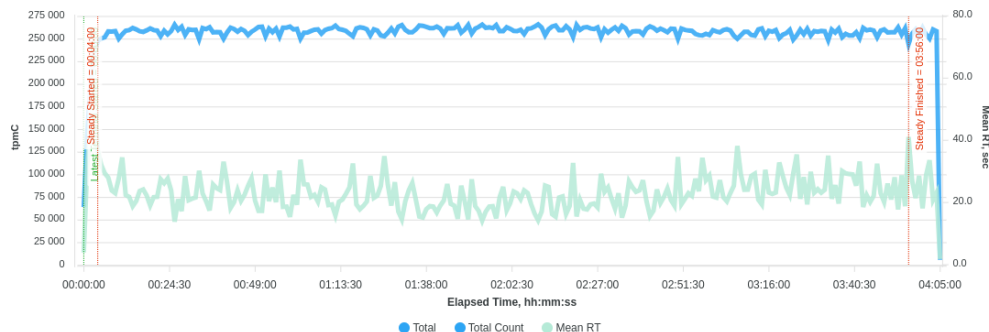


Результаты получены при нагрузке под реальным приложением в одном из проектов миграции на Квант-Гибрид

Результаты сравнительного НТ СУБД Квант-Гибрид с шифрованием всех данных и PostgreSQL по методике TPC-C

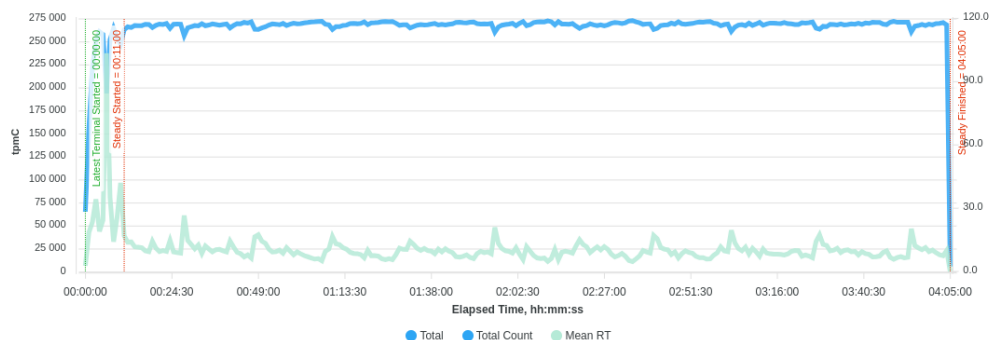
Характеристики сервера: CPU 56 ядер, RAM 500 Гб.

Параметры теста: Max_connections=3 000, Terminal_count=6 000



MQTh(tpmC)	113 980
Total tpmC (all transactions)	259 020
Transaction count	63 047 710

СУБД
КВАНТ-ГИБРИД
с шифрованием данных



MQTh(tpmC)	118 420
Total tpmC (all transactions)	269 138
Transaction count	65 412 676

СУБД Квант-Гибрид с шифрованием всех данных демонстрирует прирост производительности около 4% в сравнении с PostgreSQL

Результаты сравнительного НТ СУБД Квант-Гибрид с шифрованием всех данных и без с использованием HammerDB 4.11

Характеристики сервера: CPU 24 ядра, RAM 128 Гб.

Параметры теста: кол-во складов 1 000, использованы настройки Keying and Thinking Times* и Primary and Replica Modes **. Для организации пула соединений на стороне БД для уменьшения количества физических соединений использовался QCP. На уровне ОС был задействован функционал HugePages, позволяющий сократить накладные расходы ОС при использовании значительного объема памяти и большого количества соединений.

Сравнение по метрике NOPM (New Orders Per Minute, Новые заказы в минуту):

Число сессий	Без шифрования	С шифрованием	%, разница
40 000	49 881	47 352	5,07
50 000	57 657	58 119	-0,8
60 000	69 058	68 763	0,5

Сравнение по метрике TPM (Transactions Per Minute, Транзакций в минуту):

Число сессий	Без шифрования	С шифрованием	%, разница
40 000	121 367	115 692	4,6
50 000	142 571	142 303	1,8
60 000	170 595	170 303	0,7

Результаты демонстрируют влияние шифрования на производительность системы на уровне погрешности измерения показателей – от -0,8% до 5% (по метрике NOPM) и от 0,7% до 4,6% (по метрике TPM).

* Симуляция реальной нагрузки в системе, где пользователи затрачивают определенное время на ввод данных и делают паузы (<https://www.hammerdb.com/docs/ch04s07.html>)

** Создание максимальной нагрузки на сервер (<https://www.hammerdb.com/docs/ch08.html>)

QUANTUM CONNECTION POOL, QCP

АСИНХРОННЫЙ МАСШТАБИРУЕМЫЙ ПУЛЕР ДЛЯ СИСТЕМ С БОЛЬШИМ КОЛИЧЕСТВОМ ПОЛЬЗОВАТЕЛЬСКИХ СОЕДИНЕНИЙ

- Применяется для повышения утилизации соединений, их стабильности и общей производительности БД
- Обеспечивает подключение к СУБД Квант-Гибрид и поддержку связи с удаленными клиентами
- Переиспользует подключения на уровне сессий и в smart-режиме (разделение сессии между несколькими read-only соединениями)
- Автоматическое переключение на standby ноды
- Асинхронный режим работы и тредовая модель

МЕНЕДЖЕР КЭША ДИСКОВЫХ БЛОКОВ TARQ

САМОБАЛАНСИРУЮЩИЙСЯ МЕНЕДЖЕР КЭША ДИСКОВЫХ БЛОКОВ
С АВТОМАТИЧЕСКОЙ КОМПЕНСАЦИЕЙ НАГРУЗКИ НА ДИСКОВУЮ СИСТЕМУ

Алгоритм вытеснения из кэша используется в средах, где для ускорения операций с диском часть быстродействующей памяти используется для размещения наиболее часто востребованных блоков

Служит для определения блоков которые можно «выбросить» с наименьшими потерями для производительности, т.е. таких блоков, которые будут использоваться с меньшей интенсивностью в сравнении с остальными. Алгоритм обеспечивает устойчивость к вымыванию кэша

ОБЕСПЧИВАЕТ

- Устойчивость к вымыванию кэша
- Возможность построения In-Memory и Append-only таблиц
- Минимизация времени "разогрева кэша"

IN-MEMORY

- Гарантирует и существенно повышает вероятность нахождения информации в оперативной памяти экземпляра БД
- Увеличивает быстродействие системы за счёт снижения операций обмена с диском
- Гарантируется транзакционная целостность и обеспечение типа «хранилища» в двух режимах
- Вероятное хранение в оперативной памяти и хранение только в оперативной памяти

APPEND-ONLY

- Реализация типа хранилища, лишённого операций изменения и удаления – только запись и последующее чтение
- Достигается существенное ускорение операций за счёт упрощения транзакционной логики (отсутствует полноценный MVCC анализ)
- Является гарантом неизменения поступивших данных
- Отсутствует необходимость в автоочистке
- Поддерживаются все типы индексов

QUANTUM SECURE STORAGE, QSS

Программа предназначена для криптографической защиты конфиденциальности и целостности информации в СУБД Квант-Гибрид. Представляет собой программный продукт со встроенными библиотеками шифрования, консольными и программными интерфейсами, использующим в качестве внешних библиотек ПАК «Соболь» и Рутокен.

- Реализована возможность шифрования всех данных, отдельных таблиц, логов и дампов
- Доступно подключение и интеграция библиотек шифрования использующих аппаратное ускорение

КРИПТОГРАФИЧЕСКАЯ ЗАЩИТА СООТВЕТСТВУЕТ ТРЕБОВАНИЯМ

ГОСТ Р 34.13-2015, ГОСТ 34.13-2018, ГОСТ Р 34.12-2015, ГОСТ 34.12-2018, ГОСТ Р 34.11-2012, ГОСТ 34.11-2018 и стандартам Р 1323565.1.026–2019, Р 50.1.111-2016, Р 50.1.113-2016

QUANTUM SECURE STORAGE, QSS

ВОЗМОЖНОСТИ:

- Шифрование/расшифрование с генерацией/валидацией имитовставки (с использованием разделяемой памяти/ через сокеты)
- Авторизация по паролю с опциональным использованием криптографических ключей, хранящихся на отчуждаемых носителях
- Управление ключами (создание, удаление, блокирование, добавление)
- Хранение ключей в зашифрованном виде на диске
- Защита ключей: стирание ключей из оперативной памяти при штатном и нештатном завершении процесса
- Управление пользователями (добавление, удаление, редактирование)
- Получение списка подключенной разделяемой памяти
- Отсоединение разделяемой памяти администратором
- Смена и регенерирование ключа хранения

UNIVERSAL MEMORY CACHE ANALYZER, UMCA

Экземпляр QNB может работать достаточно долгое время. В процессе он накапливает данные о загруженных дисковых блоках. Если по каким-либо причинам появится необходимость остановить его работу, то эти данные будут потеряны. В этом случае после запуска новому экземпляру придется заново собирать эти данные.

Позволяет сохранять содержимое менеджера кэша дисковых блоков в файловой системе, а затем загружать его обратно, тем самым восстанавливая прежнее состояние. За счет этого время на прогрев экземпляра не затрачивается.

КАКУЮ ЗАДАЧУ РЕШАЕМ?

- Хотим знать что происходит "внутри" работающего экземпляра СУБД, какие операции самые продолжительные, как меняется продолжительность и количество операций со временем.
- Мониторинг должен минимально влиять на работу СУБД
- Хорошо бы иметь возможность включать\выключать мониторинг на лету
- Хорошо бы собирать доступные метрики операционной системы

METRICSD – СЕРВЕР МЕТРИК

- Сбор, агрегация и пересылка метрик QNB
- Предельно малое время сбора - практически не влияет на быстродействие сервера
- Существенное расширение числа параметров мониторинга
- Выгрузка метрик в графит, прометеус и вектор

ВАРИАНТЫ АНАЛИЗА ПРОИЗВОДИТЕЛЬНОСТИ ЭКЗЕМПЛЯРА

ТАБЛИЦЫ С ДАННЫМИ О ПРОИЗВОДИТЕЛЬНОСТИ (СОБИРАЮТСЯ STATSD)

- Низкая грануляция
- Таблицы должны быть запрошены через «обычный путь»
- Для вывода нужны плагины, агенты etc

ТОЧКИ ТРАССИРОВКИ (DTRACE) И СРЕДСТВА LINUX BPF

- Не готовое решение
- Требуется перекомпиляция
- Требуется дополнительных компетенций

METRICSD

- Включается\выключается по запросу
- Минимальное влияние на рабочую СУБД
- Внешний агрегатор для метрик самой низкой гранулярности

СХЕМА СБОРА МЕТРИК MetricsD

СУБД

При включении метрик определённые операции (функции) ядра СУБД передают временные и количественные значения в неблокируемую очередь, количество метрик одного процесса может достигать десятков тысяч в секунду

MetricsD

Агрегатор потребляет значения из очереди и консолидирует их в зависимости от природы наблюдаемой величины, после чего, передаёт потребителям. Также агрегатор собирает метрики с операционной системы

ПОТРЕБИТЕЛИ

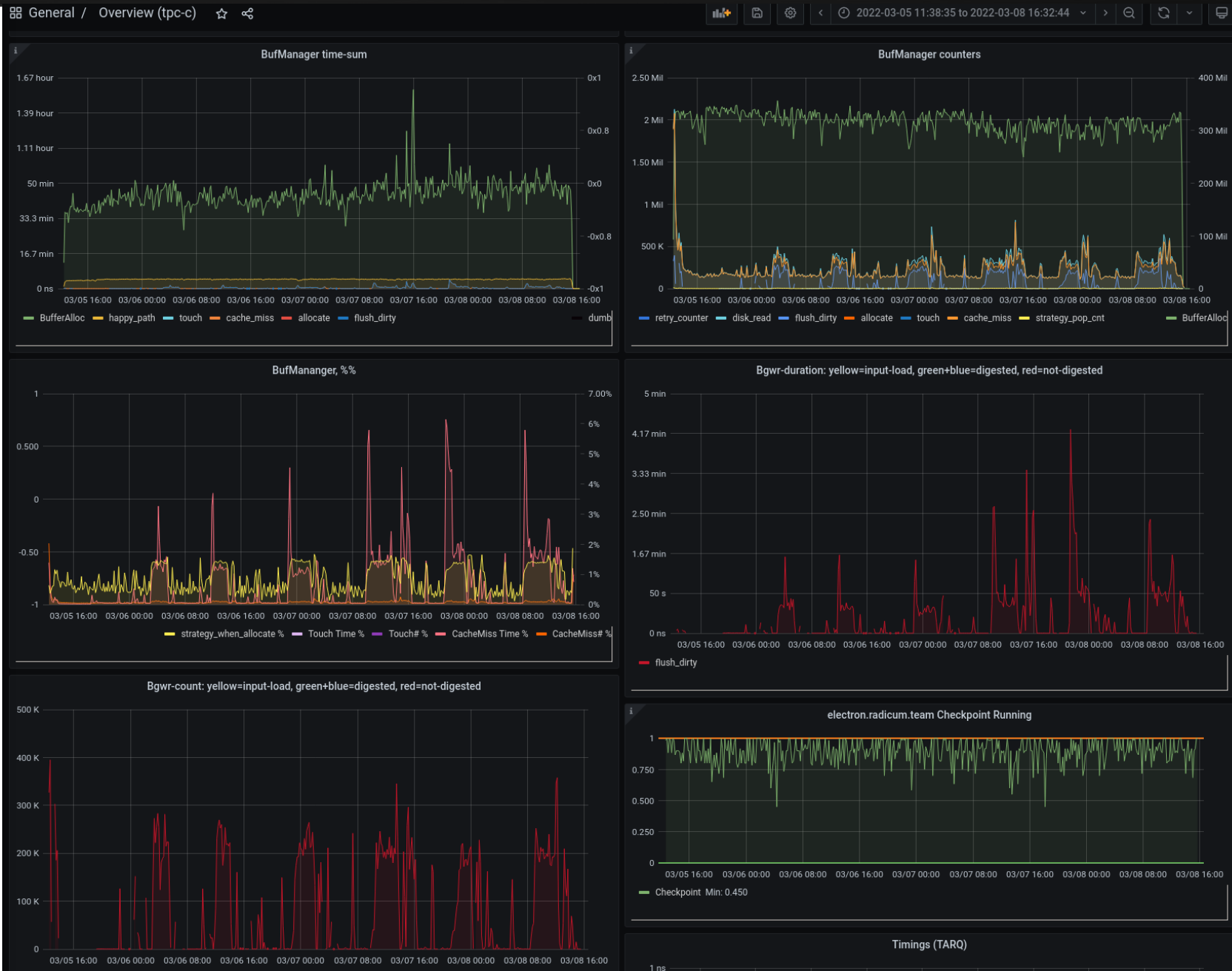
Получают консолидированные значения и используют полученные агрегаты для отображения пользователю



Метрики буфера страниц

- 03/07 в 16:00 видно увеличение времени аллокаций буферов (BufManger time-sum)
- При этом количество аллокаций не менялось (BufManager Counters)
- Из всего остального возросла активность Bgwr

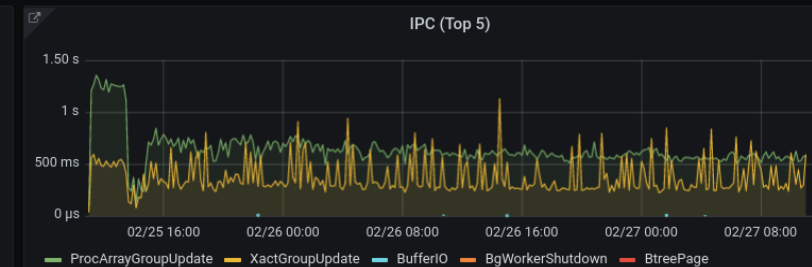
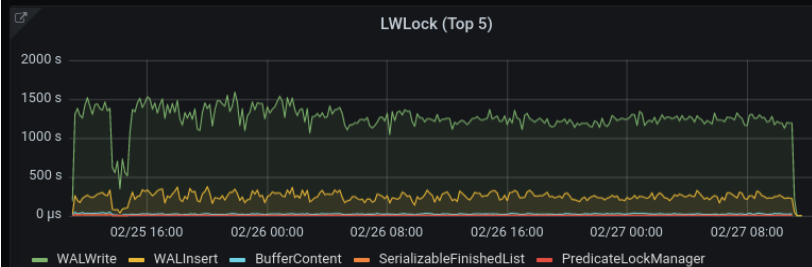
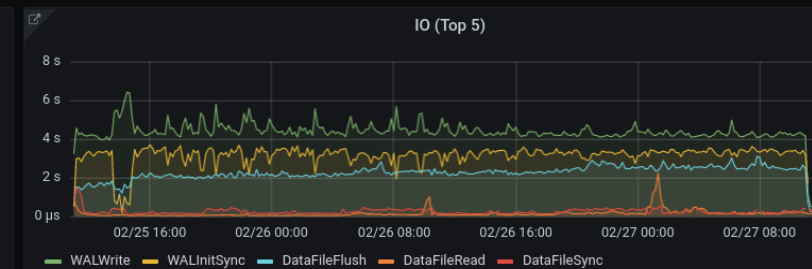
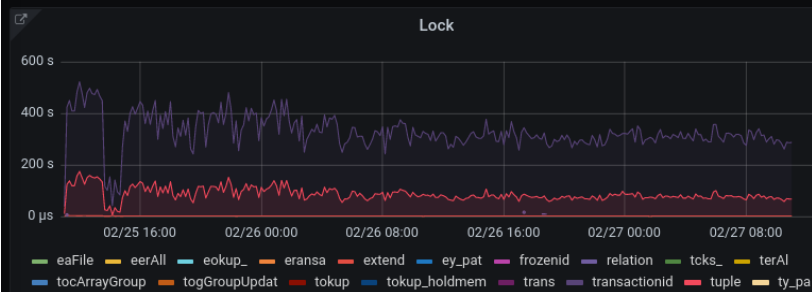
Вывод : в указанный период либо были штатные операции изменяющие большое количество блоков, либо деградация дисковой системы.



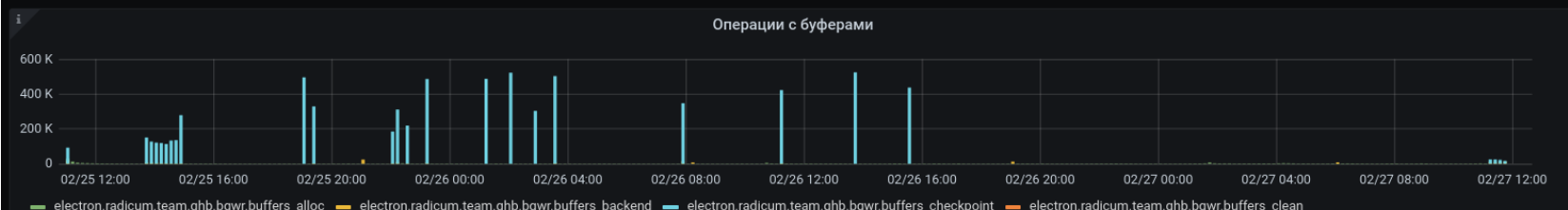
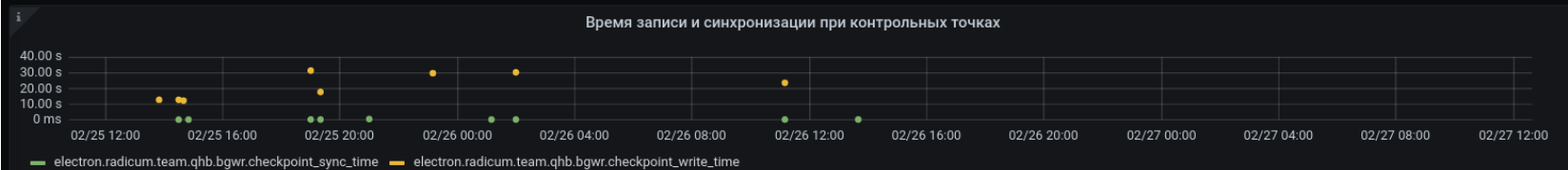
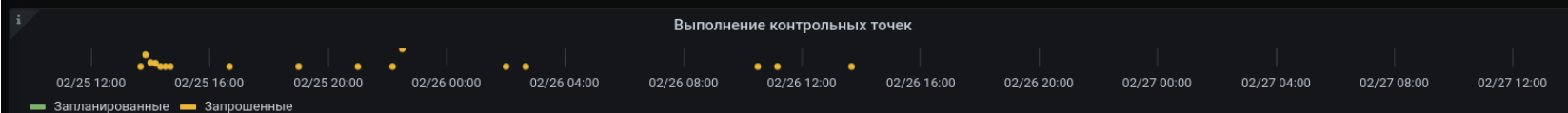
Блокировки и контрольные точки

- Lock — обычные (тяжёловесные) блокировки, больше всего времени уходит на транзакции (trans) и строки (tuple)
- LWLock — легковесные блокировки (защелки), больше всего времени занимает запись в журнал и буфер
- В 16:00 02/25 Наблюдался рост активности IO и падение конкуренции за журнал, скорее всего это массированный апдейт большой таблицы или перестроение индекса, об этом говорит и выросшее кол-во контрольных точек.

События ожиданий



Контрольные точки и операции с буферами



Метрики ОС

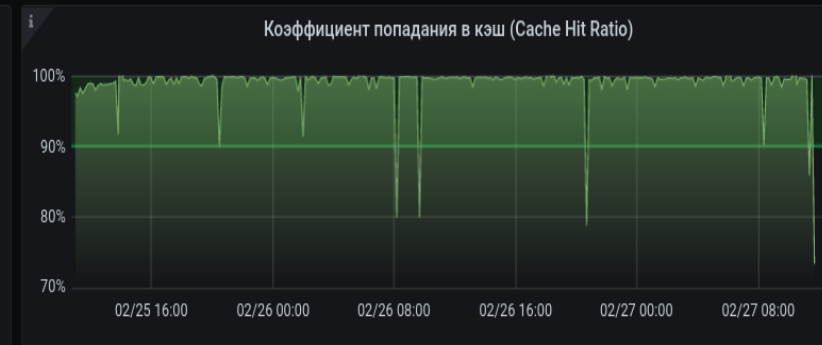
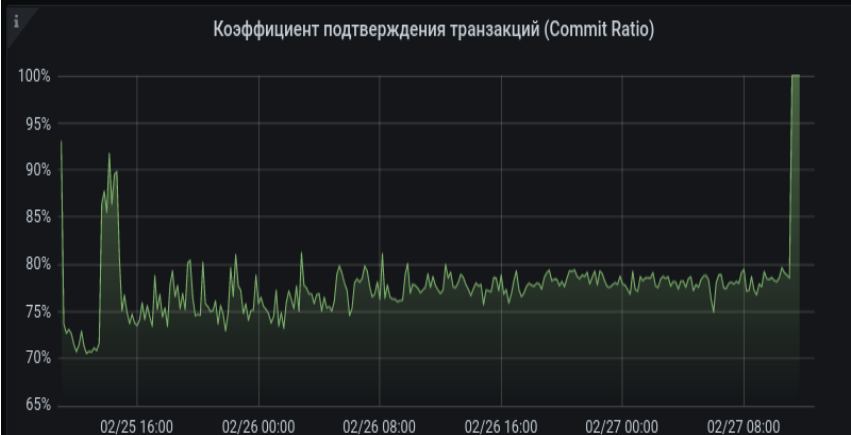
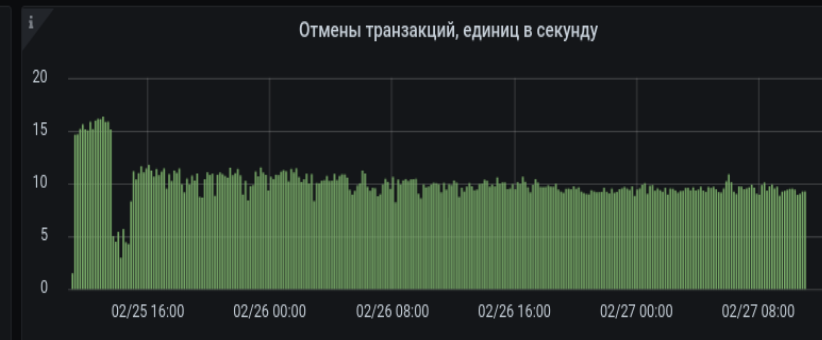
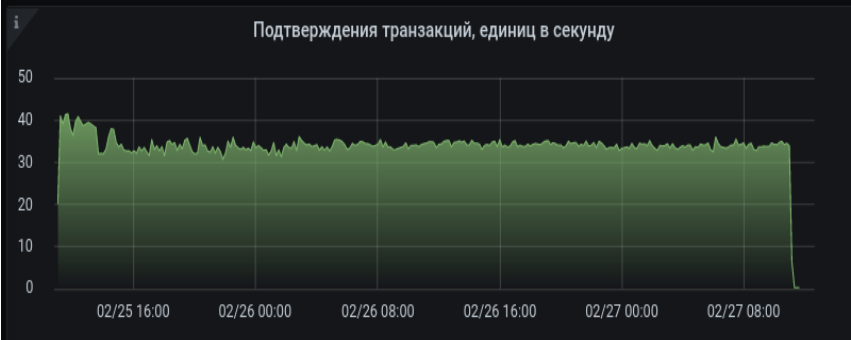
- Виден очень продолжительный «прогрев» 512Гб доступной памяти
- Скорость потребления дискового пространство при сохраняющейся нагрузке приведёт к недостатке свободного места примерно через 2-е суток



Метрики пользователя

- Метрики определены пользователем который под «транзакцией» подразумевал бизнес-операцию которая включает несколько фиксаций транзакции в БД
- Видно, что отмены и подтверждения, похоже независимы

Транзакции



Чтение и запись блоков



КАКУЮ ЗАДАЧУ РЕШАЕМ?

- Запрос выполняется драматически долго и непонятно, сумеем ли мы дождаться его завершения
- В запросе используется масса временных таблиц и представлений и его план невозможно получить при помощи обычного explain
- Запрос формируется динамически (так делать не стоит, но так часто делают)

SQL Profiler - ТРАССИРОВКА ДОЛГИХ ЗАПРОСОВ

- Обеспечивает отображение процесса выполнения запроса
- Использует встроенный брокер сообщений

СХЕМА УСТРОЙСТВА SQL TRACER

РАЗБОР ЗАПРОСА

Парсер разбирает входящий поток символов, превращает его в абстрактное синтаксическое дерево (Abstract Syntax Tree)

ПЛАН

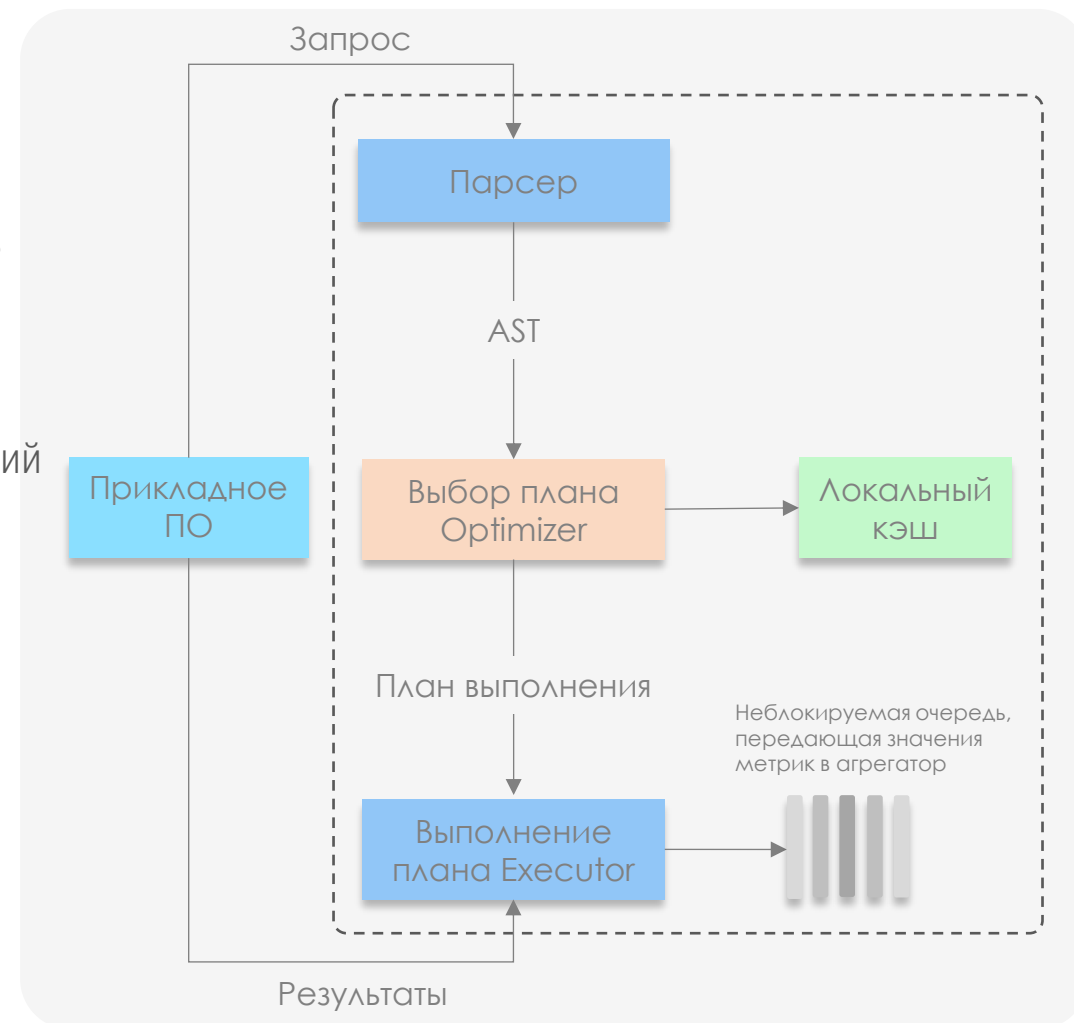
Так как SQL декларативный язык, то способов "выполнения" инструкций может быть несколько, планировщик выбирает пригодный на основе статистики

ИСПОЛНИТЕЛЬ

По выбранному плану запускает соответствующие процедуры для операций с данными

ТРАССИРОВКА

Каждый шаг планировщика передаёт статистику в неблокирующую очередь трассировщика



QUANTUM BACKUP, QBACKUP

РАСШИРЕНИЕ ДЛЯ РЕЗЕРВНОГО КОПИРОВАНИЯ ДАННЫХ БД

- Полное и инкрементальное резервное копирование без остановки экземпляра
- Параллельное многопоточное бэкапирование
- Сжатие резервных копий
- Удалённое резервное копирование
- Ускоренное восстановление (быстрее, чем воспроизведение WAL-логов)
- Восстановление на момент времени
- Ведение каталога резервных копий (в текстовом виде или JSON)
- CDC (change data capture)
- Создание резервных копий по табличным пространствам
- Валидация резервных копий
- Кумулятивный бэкап
- Удаленное восстановление

ВСТРОЕННАЯ ОЧЕРЕДЬ СООБЩЕНИЙ IN-MEMORY

- Очереди сообщений представляет собой подсистему СУБД для отправки типизированных сообщений и их получения заинтересованными (подписанными) клиентами.
- Существует два вида очередей: персистентные и in-memory. Если последние хранят сообщения до обработки всеми подписанными клиентами исключительно в памяти, то персистентные могут сохранять данные на диске, благодаря чему данные будут доступны после перезапуска базы.
- В Квант-Гибрид реализована in-memory очередь. Она хранит элементы в буфере фиксированного размера, и если тот заполнится, функция будет перезаписывать наиболее старые элементы

ВСТРОЕННЫЙ ПЛАНИРОВЩИК ЗАДАНИЙ QSCHEД

Пользовательские задания выполняются набором фоновых процессов, уникальных для каждой БД. Основой для работы планировщика является набор системных таблиц с префиксом `pg_catalog.qsched_`, который доступен для просмотра администратору базы данных и пользователю-владельцу БД.

Для создания и планирования заданий предусмотрено право `SCHED`, которое присваивается администратором базы данных пользователям, имеющим право работать с этой базой.

ВОЗМОЖНОСТИ

- Создание, управление расписаний / задач
- Ведение логов планировщика
- Включение / выключение динамического расчёта времени следующего выполнения
- Управление параллельными процессами планировщика

ПЕРЕМЕННЫЕ УРОВНЯ СЕССИИ QHB-VARIABLES

ПОЗВОЛЯЕТ СОЗДАВАТЬ ПЕРЕМЕННЫЕ, КОТОРЫЕ БУДУТ УДАЛЕНЫ ПО ЗАВЕРШЕНИИ СЕССИИ

- Механизм является альтернативой временным таблицам в местах, где их производительности недостаточно.
- Существует два типа переменных: скалярные и табличные. В скалярных переменных за раз можно хранить только одно значение любого типа, в табличных же можно хранить множество строк.
- Переменные также подразделяются на транзакционные и нетранзакционные. Состояние первых не зависит от результата транзакции, а вторые меняют своё значение учитывая COMMIT и ROLLBACK.

КАКУЮ ЗАДАЧУ РЕШАЕМ?

- Нужно максимально быстро загрузить данные из CSV в таблицу без влияния на производительность экземпляра СУБД
- В идеале сформировать таблицу с данными без экземпляра

QDL (Quantum Direct Loader)

- Загружает данные из внешнего источника непосредственно в таблицу СУБД
- Обход транзакционного ядра, отсутствие блокировок
- Многопоточная загрузка
- Ускорение в несколько раз

СХЕМА УСТРОЙСТВА QDL

БИНАРНАЯ УТИЛИТА

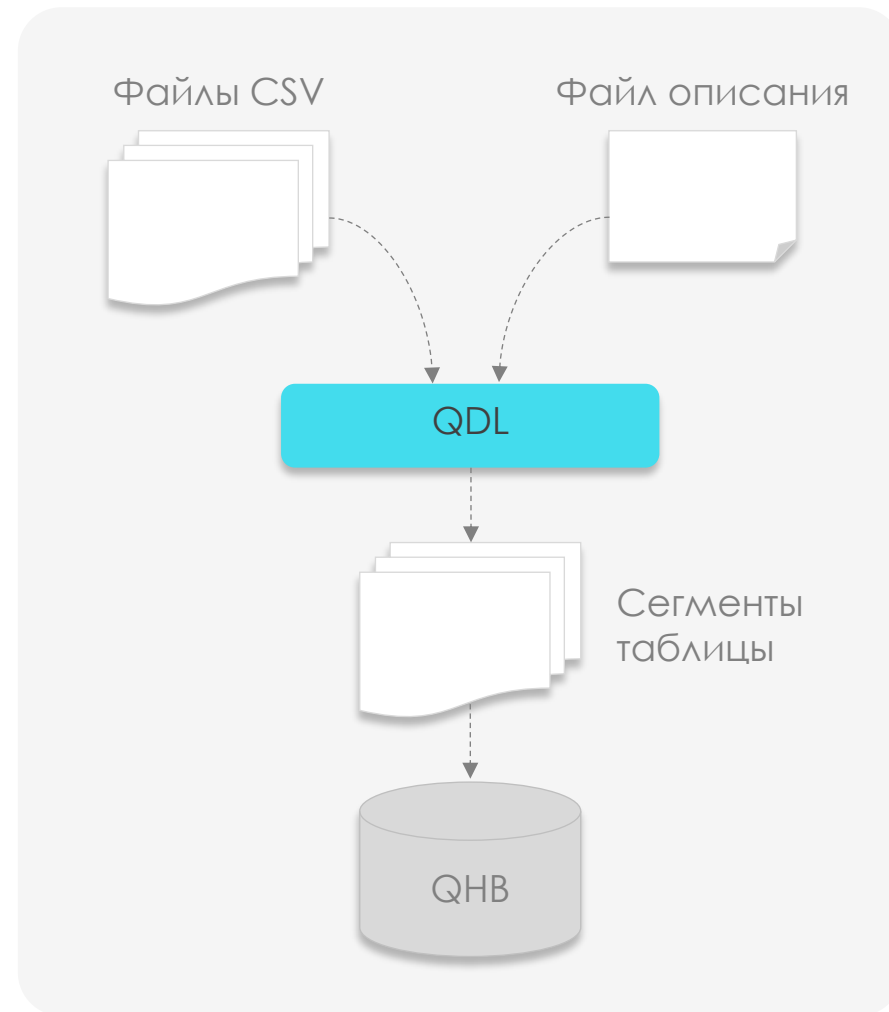
- QDL принимает на вход данные в виде файлов CSV и файл описания (типы полей + идентификаторы таблицы)
- QDL создаёт на диске сегменты таблицы - бинарные файлы

ПОДКЛЮЧЕНИЕ

- Бинарные файлы переносятся в каталог СУБД и подключаются набором команд

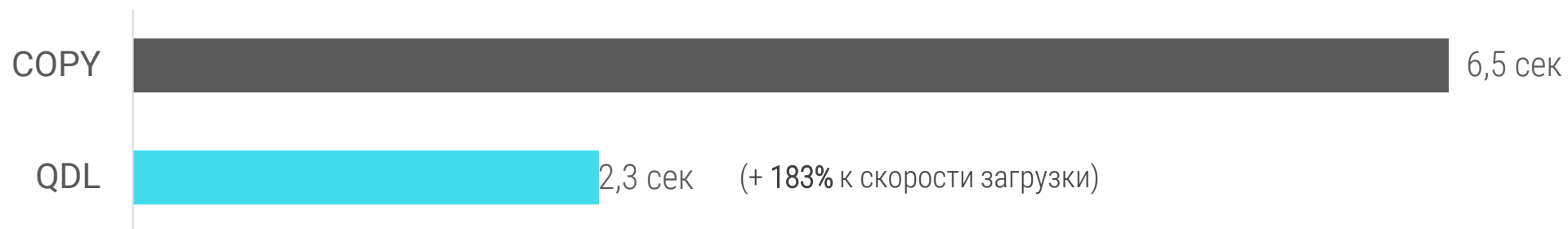
ВОЗМОЖНОСТИ

- Кроме стандартных типов, поддерживаются области переполнения (TOAST) и большие объекты.

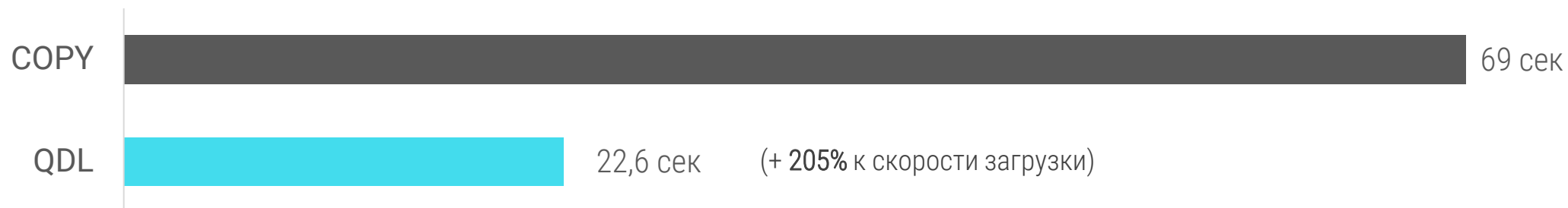


СПРАВНЕНИЕ QDL и COPY

- Загрузка небольшого объема данных (CSV, 313 Мб, ~900 тыс. строк)



- Загрузка существенного объема данных (CSV, 2,7 Гб, ~8 млн. строк)



ВНЕШНЕЕ ХРАНЕНИЕ ДВОИЧНЫХ ДАННЫХ RBYTEA

РЕАЛИЗУЮЕТ ХРАНЕНИЕ БИНАРНЫХ ОБЪЕКТОВ ВО ВНЕШНЕМ ХРАНИЛИЩЕ
С СОХРАНЕНИЕМ СПОСОБА ИХ ОБРАБОТКИ

- Обращение к бинарным объектам во внешнем хранилище реализовано также как и к данным СУБД (типы BLOB / CLOB / BYTEA)
- Расширение обеспечивает разгрузку транзакционных СУБД и упрощает их обслуживание
- Расширение совместимо с QSS - хранимые данные во внешнем хранилище могут шифроваться как и данные БД

Целью расширения является вынос двоичных данных из таблиц базы данных в нетранзакционное хранилище для разгрузки БД

БИТЕМПОРАЛЬНАЯ МОДЕЛЬ ДАННЫХ QVIM

РЕАЛИЗУЕТ ХРОНОЛОГИЧЕСКОЕ ХРАНЕНИЕ ДАННЫХ, ОТНОСЯЩИХСЯ К ПРОШЛЫМ, А ТАКЖЕ БУДУЩИМ ПЕРИОДАМ ВРЕМЕНИ

- Расширение Qvim представляет собой конструктор для разработки ERM на темпоральной основе, то есть, моделируемые объекты представляются временными срезами, последовательностями своих состояний во времени
- Основное внимание и акцент в данном расширении сделан не на "декомпозиции до конца", а на поддержании темпоральности данных
- Расширение реализует краткий набор логики хранения (создания, изменения) объектов в реляционной СУБД (на примере qhb), реализованный в слое хранимых процедур

РАСШИРЕНИЕ 2В

ОБЕСПЕЧИВАЕТ ПОДДЕРЖКУ ПРОДУКТОВ 1С

- Реализована в ядре QHB для возможности использования с «коробочными» продуктами 1С
- Дополнительные типы данных и операторы равенства для совместимости с Microsoft SQL
- Быстрое усечение временных таблиц и моментальное обновление статистики отдельных таблиц
- Поддержка указаний для планировщика, позволяющих отключать/подключать определённые индексы при выполнении запросов

Фирмой 1С было проведено тестирование успешное СУБД "Квант-Гибрид" с платформой «1С:Предприятие» версии 8.3.20 на совместимость.

ВСТРОЕННЫЙ В КВАНТ-ГИБРИД ИНСТРУМЕНТАРИЙ ДЛЯ ЗАДАЧ РАСЧЕТА ВЕРОЯТНОСТНЫХ МОДЕЛЕЙ С ПРИМЕНЕНИЕМ СЕТЕЙ БАЙЕСА

- Исходные данные для различных вычислений хранятся в таблицах и представлениях внутри QNB. К таким данным относятся исторические факты (срабатывание или несрабатывание гипотез) и соответствующие им причины (параметры)
- В качестве параметров рассматриваются логические (бинарные) значения или скалярные величины, которые методами расширения приводятся к дискретным значениям и разбиваются на диапазоны, определяемые пользователем. Значения гипотез всегда предполагаются логическими (бинарными)
- После предварительной статистической обработки исторических фактов (обучения), модель позволяет рассчитывать условные вероятности тех же самых гипотез при новых условиях (новых значениях параметров)

РЕАЛИЗОВАНЫ АЛГОРИТМЫ

■ Наивный байесовский классификатор

Простой классификатор, основанный на предположении о независимости параметров

■ Классификатор, основанный на функции взаимной информации

Взаимная информация — это статистическая функция двух случайных величин, описывающая количество информации, содержащееся в одной случайной величине относительно другой. Метод позволяет определить для гипотез зависимые параметры и в дальнейшем использовать эти зависимости для предсказания гипотез

■ Методы Монте-Карло с цепями Маркова

Класс алгоритмов, моделирующих некоторое распределение вероятностей. В расширении Qbayes для установления зависимых параметров для сэмплирования гипотез применяется реализация алгоритма Метрополиса — Гастингса