

avito.tech

Москва — 2023

Возможна ли жизнь с Open Source BI в большой компании





Что у нас было

O Tableau

Tableau в 2023 году 11-й год подряд становится лидером рейтинга компании Gartner для отрасли аналитических и BI-платформ.

Figure 1: Magic Quadrant for Analytics and Business Intelligence Platforms



Source: Gartner

Немного цифр

1200+ ~3000

DAU

all users

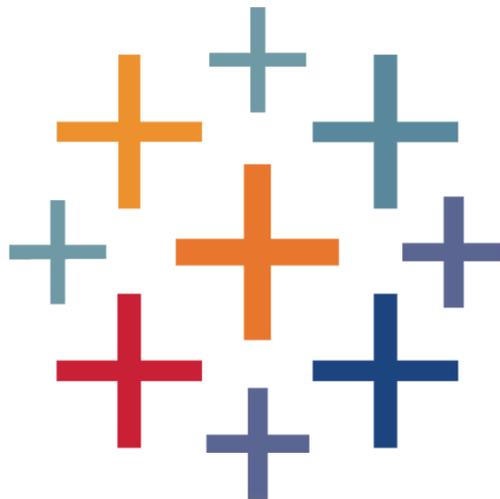
200+ ~3000

creator

workbook

500Gb

datasource size





Как выбирали замену

Первые вопросы

- Продление лицензии для Tableau Server.
- Продление лицензии для Tableau Desktop.

tableau.com is not available in your region.

5 месяцев на выбор нового ВІ и перенос отчётов*



Наши действия

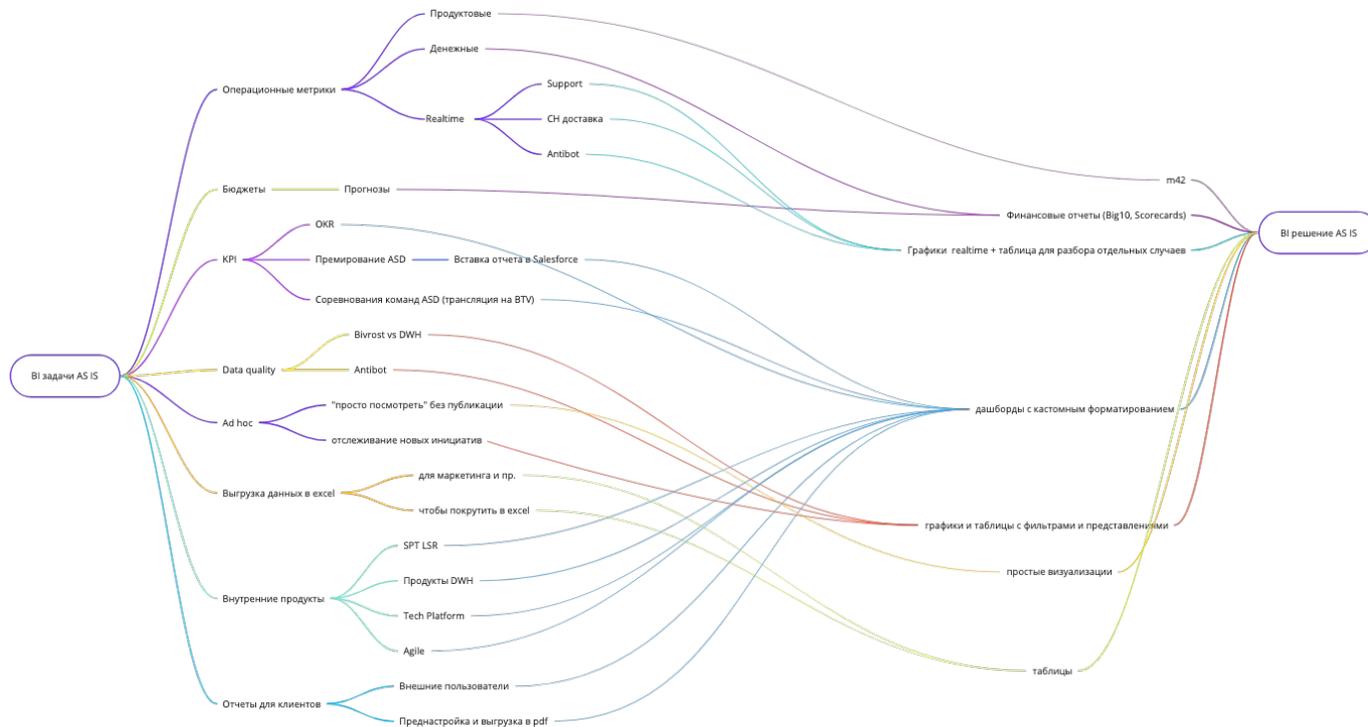
Анализ аналогов

- Лидеры рынка:
 - Microsoft Power BI.
 - Qlik.
- Open Source:
 - Apache Superset.
 - Redash.
 - Metabase.
- Yandex DataLens.
- FineBI.

Анализ использования Tableau

- Как используется Tableau.
- Список «активных» отчётов.
- Список «важных» отчётов.
- Кластеризация отчётов.
- Используемые фичи в отчётах.

Кластеризация отчётов



Кластеризация отчётов — Ad hoc

ВІ задачи AS IS

Data quality

Ad hoc

```
master[idap proxy:5433] Run Cancel Select...
48 hasNfrToml,
49 hasMiddleware,
50 requests,
51 budget,
52 latency,
53 errors,
54 handlersTotal,
55 handlersInNfr,
56 handlersWithNoData,
57 handlersWithMiddleware,
58 nfrCoverage,
59 middlewareCoverage,
60 nfrVersion
61 from dev_ext.service_nfr_api
62 order by dt
63 with options (
64   @service_name=['dwh-redash', 'data-docs'],
65   @date_start=(select to_char(now() - interval '1' day, 'YYYY-MM-DD')),
66   @date_end=(select to_char(now() - interval '1' day, 'YYYY-MM-DD')),
67   is_unsegmented = true,
68   launches_to_keep = 14,
69   send_errors_to = ['@ianikin']
70 )
71 schedule every day;
72
73
74
75 select * from dev_sae.service_nfr_api
76
77
78
79
```

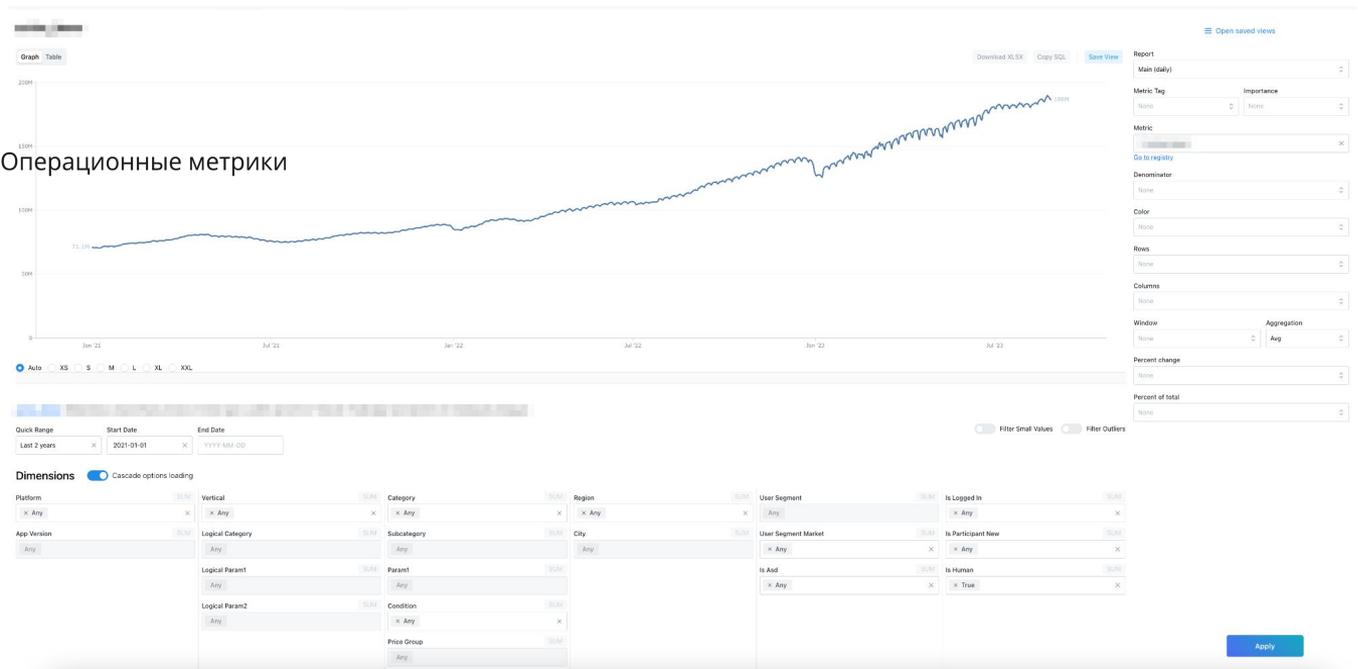
0.091 seconds 1 rows .csv .xlsx .twb | select * from dev_sae.service_nfr_api

launch_id	paas_service	dt	slo	sli	sli1	sli7	sli14	sli30	hasNfrToml	hasMiddleware	re
1	11 213 427	service-data-docs	null	null	null	null	null	null	null	null	

Кластеризация отчетов — m42

BI задачи AS IS

Операционные метрики



Используемые фичи

Группа	Требование	Приоритет
Filters	One on dashboard (Global)	3
Filters	Drop down	3
Filters	Data filter relative	2
Permissions	Roles/Groups (creator, viewer)	2
Dashboard	Text blocks	2
Analysts experience	Usability	2
Datasources	Automatic Data format detection	2
Datasources	Scheduling	2
Datasources	Cache/Extract/CH	3
Datasources	Vertica	3
Datasources	Clickhouse	3
End-user experience	Share custom view (preset)	3
Authorization	LDAP	2
Speed	Datasource communication	3
Visualization	Colors palette	2

Open Source

- superset (ver 1.4.1)
- redash (ver 10.1.0)
- metabase (ver 1.42.2)



Покрытие нашим требованиям

superset 39%

redash 35%

metabase 27%

Что в итоге??!

Анонс замены Tableau на Redash к концу июня 2022 года:

- Актуализировали на тот момент всю доступную нам информацию по работе с Redash в CF.
- Описали статью, которая поможет пользователям перевести отчёты.
- Создали отдельный канал для поддержки наших пользователей, которые создают отчёты в Redash.

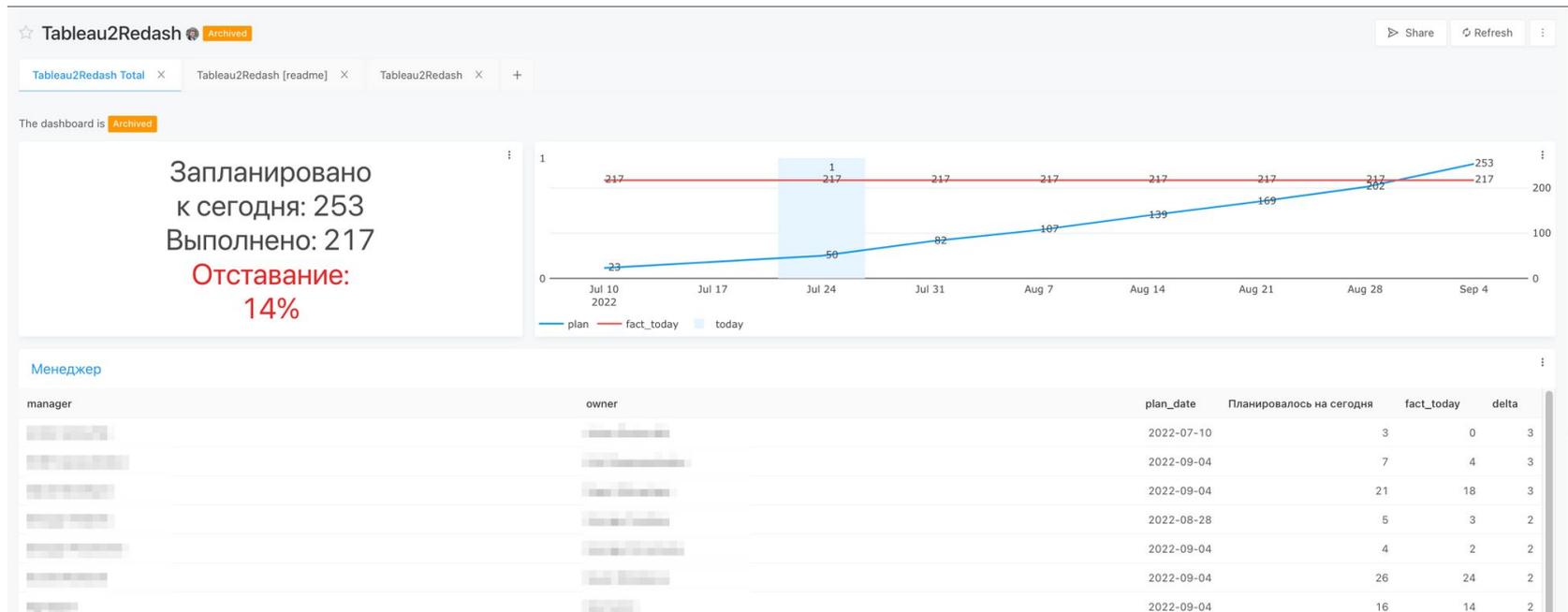


45 дней до дня X



Перевод отчётов

Критичные дашборды

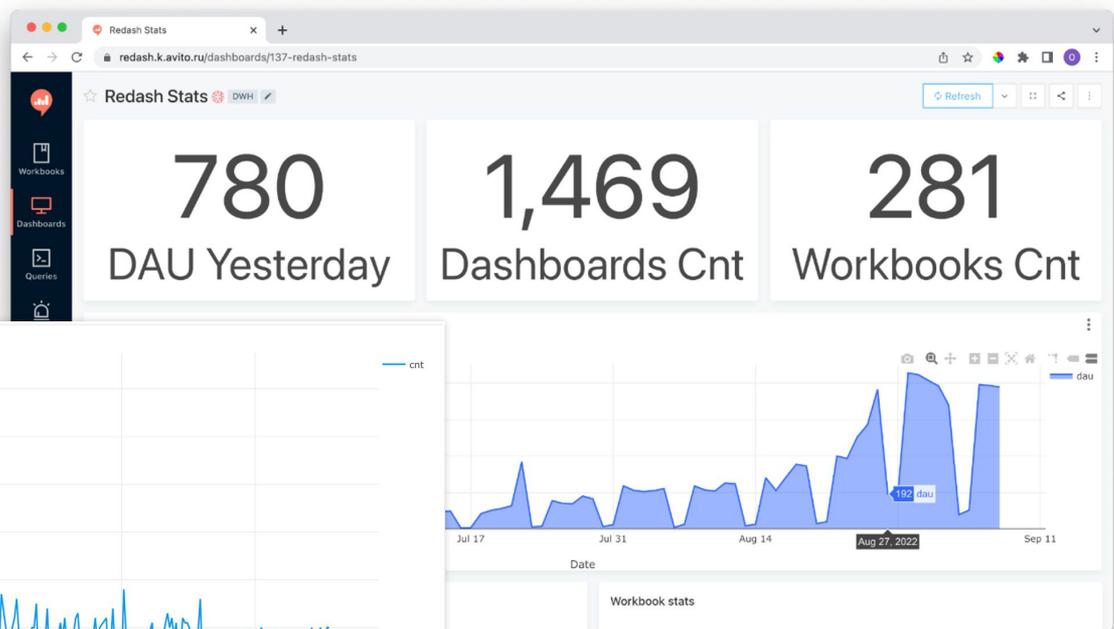




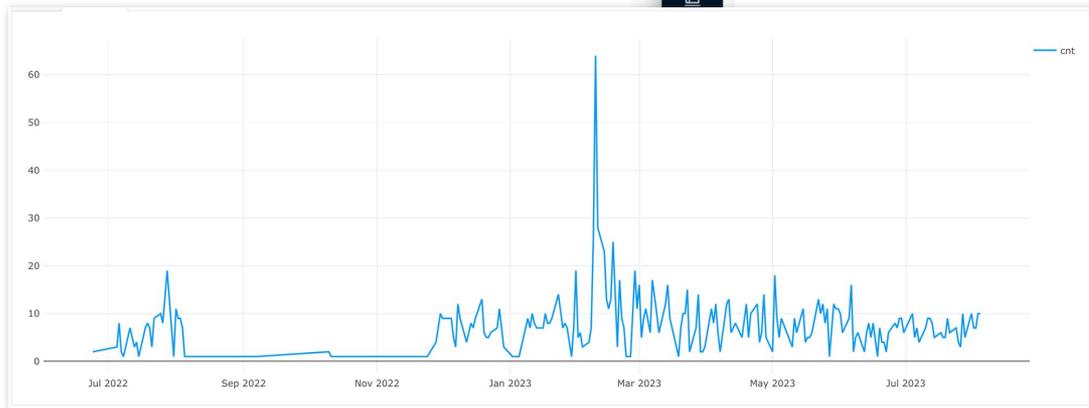
Поддержка

Поддержка

К августу нас было почти 800:



Вопросов в чате:





Инфраструктура



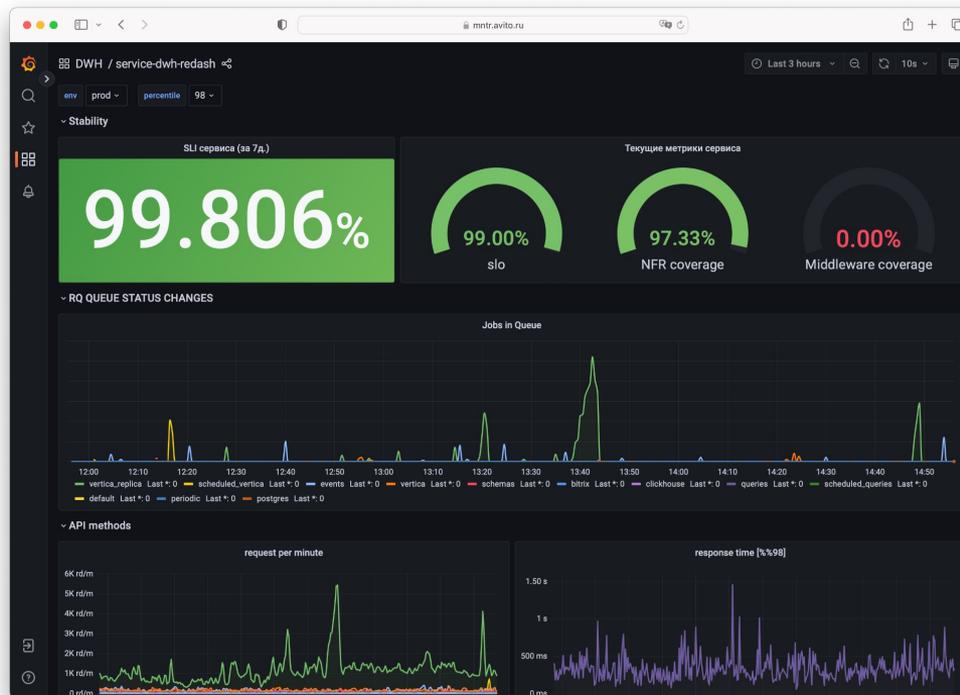
Инфраструктура

- отказоустойчивость в 3 dc k8s;
- метаданные в PostgreSQL;
- кэш данных в ClickHouse;
- очереди реализованы через Redis.

DWH в Avito:

- 1Pb данных в Vertica.
- 100Tb данных clickstream в ClickHouse.

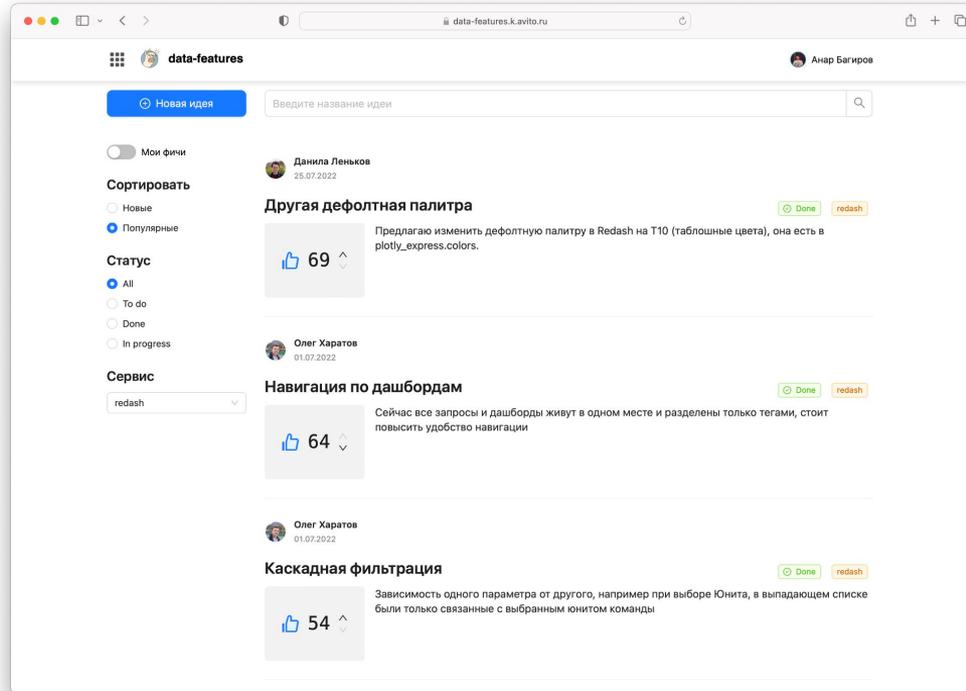
Доступность





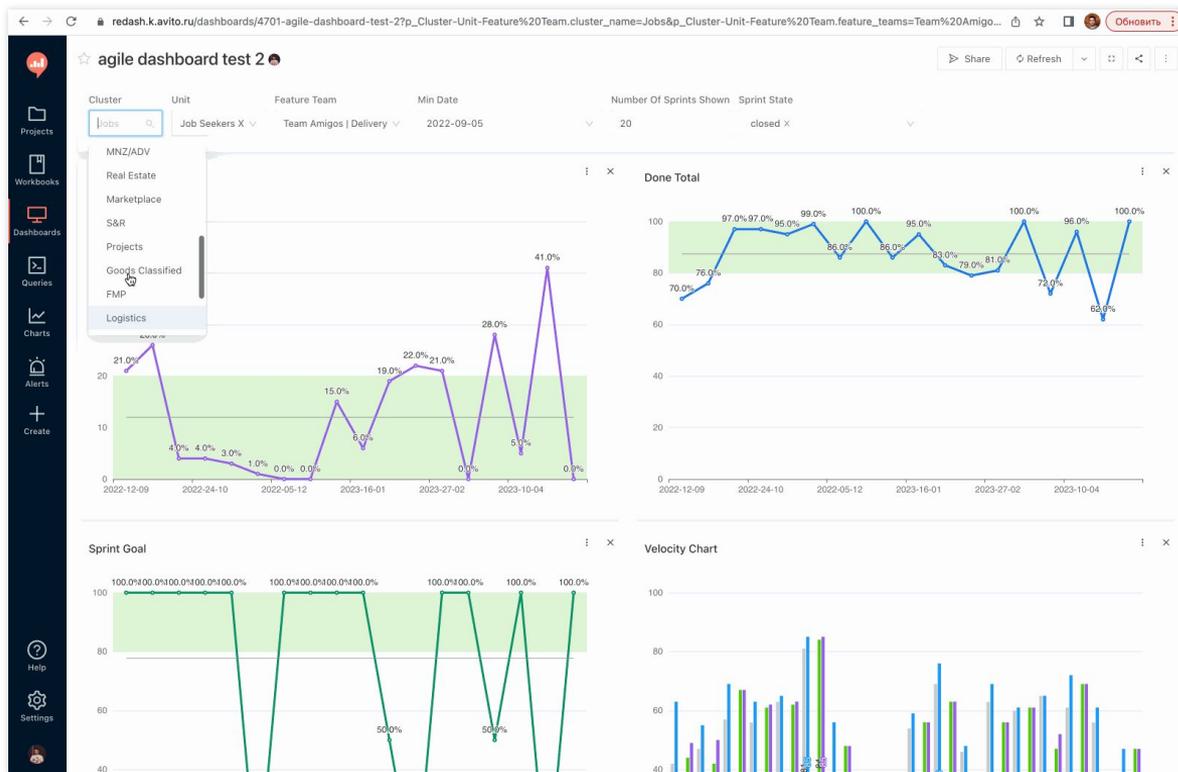
Разработка

Приоритизация



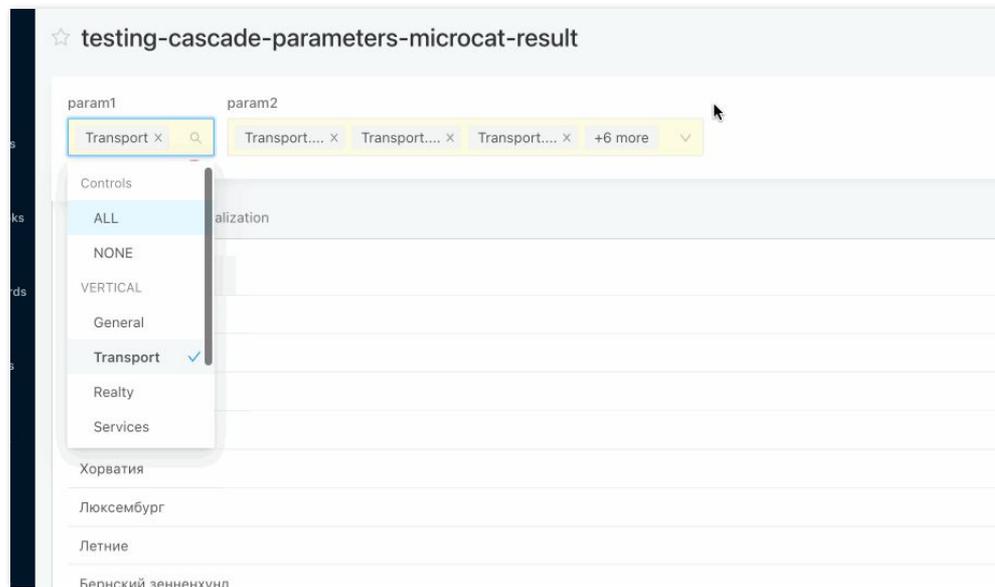
Крупные фичи

- row level security;
- workbook;
- projects;
- preset;
- NoSQL-визуализация;
- dataset.



Улучшение работы с параметрами

- ALL в dropdown параметрах.
- Каскадные параметры.



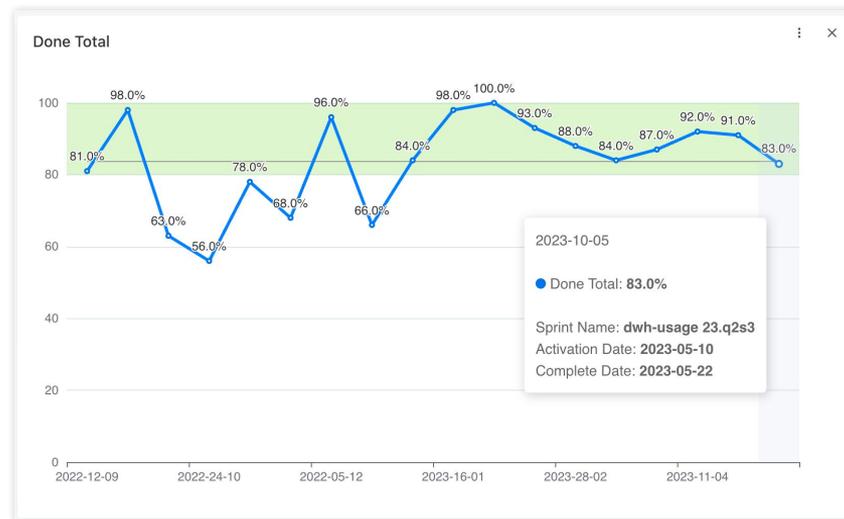
Улучшение визуализации

- Subplots.
- Новые сводные таблицы.



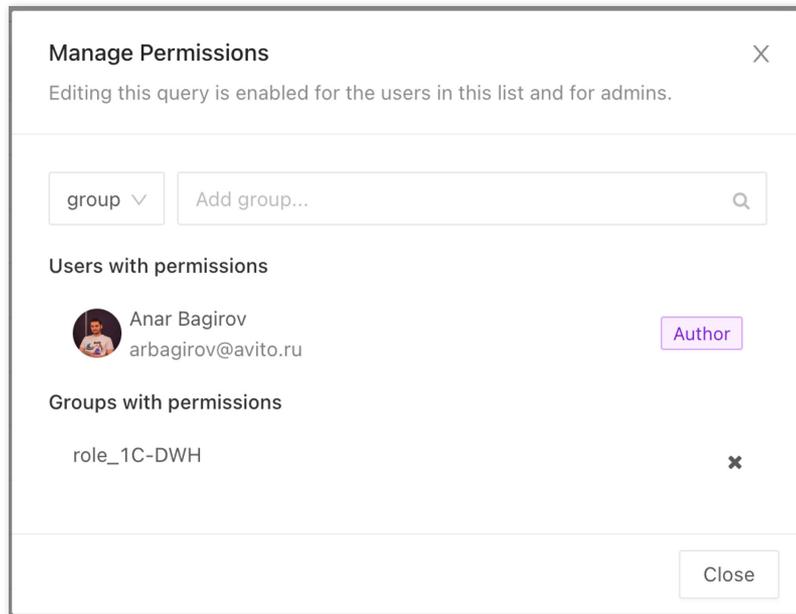
Улучшение визуализации

- Перевод визуализации на использование Apache ECharts:
 - не все фичи мы могли реализовать;
 - не все баги мы могли исправить;
 - блокируется UI во время рендера.



Групповые доступы

- Добавили поддержку групп в manage permissions.
- Группы в projects.



Алерты

- Два вида рассылок:
 - default;
 - batch.
- Обновили страницу просмотра.
- Изменили варианты отправки алертов.
- Добавили быстрые подписки.

When triggered, send notification:

Just once ▾

Name	Type	Created By	State	Last Updated At	Created At
PSDD tracking: smooth_fact_psdd < 86	DEFAULT	Aleksandr Afenov	UNKNOWN	44 minutes ago	2023-08-15 14:22
non_analytics_questions: group_role > 1	DEFAULT	Aleksandr Vanin	UNKNOWN	an hour ago	2023-08-15 14:14
Alert CPV new categories: qty >= 1	BATCH	Aleksandr Zvorygin	UNKNOWN	a day ago	2023-08-14 16:33
for dds load alert new: table_name !=	DEFAULT	Anastasiya Voloboeva	TRIGGERED	9 hours ago	2023-08-11 12:04
1056_6254_outsource_wfm_online_flow: Cnt > 1000	DEFAULT	Viktoriya Nayman	TRIGGERED	5 minutes ago	2023-08-11 09:07
autoteka_on_avito_revenue_checker: delta != 0	DEFAULT	Artur Verdyan	OK	3 days ago	2023-08-11 08:09
[Stability DM] Alert Super Critical Finish Time: late_cnt > 1	DEFAULT	Pavel Puchkov	OK	3 hours ago	2023-08-07 15:43
seo_metrics_alerts	DEFAULT	Elena Kotova	UNKNOWN	7 days ago	2023-08-04 16:54
Количество колл-ов в BMWG: count > 0	DEFAULT	Pavel Sayapin	OK	12 days ago	2023-08-03 14:30
empty attribute description for alert: owner !=	BATCH	Oleg Kharatov	TRIGGERED	12 days ago	2023-08-03 11:33
Mistakes in traction models filling for PMO	DEFAULT	Igor Tyushko	TRIGGERED	just now	2023-08-03 08:51
non_analytics_questions: group_role !=	BATCH	Aleksandr Vanin	TRIGGERED	12 days ago	2023-08-02 21:30
analytics_questions: group_role !=	BATCH	Aleksandr Vanin	TRIGGERED	12 days ago	2023-08-02 21:17
Alert AF CPC False-positive: cookie_fp >= 0.25	DEFAULT	Maksim Nekrasov	OK	12 days ago	2023-08-02 16:15
Mistakes in traction models filling for initiatives owners	BATCH	Igor Tyushko	TRIGGERED	2 hours ago	2023-08-02 10:01
Delivery_term_usage_checker: is_normal_delta == 0	DEFAULT	Alsher Zhurav	OK	13 days ago	2023-08-01 10:49
Bayes SPT due date: due_date_calculated < 14	DEFAULT	Valeriya	OK	17 days ago	2023-07-28

Технический долг

- Проект разделён на три части:
 - backend;
 - client;
 - библиотека визуализации.
- Клиент более чем на 50% переведен на TypeScript.
- Улучшили покрытие unit-тестами.
- Улучшили производительность.



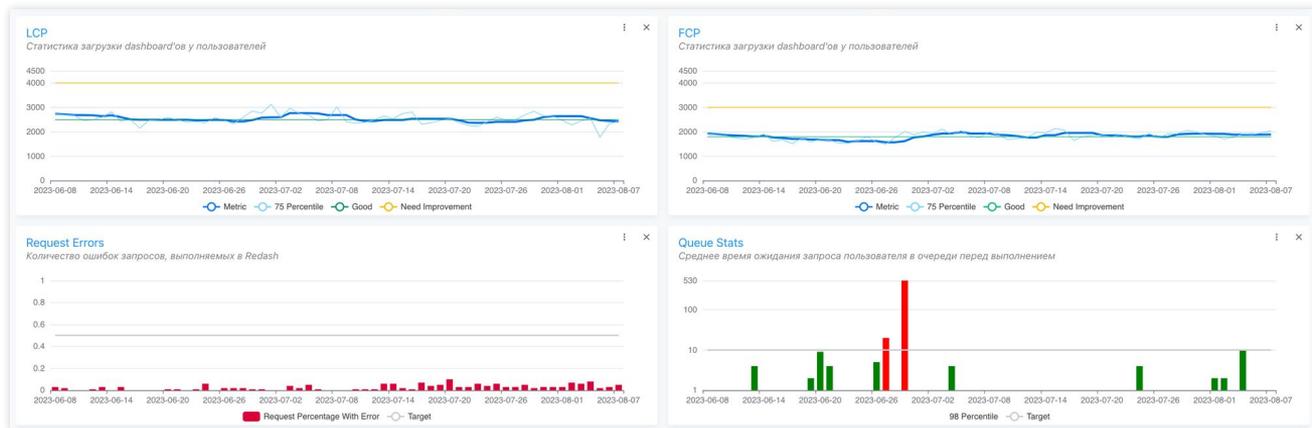
К чему мы пришли

Опрос пользователей

	Y22Q4	Y23Q1	Y23Q2
Вопрос_1/11_Оцените_удобство_создания_источников_данных_(query)	7	7	8
Вопрос_2/11_Оцените_удобство_создания_графиков	6.3	6.6	7.3
Вопрос_3/11_Оцените_полноту_функционала_в_Redash	5.2	6.0	6.6
Вопрос_4/11_Оцените_скорость_загрузки_(отрисовки)_дашбордов	6.8	6.7	7.1
Вопрос_5/11_Оцените_удобство_создания_дашбордов	6.1	6.5	6.8
Вопрос_6/11_Оцените_удобство_навигации	5.6	5.5	6.3
Вопрос_7/11_Оцените_удобство_кручения_(анализа)_данных	4.6	5.9	5.9
Вопрос_8/11_Оцените_как_вам_Redash_с_учетом_сделанных_доработок	6.6	6.9	7.6

Направления развития

- Исправление багов и отказоустойчивость.
- Поиск данных и скорость работы:
 - LCP — скорость загрузки основного контекста.
 - FCP — первая отрисовка контекста.
- Права.
- Embedded analytic.



Команда

- 3 фронтенд-разработчика.
- 2 бэкенд-разработчика.
- 1 техлид.





Итог

Минусы

- Больно после лидера рынка.
- Нужно готовиться к большому количеству вопросов, т.к. материалов очень мало.
- Нужны немалые ресурсы на доработку.
- Пользователи должны перестроиться для работы с новым инструментом.
- Отчеты на live-подключении работают не оптимально.

Плюсы

- Развяжет руки по лицензиям и количеству пользователей.
- Позволит добавить внешних пользователей, если продумать доступы.
- Можно доработать инструмент под нужды компании.

1990+ ~5500

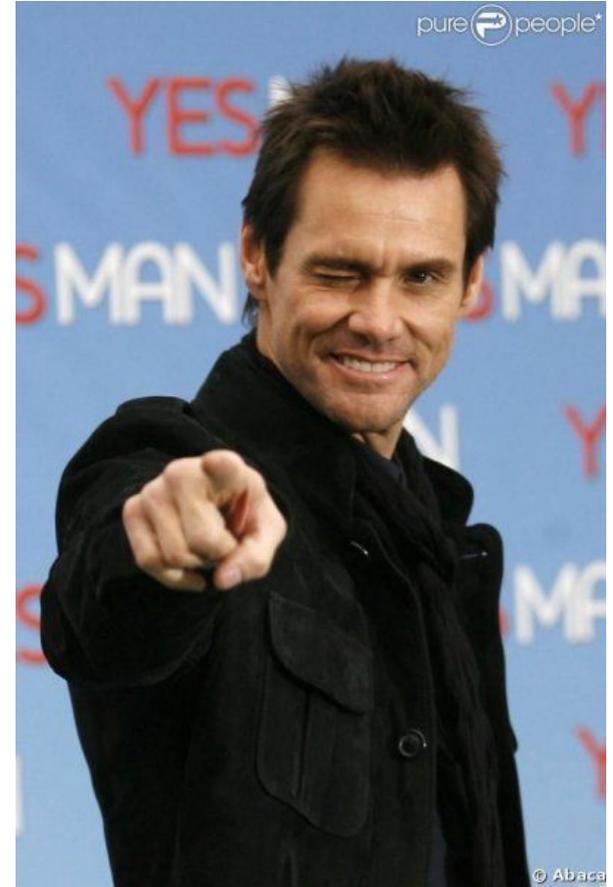
DAU

all users

500+

creator

avito.tech



avito.tech

АНИКИН Илья

 ianikin@avito.ru

 [@bora86bora](https://t.me/bora86bora)

Ссылки

Open Source Redash