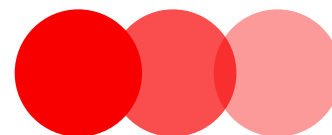


Разработка корпоративной платформы отчетности и BI

Вопросы производительности



Владимир Сухов

Руководитель сектора разработки BI и Big Data, розничная сеть «Магнит»

Что делает Магрепорт?



Предоставляет
пользователю
интерфейс задания
значений фильтров
отчёта



Формирует
запрос к БД



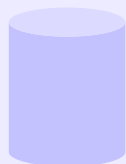
Взаимодействует
с БД



Сохраняет
выгруженные
данные



Предоставляет
пользователю
возможность
работы
с выгруженными
данными



БЫСТРАЯ БД ДЛЯ
СПРАВОЧНИКОВ



БД С ДАННЫМИ
ОТЧЁТОВ

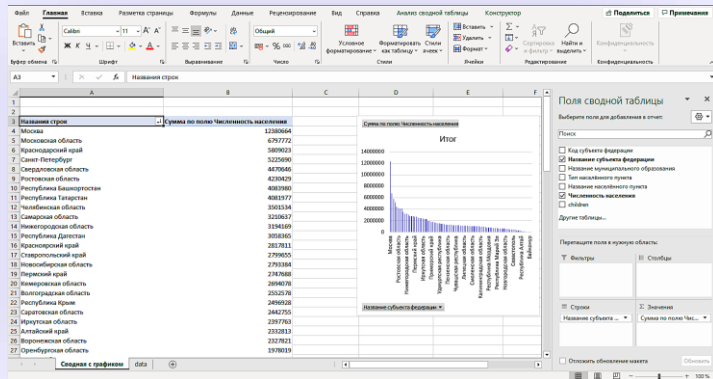


ДИСКОВЫЙ РЕСУРС

Формат AVRO:
3-10 сжатие по сравнению
с CSV экономия пространства и IO

Работа с выгруженными данными

1 ЭКСПОРТ В EXCEL



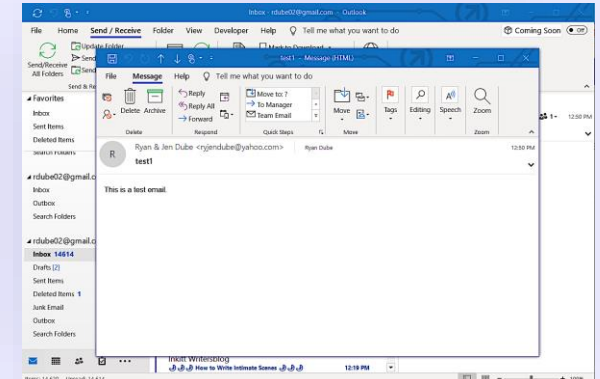
↓
Стандартный
шаблон

↓
Специфичный
шаблон

2 РАБОТА С OLAP

Название субъекта Федерации	Название муниципального образования	Тип образовательного пункта	Сумма Численность населения
Алтайский край	Алтайск	Сумма Численность населения	29249
Алтайский край	Барнаул	Сумма Численность населения	633432
Алтайский край	Белогорск	Сумма Численность населения	15160
Алтайский край	Бийск	Сумма Численность населения	200576
Алтайский край	Заринск	Сумма Численность населения	46597
Алтайский край	Зеленогорск	Сумма Численность населения	10321
Алтайский край	Катавский	Сумма Численность населения	40795
Алтайский край	Летовский	Сумма Численность населения	12654

3 ОТПРАВКА ОТЧЁТА ПО ЭЛЕКТРОННОЙ ПОЧТЕ



Аспекты разработки

Концепция

Безопасность

Администрирование

Архитектура

Производительность

Простота и удобство

Функциональность

Эволюция

Производительность: составляющие

Качество формируемых SQL-запросов к СУБД

Управление собственными ресурсами системы и управление конкурентной нагрузкой пользователей

Эффективность реализации высокоинтенсивных вычислений

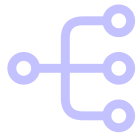
Эффективность работы с собственным репозиторием метаданных

Эффективность алгоритмов файлового экспорта

Эффективность решений фронтенда:

- Кэширование
- Пейджинация
- Эффективный рендеринг многокомпонентных объектов

Качество SQL-запросов



Партиционирование



Индексы



Суррогатные ключи

Использование ID вместо бизнес-ключей

Список бизнес-ключей

123746, 647382, 436473,
726482, 373929, 102945

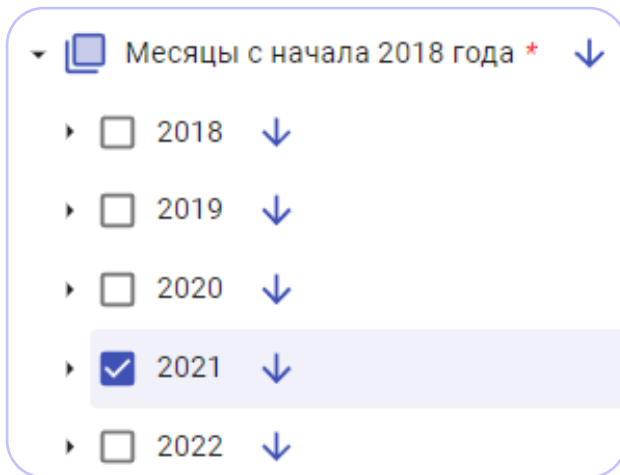


ТОВАР_ID IN
(14, 17, 3, 29, 54, 41)

Результаты экспериментов на СУБД Терадата:

- Потребление CPU: сокращение в 3 – 3,5 раза
- Потребление IO: сокращение в 1,5 – 2,5 раза
- Потребление Spool: сокращение в **десятки** и **сотни** раз

Разыменование партиций



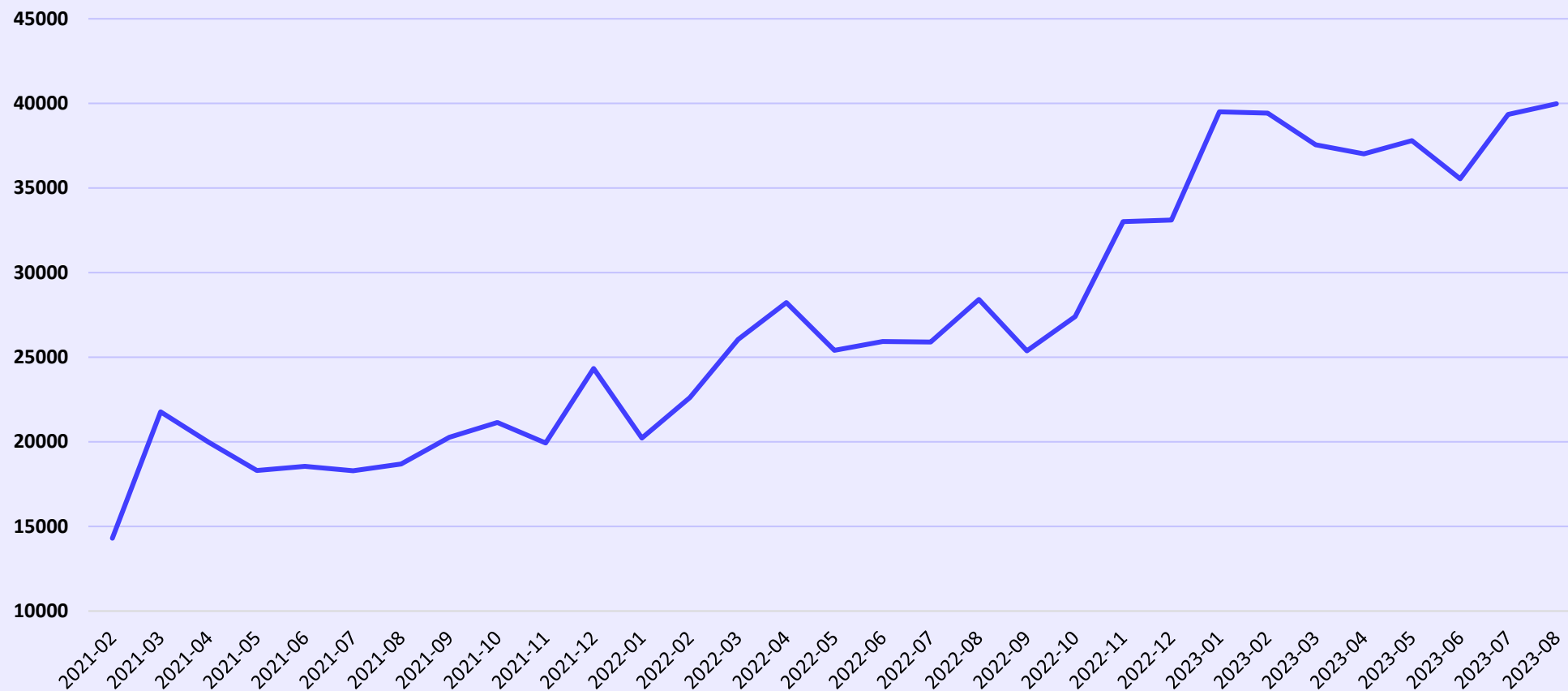
МЕСЯЦ_ID IN
(202101, 202102, 202103, 202104, 202105,
202106, 202107, 202108, 202109, 202110,
202111, 202112)

Результаты экспериментов на СУБД Терадата:

- Потребление CPU: сокращение в 5 раз
- Потребление IO: сокращение в 5 раз
- Потребление Spool: сокращение в **10 раз**

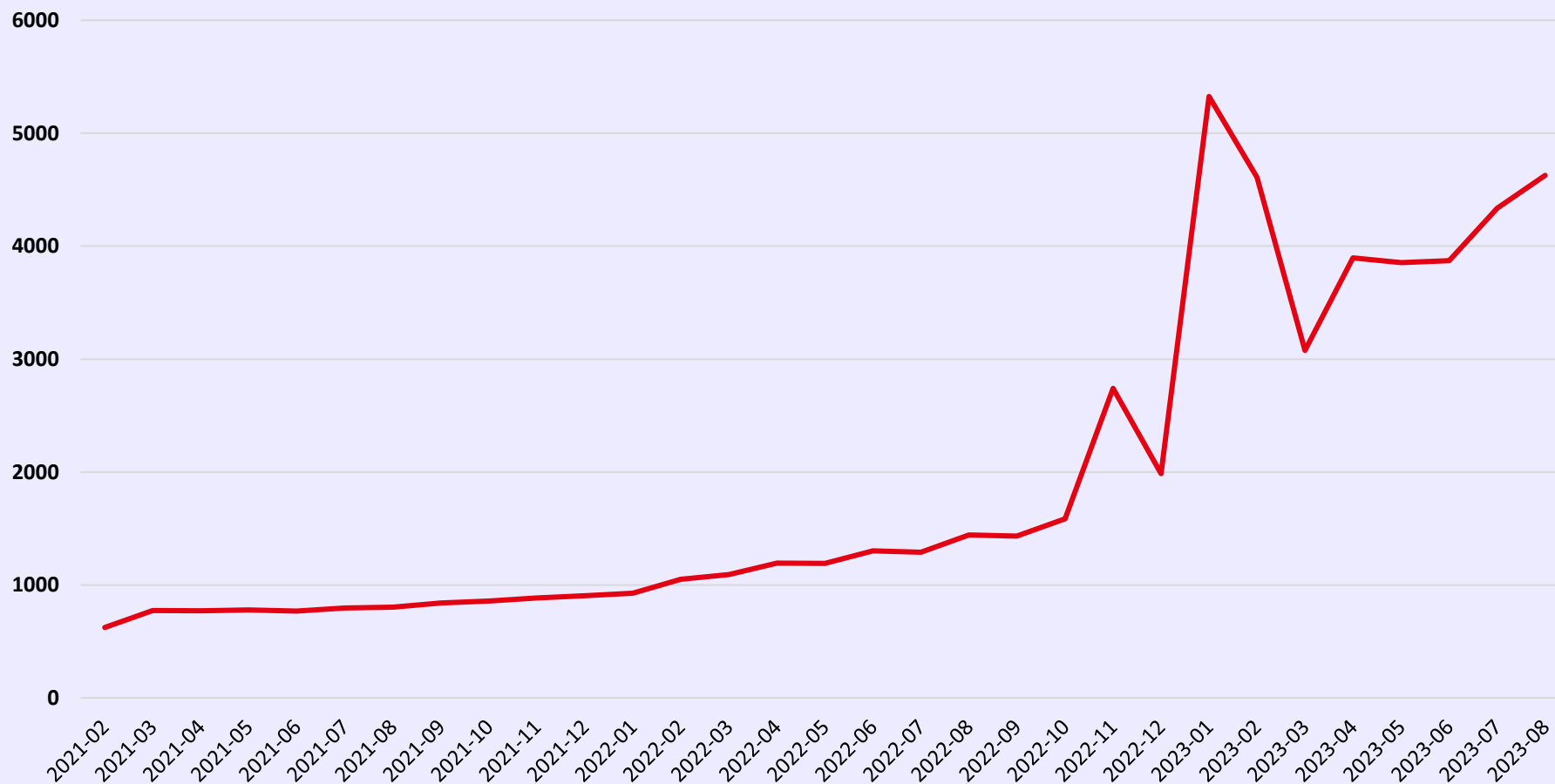
Конкурентная нагрузка пользователей

Кол-во отчётов в месяц



Конкурентная нагрузка пользователей

Количество пользователей в месяц



Конкурентная нагрузка пользователей

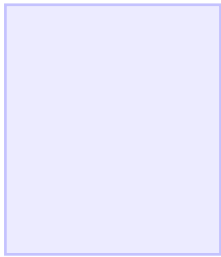
Пиковая нагрузка:

300 – 600
отчётов в час

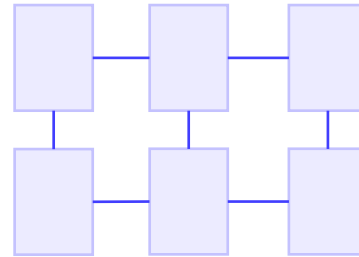
Экстенсивный и интенсивный ответы

Экстенсивный ответ: масштабирование

Было



Стало



Преимущества

- Технологично
- Управляемо
- Предсказуемо

Интенсивный ответ: оптимизация

Было



Стало



Преимущества

- Экономично

Конкурентная нагрузка пользователей

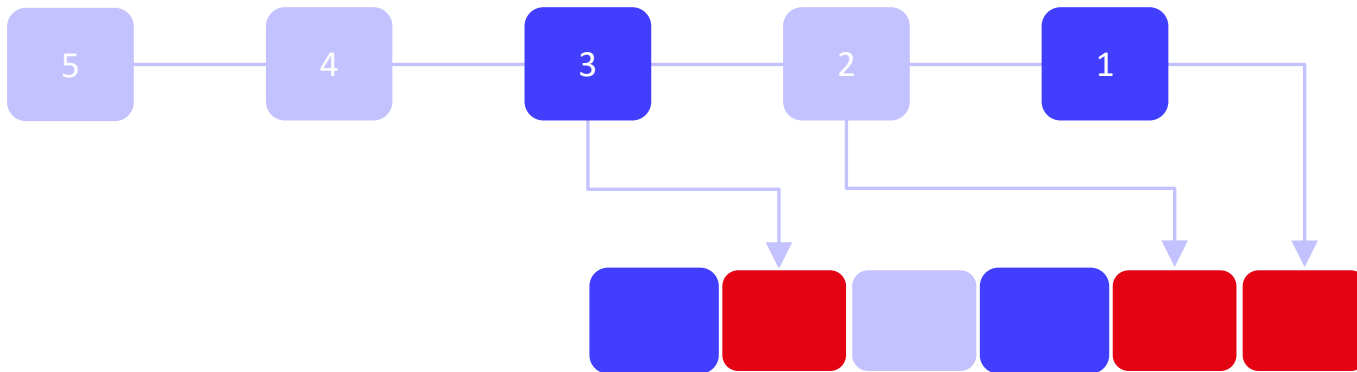
СИНХРОННЫЕ ЗАПРОСЫ



АСИНХРОННЫЕ ЗАПРОСЫ



Система очередей и ограничений



Очереди

- Очередь выполнения запросов к БД (8)
- Очередь экспорта полученных данных в Excel (16)

Ограничения

- Количество одновременно загруженных в память OLAP-кубов (3)
- Количество одновременно выполняющихся OLAP-запросов (3)
- Количество одновременных экспортов сводной в Excel (2)

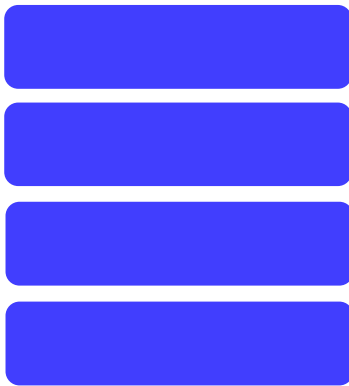
В скобках указаны действующие значения для нашего сервера с характеристиками:

- CPU 8 cores Intel Xeon GOLD 6154 3GHz
- 24 GB RAM

Пулы коннектов к СУБД

Пулы коннектов к субд

Источник
данных



Отдельные коннекты для справочников и для отчётов

Справочники

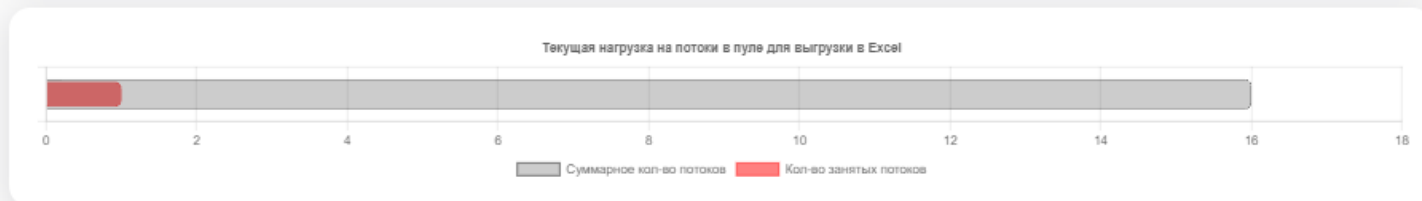
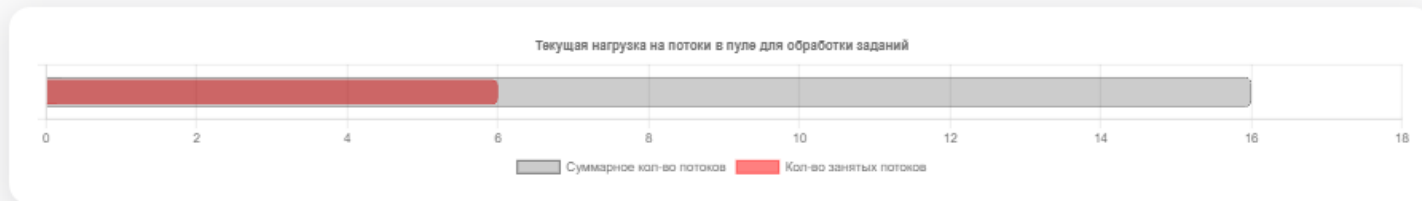


Отчёты

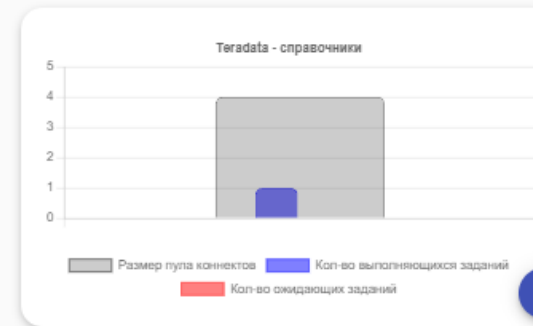
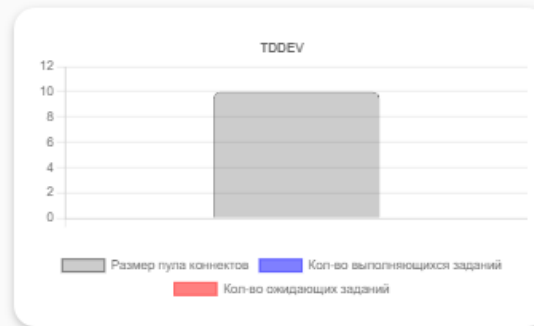
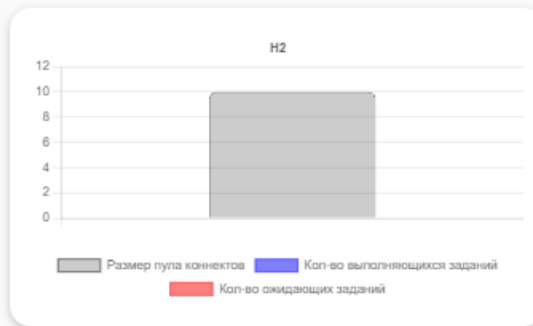
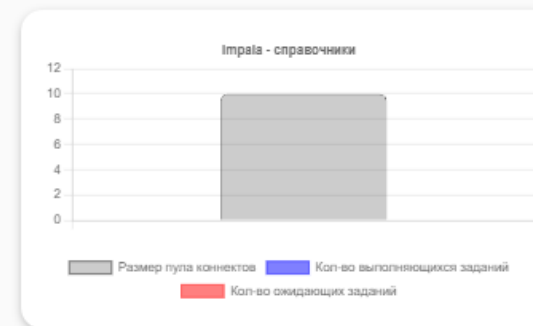
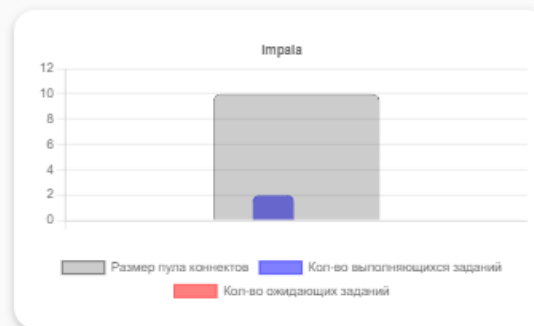
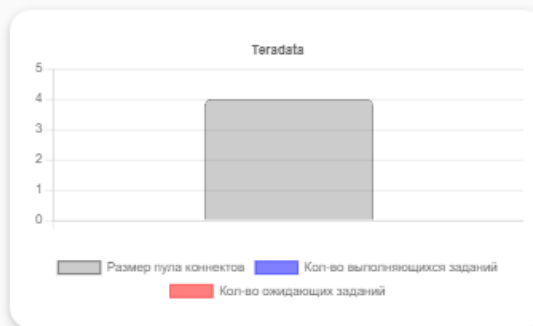


Мониторинг нагрузки

Мониторинг нагрузки



В разрезе каждого источника данных



Высокоинтенсивные вычисления

Обработка olap-запроса

Месяц	Магазин	Группа	Товар	Продажа	Потеря	Остаток
202201	M1	Г1	T1	3	1	7
202201	M1	Г2	T2	4	2	8
202202	M2	Г1	T3	1	0	5
202203	M3	Г4	T4	2	0	6
202201	M4	Г3	T5	5	0	8
202202	M3	Г5	T6	7	0	7
202204	M2	Г6	T7	8	1	6

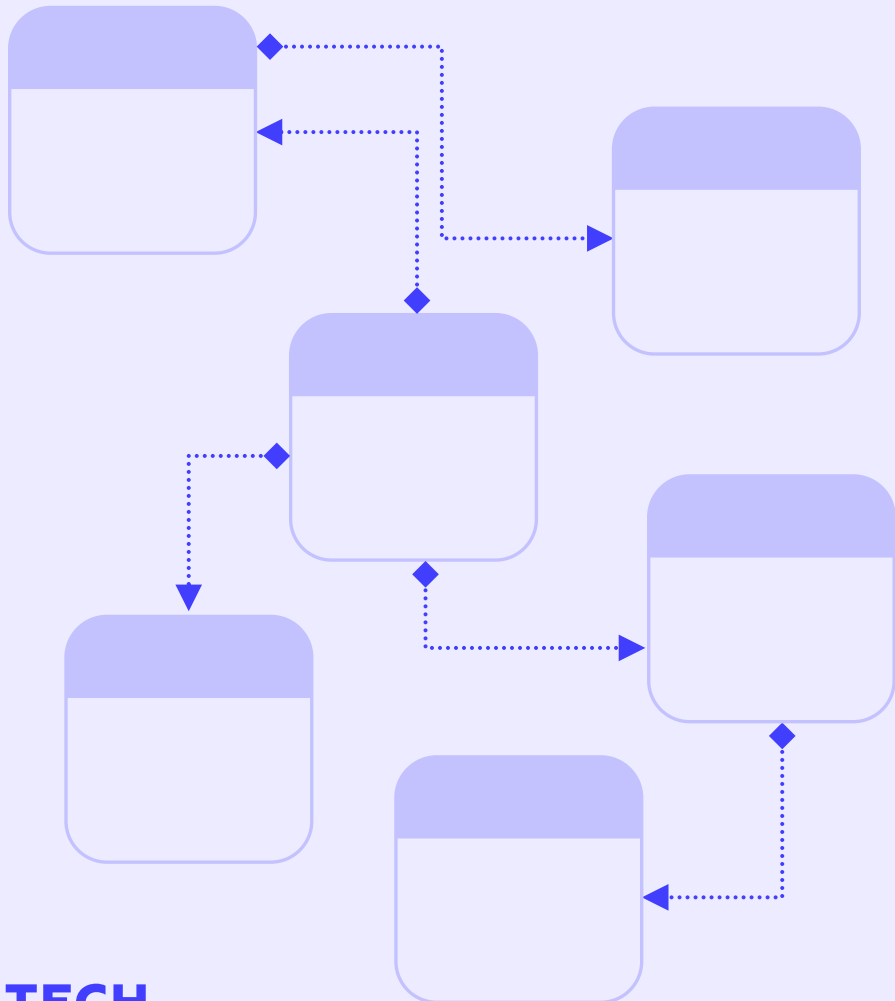
↓

Месяц	202201	202204	202205
Магазин	Группа		
M1	Г1		
M1	Г2		
M1	Г3		
M1	Г4		
M4	Г1		
M4	Г3		
M4	Г4		
M4	Г6		

Пространство для оптимизации

- Оптимизация структуры куба:
сокращение объема занятой памяти
вплоть до 10 раз
- Оптимизация расчёта метрик:
 - Сокращение объема используемой памяти
 - Увеличение производительности вычислений
 - Параллелизм вычислений

Эффективность работы с репозиторием



- Нормализация
- Индексы
- Контроль объема репозитория:
статистику и логи
следует накапливать
в отдельной БД

Эффективность файлового экспорта

			Год 2022				2023				
Регион	Филиал	Формат	Год-неделя	202249	202250	202251	202252	202301	202302	202303	202304
Волжский округ	Ижевск Север	МД	Сумма Продажи с учетом скидок, руб	5 361 636.86	6 484 586.91	6 307 447.72	9 222 828.67	4 474 885.67	4 670 739.93	4 956 218.06	4 711 306.78
			Кол-во уникальных Товар	2152	2157	2145	2117	2091	2058	2045	2026
Южный округ	Армавир	БФ	Сумма Продажи с учетом скидок, руб	10 443 512.26	12 851 012.03	15 971 035.49	26 631 574.81	9 262 531.13	8 463 394.29	8 304 208.47	7 966 179.67
			Кол-во уникальных Товар	3994	4001	3982	3963	3958	3952	3948	3919
	Краснодар Запад	БФ	Сумма Продажи с учетом скидок, руб	21 249 253.87	24 570 474.30	27 359 538.69	43 208 308.55	18 772 186.01	17 657 632.82	17 641 617.76	17 255 552.38
			Кол-во уникальных Товар	4474	4503	4543	4540	4507	4484	4493	4457

- Очень CPU-интенсивная операция
- Поточковая запись файла

Скриншот интерфейса Microsoft Excel, отображающий таблицу с данными по продажам и уникальным товарам за 2022 и 2023 годы. Видна панель меню (Файл, Главная, Вставка, Разметка страницы, Формулы, Данные, Рецензирование, Вид, Автоматизировать, Справка) и панель инструментов (Буфер обмена, Шрифт, Выравнивание, Число). Таблица занимает область A1:H8.

Эффективность решение фронтенда

- Кэширование
- Пейджинация
- Эффективный рендеринг многокомпонентных объектов

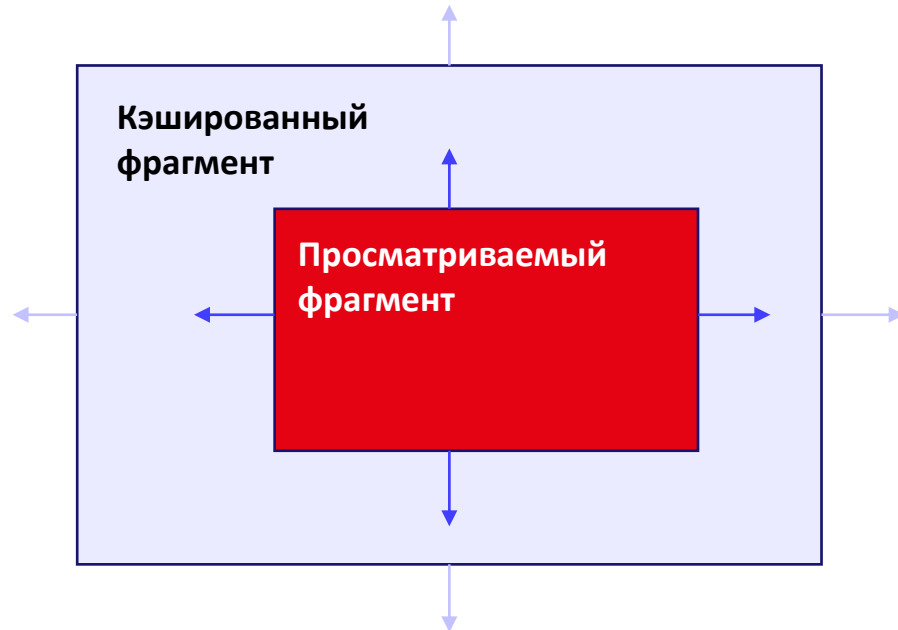
The screenshot shows a web application interface with a sidebar on the left and a main data table. The sidebar contains a 'Каталог отчёта' (Report Catalog) section with two sub-items: 'Количество ID задания' (Task ID Count) and 'Кол-во уникальных Логин' (Unique Logins Count). The main table displays data for various report categories across months from 2021 to 2022. The table has columns for 'Год' (Year) and 'Месяц-год' (Month-Year), followed by columns for each month from 2021-12 to 2022-08. The rows represent different report categories, each with two sub-rows for 'Количество ID задания' and 'Кол-во уникальных Логин'. The table includes a pagination bar at the top showing '1' and a toolbar with various icons for filtering and sorting.

Каталог отчёта	Месяц-год	Год		Месяц-год							
		2021	2022	2021-12	2022-01	2022-02	2022-03	2022-04	2022-05	2022-06	2022-07
Чеки (id: 74)	Количество ID задания				33	77	89	179	354	218	212
	Кол-во уникальных Логин			12	31	27	34	55	47	58	
Уровень сервиса (id: 17)	Количество ID задания	402	478	311	398	347	383	313	359	342	
	Кол-во уникальных Логин	99	121	104	98	108	103	98	110	101	
Доли (id: 16)	Количество ID задания	59	107	171	107	150	129	149	103	122	
	Кол-во уникальных Логин	36	47	69	56	60	70	78	58	64	
Показатели (id: 20)	Количество ID задания	110	115	125	104	139	186	144	176	170	
	Кол-во уникальных Логин	50	52	66	58	78	98	82	91	83	
АУМ (заморожен) (id: 81)	Количество ID задания						23	84	1		
	Кол-во										

Эффективность решение фронтенда

Кэширование

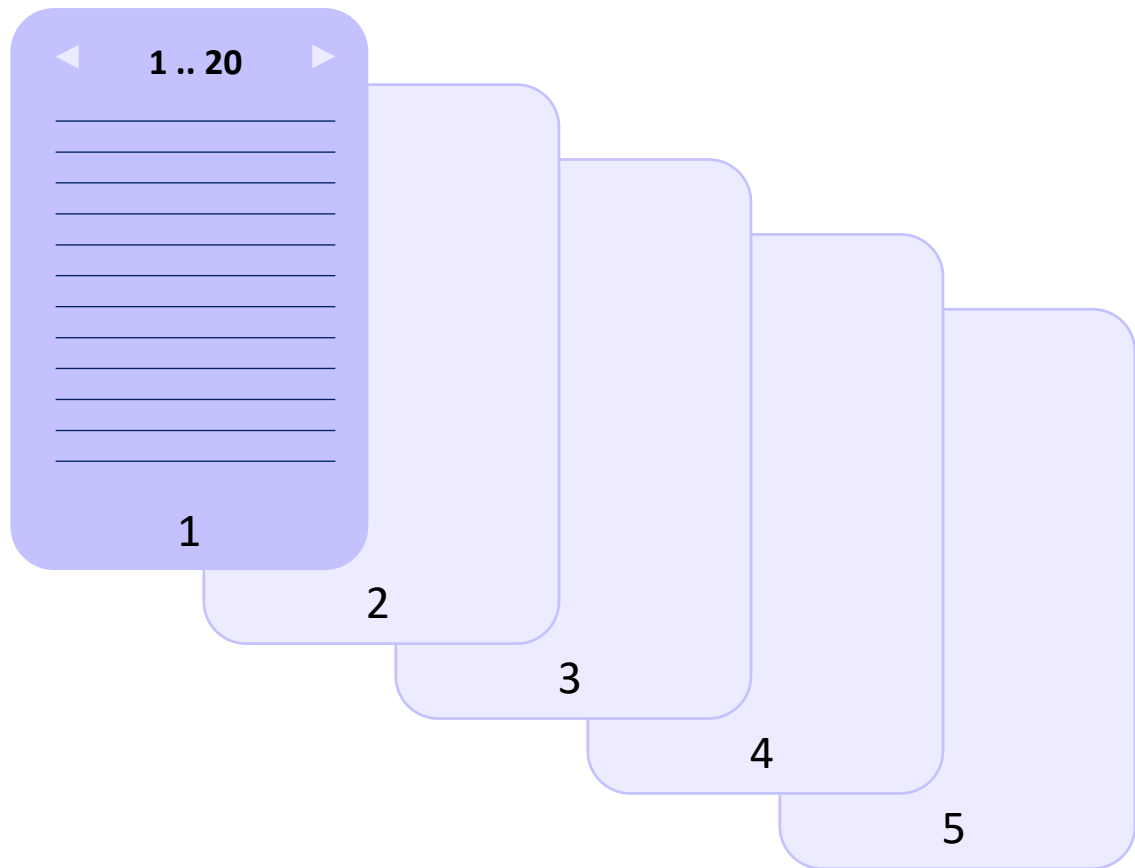
Вся таблица



- Мгновенный отклик на малые перемещения
- Плавная прокрутка

Эффективность решение фронтенда

Пейджинация



- Может быть как с кэшированием, так и без него
- Сокращает время получения данных от сервера
- Облегчает компонент и сокращает время прорисовки

Эффективность решение фронтенда

Эффективный рендеринг многокомпонентных объектов



Для таблиц с количеством ячеек порядка сотен использование техники событийного оснащения ячеек скрытым содержимым позволил уменьшить время рендеринга с **2–3 секунд** до **0.1–0.2 секунд**

Читай нас на:

HABR



[Habr.com/ru/
company/magnit](https://habr.com/ru/company/magnit)

VC



[Vc.ru/magnit-tech](https://vc.ru/magnit-tech)

Magnit



www.magnit.tech

Присоединяйтесь к нашей команде:



Спасибо за внимание!

Сухов Владимир

Руководитель сектора разработки BI и Big Data

 @V_Sukhov

 suhov_vb@magnit.ru