

**РУКОВОДСТВО ПОЛЬЗОВАТЕЛЯ
СИСТЕМЫ БИЗНЕС-АНАЛИТИКИ С КОНСТРУКТОРОМ
НЕЙРОСТЕЙ**

Санкт-Петербург
2023

Оглавление

1. Введение	4
2. Основные понятия системы	5
3. Роли в системе.....	7
3.1 Администрирование сервера. Первая авторизация в системе. Смена пароля.....	8
3.2 Настройка SMTP-аккаунта	9
3.3 Новый домен	11
3.4 Список доменов	13
4. Авторизация в системе	13
4.1 Первый вход в систему	13
4.2 Восстановление пароля.....	15
5. Приложения.....	15
5.1 Назначение	15
5.2 Создание приложений.....	16
5.3 Публикация приложений	17
5.4 Создание дашбордов	18
5.5 Наполнение дашбордов графиками	20
5.6 Кросс-фильтрация	22
5.7 Глобальные фильтры.....	23
6. Графики.....	24
6.1 Создание группы графиков	24
8.1 Создание графика	26
8.2 Редактор графиков.....	37
7. Шаблоны.....	52
7.1 Назначение	52
8. Конструктор.....	67
8.1 Начало работы в конструкторе после создания нового домена.....	67
8.2 ETL-процесс	72
8.2.1 Подключение к внешнему источнику данных, получение схемы данных	74
8.2.2. Коннекторы и сервисы.....	75
8.2.3 Извлечение таблицы данных в ETL хранилище: создание спецификации, шаблона и ETL-задачи.....	76
8.2.4 Трансформация (создание представлений, объединение, группировка, сортировка, условия)	80
8.2.5 Отправка данных в шину данных: создание топика, шаблона и ETL-задачи	86
8.2.6 Отправка данных в аналитическое хранилище: создание топика, шаблона и ETL-задачи	91
8.2.7 Импорт данных в форматах csv, json, excel. Назначение типов, перезапись и дозапись данных.....	97

8.3	Настройка расписания.....	99
8.4	Перезапись данных, инкрементальное обновление данных	102
8.5	Настройка кросс-фильтрации.....	104
8.6	Настройка глобальных фильтров (виртуальные типы данных, настройка)	111
9.	Безопасность (создание пользователей, состав ролей, создание групп, назначение ролей на группу, разграничение доступа к приложениям, разграничение доступа к данным в ETL-хранилище и аналитическом хранилище).....	116
9.1	Добавление пользователя.....	116
9.2	Роли	118
9.3	Группы доступа.....	119
9.4	Назначенные группы доступа.....	119
9.5	Роли по группам.....	120
10.	Конструктор нейросетей	122
10.1	Шаг 1. Создание шаблона ML-задачи.....	122
10.2	Шаг 2. Запуск ML-задачи.....	127
10.3	Шаг 3. Мониторинг.....	128
11.	Заключение	130

1. Введение

Данный документ представляет из себя описание функциональных возможностей системы.

Изучив данный документ, пользователь сможет полноценно работать в системе: создавать новые домены, приложения, графики, дашборды, добавлять новых пользователей, конструировать нейрости.

2. Основные понятия системы

Таблица № 1. Основные понятия

Термин	Описание
ETL-процесс	Процесс извлечение, преобразования и отправки данных в аналитическое хранилище
ETL-сервисы	Kafka.broker в качестве шины данных для передачи в аналитическое хранилище, druid.broker в качестве аналитического хранилища, коннекторы для подключения к источникам и извлечения схемы данных и самих данных
ETL-хранилище	Хранилище на postgresql, в которое выгружаются данные из внешних источников, для дальнейшего преобразования и отправки в аналитическое хранилище
Аналитическое хранилище (Apache Druid)	СУБД, обеспечивающая колоночное хранение большого объема данных
Виджет (Библиотека графиков)	Это график, построенный на основе выбранного шаблона визуализации и данных
Внешние источники данных	Это внешние системы, к которым с помощью коннекторов может подключиться Инженер данных
Глобальные фильтры	Работают в рамках приложения. Глобальные фильтры позволяют устанавливать фильтрацию по разным значениям полей
Дашборд	Информационная панель, которая состоит из совокупности графиков, имеет признак интерактивности
Домен	Автономная область, в том числе безопасности, которая характеризуется следующими признаками: уникальными пользователями, ETL-хранилищами, аналитическим хранилищем, приложениями, дашбордами, данными

Источники данных	Это извлеченные таблицы данных, которые хранятся в ETL-хранилище и Аналитическом хранилище
Конструктор	Раздел в системе для конструирования ETL-процесса, управления безопасностью, управлению настройками визуализации, и администрирования системы
Кросс-фильтрация	Функционал на дашборде, при котором пользователь выбирает на графике значение(ия) и другие графики фильтруются согласно этому запросу. Кросс-фильтрация работает только в рамках одного дашборда
Приложение	Обособленная предметная область, которая состоит из совокупности дашбордов, и обладает собственной безопасностью
Служебная база данных	База postgresql, предназначенная для хранения объектов: описание источников данных, описание ETL-хранилищ, описание аналитических хранилищ, описание приложений (дашбордов, графиков и правил по работе с ними), описание ETL-задач, информацию по подсистеме прав доступа (пользователи, группы, роли), служебные представления (деревья, история поиска и выбора пользователей и т.д.), фильтры
Шаблон (шаблон визуализации)	Тип визуального представления информации. Например: кольцевая диаграмма, таблица, карточка с текстом, столбчатый график
Шина данных (Apache Kafka)	Распределенный брокер сообщений, используется для доставки данных из внутреннего ETL-хранилища до аналитического хранилища

3. Роли в системе

Функциональные возможности работы пользователей с системой определяются такой сущностью, как «Роль». Доступ к информации, содержащейся в системе, и аналитическим приложениям определяются такой сущностью, как «Группы». Каждому пользователю может быть назначена одна или несколько следующих ролей:

1. «Мастер-администратор» — управляет доменами, настройками исходящей почты.
2. «Администратор домена» — назначение остальным пользователям ролей и групп, регистрация ETL хранилища, Аналитического хранилища, ETL-сервисов.
3. «Инженер данных» (Data Developer):
 - настройка процесса извлечения, преобразования данных и их загрузки в аналитическое хранилище, настройка расписания процесса (ETL-процесс);
 - работа с моделями данных в хранилище, создание дополнительных структур данных.
4. «Дизайнер приложений» (Data Designer) — конструирование аналитических приложений, предназначенных для работы остальных пользователей, на основе информации в аналитическом хранилище.
5. «Аналитик» — работа с назначенными по политике безопасности аналитическими приложениями, возможность создания собственных дополнительных дашбордов и графиков на них.
6. «Пользователь» — работа с аналитическими приложениями без возможности создания собственных дополнительных дашбордов и корректировки существующих, только режим просмотра.

Каждому пользователю в подсистеме может быть назначено неограниченное число ролей, разрешенное ему по политике безопасности.

Каждой «Группе» в подсистеме может быть назначены для доступа неограниченное количество аналитических приложений и источников (моделей данных) в аналитическом хранилище.

После создания необходимых групп в них добавляются конкретные пользователи, которые получают доступ к соответствующей информации, содержащейся в подсистеме.

Сочетание назначенных пользователю ролей и групп формирует набор функциональных возможностей по работе с системой, доступных пользователю, и конкретных разделов аналитического хранилища, к которым пользователь имеет доступ, и возможности управлять и настраивать ETL-процесс.

3.1 Администрирование сервера. Первая авторизация в системе. Смена пароля

Осуществить первое подключение к серверу, а также создать новые домены и сделать их первичную настройку можно с помощью подключения к Администрированию сервера, авторизовавшись через учетную запись Мастер-администратора. Данные для авторизации вы можете получить у вашего IT-департамента/службы технической поддержки.

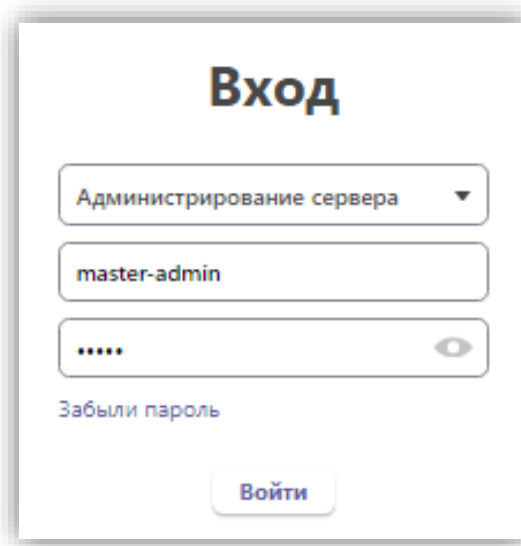


Рисунок 1. Авторизация в системе для Мастер-администратора

Далее после смены пароля вам станут доступны следующие пункты меню:

- Настройка SMTP аккаунта

- Новый домен
- Список доменов

3.2 Настройка SMTP-аккаунта

Simple Mail Transfer Protocol (SMTP) — простой протокол связи, применяемый с целью пересылки электронных писем с сервера отправителя на сервер получателя. Этот протокол не рассчитан на обработку входящих сообщений, его используют для отправки и последующей доставки писем адресату.

SMTP-сервер — сервер, который работает по протоколу SMTP. Его главная задача — выступать ретранслятором (передатчиком) между серверами отправителя и адресата. Каждый SMTP-сервер обладает собственным адресом в формате `smtp.serveraddress.com`. Это позволяет безошибочно определять нужный сервер при пересылке почты.

Настройка SMTP-аккаунта осуществляется один раз. Данные для настройки SMTP-аккаунта вы можете получить у вашего IT-департамента/службы технической поддержки. Данный сервис будет осуществлять рассылку учетных данных для новых подключенных пользователей. В качестве отправителя пользователи будут видеть адрес электронной почты, который будет указан в настройках.

На рисунке ниже вы можете увидеть окно настройки SMTP-аккаунта.

Настройки SMTP аккаунта

Адрес сервера

Порт сервера

Почтовый протокол

Использовать SSL/TLS

Имя пользователя

Пароль

Рисунок 2. Настройки исходящей почты

- Адрес сервера – адрес электронной почты в формате smtp.yandex.ru;
- Порт сервера – номер порта для сервера электронной почты
- Почтовый протокол – протокол передачи почты (IMAP, POP3, SMTP, HTTP);
- Использовать SSL/TLS (SSL — Secure Socket Layer, уровень защищенных сокетов; TLS — Transport Layer Security, безопасность транспортного уровня) – использование протоколов для обеспечения защищенной передачи данных;
- Имя пользователя – учетное имя пользователя на почтовом сервере;
- Пароль – пароль для доступа к почте.

После того, как будут заполнены все поля, необходимо нажать на кнопку «Проверить соединение». Если соединение установлено, отобразится индикатор зеленая галочка возле кнопки.



Рисунок 3. Успешное соединение с сервером

В случае, если возникли проблемы с установкой соединения, отобразится индикатор красный крестик. Необходимо обратиться в ваш IT-департамент/службу технической поддержки для решения вопроса.



Рисунок 4. Не удалось установить связь с сервером

3.3 Новый домен

Для того, чтобы приступить к созданию нового домена, в первую очередь необходимо подключиться к служебной базе на PostgreSQL.

Рисунок 5. Форма создания нового домена

Информацию для подключения к служебной базе вы можете получить у вашего IT-департамента/службы технической поддержки.

После того, как будут заполнены все поля, необходимо нажать на кнопку «Проверить соединение». Если соединение установлено, отобразится индикатор зеленая галочка возле кнопки.



Рисунок 6. Успешное соединение со служебной базой

В случае, если возникли проблемы с установкой соединения, отобразится индикатор красный крестик. Необходимо обратиться в ваш IT-департамент/службу технической поддержки для решения вопроса.



Рисунок 7. Не удалось установить связь со служебной базой

После установки соединения с базой осуществляется настройка названия домена. Ограничений в названии нет, название может быть написано как русскими, так и английскими буквами, содержать в себе цифры и знаки.

Далее необходимо ввести **параметры** для администратора домена. Впоследствии администратор будет осуществлять настройку самого домена, добавлять новых пользователей. Если пользователь, которому необходимо назначить роль Администратора домена, уже был ранее добавлен в систему в другие домены, то при вводе его почты выпадет список со всеми совпадениями по почте.

! Мастер-администратор имеет доступ исключительно к разделу Администрирование. У Мастер-администратора нет доступа к доменам, соответственно, никаких настроек внутри домена Мастер-администратор производить не может.

Один пользователь может быть администратором в нескольких доменах, если для предусмотрено наличие нескольких доменов.

После установки необходимых параметров необходимо нажать на кнопку «Создать». Домен будет создан, он будет отображен в окне «Список доменов», а также станет доступен для выбора в списке доменов в окне первого запуска.

После создания домена новому Пользователю, который назначен Администратором домена, на почту придет письмо, в котором указаны данные для доступа, а именно: домен, роль, логин, пароль, ссылка на вход. Если

Пользователь уже существует в Системе, то ему придет письмо с указанием, в какой домен его добавили.

3.4 Список доменов

В данном разделе Мастер-администратор может:

- посмотреть список всех созданных доменов (поля: идентификатор, дата создания, дата обновления, имя домена в базе, имя домена для пользователей, служебная база данных, имя пользователя служебной базы данных, статус)
- заблокировать/разблокировать домен путем установки или снятия галочки в поле Статус

Индетификатор	Дата создания	Дата обновления	Имя	Имя домена	Статус
1873e17d-9d8e-4283-	2022-10-03 08:05:05		domen-0310	Домен 0310	<input checked="" type="checkbox"/>

Рисунок 8. Блокировка домена

4. Авторизация в системе

4.1 Первый вход в систему

Вход в учетную запись осуществляется с главной страницы. Для успешной авторизации в системе необходимо выбрать домен (домен - автономная область, в том числе безопасности, которая характеризуется следующими признаками: уникальными пользователями, ETL-хранилищами, аналитическим хранилищем, приложениями, дашбордами, данными), в котором необходимо авторизоваться, пароль, который был прислан пользователю на почту/пользователь ранее сам установил.

Вход

Администрирование сервера

Логин

Пароль

Забыли пароль

Войти

Рисунок 9. Окно авторизации

В случае, если при авторизации домен был выбран неправильно, система покажет ошибку: «Неверный логин или пароль».

После входа в учетную запись Пользователь может сменить пароль. Для этого необходимо во вкладке “Профиль”, которая расположена в верхнем правом углу, кликнуть по кнопке “Сменить пароль”.

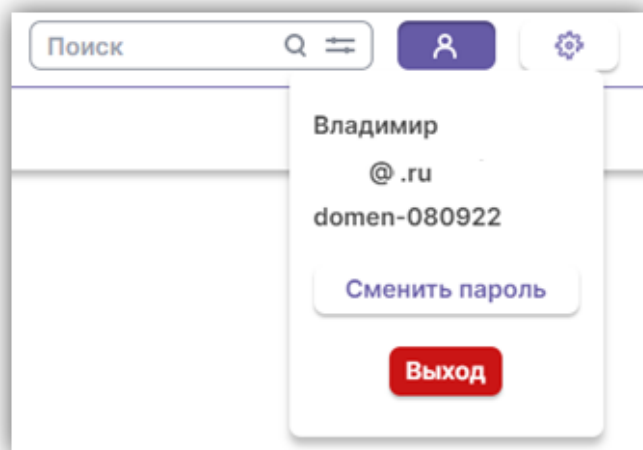


Рисунок 10. Смена пароля

После этого откроется окно, в котором нужно указать действующий пароль и ввести новый.

A screenshot of a password change form. The form has a title 'Смена пароля' and a close button 'X' in the top right corner. It contains three input fields: 'Старый пароль', 'Новый пароль', and 'Подтвердите пароль'. Each field has a toggle icon (an eye) to the right of it. At the bottom of the form is a button labeled 'Подтвердить'.

Рисунок 11. Форма смены пароля

! Уведомление, которое придет на почту после смены пароля не содержит нового пароля.

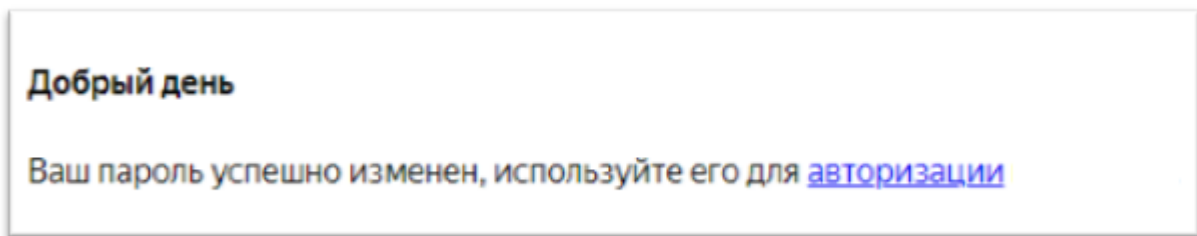


Рисунок 12. Письмо о смене пароля

4.2 Восстановление пароля

Если Пользователь забыл пароль и не может найти письмо на почте, то для восстановления пароля на главной странице необходимо указать логин и кликнуть по кнопке “Забыли пароль”.

После этого Пользователю придет на почту, указанную при регистрации, новый пароль. Логин остается прежним.

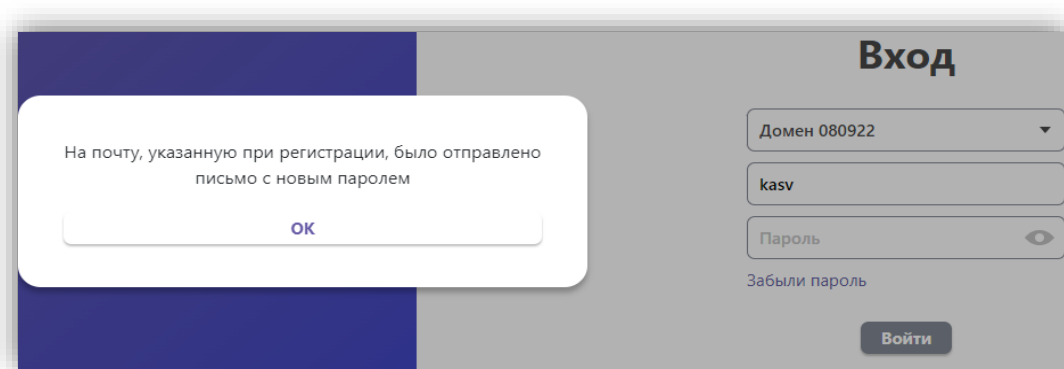


Рисунок 13. Уведомление о новом пароле

5. Приложения

5.1 Назначение

Приложение — предметная область, объединяющая неограниченное количество дашбордов (листов), на каждом из которых может располагаться неограниченное количество графиков.

Свойства приложений:

1. Произвольное количество дашбордов из числа доступных для пользователя
2. Их произвольное расположение относительно друг друга

3. Общая фильтрация по дашбордам согласно ручным настройкам (как правило дата, период, общий реквизит)
4. Обмен настройками одинаковых панелей между пользователями

5.2 Создание приложений

Для создания нового приложения необходимо:

- нажать кнопку «Создать», расположенную в верхней части экрана;
- в появившемся окне «Создание нового приложения» ввести «Название приложения», добавить «Описание», выбрать цвет, при необходимости выбрать изображение (из собственного файла), после чего нажать кнопку «Сохранить».

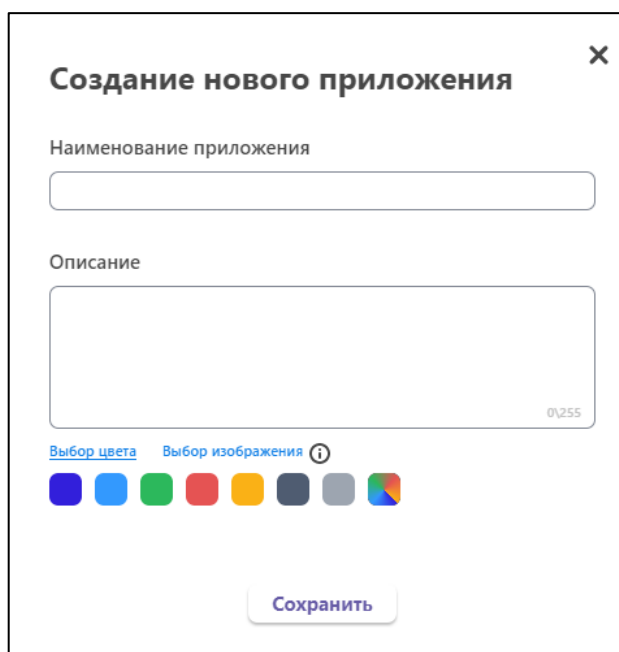


Рисунок 14. Форма создания нового приложения

- для редактирования / удаления приложения необходимо нажать кнопку «Свойства приложения», расположенную в правом верхнем углу каждого созданного приложения, и выбрать редактировать / удалить.

- при редактировании приложения появится окно, аналогичное окну «Создание нового приложения», в котором необходимо внести все корректировки и нажать кнопку «Сохранить».

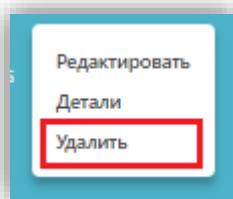


Рисунок 15. Удаление приложения

Для удаления приложения необходимо нажать кнопку «Удалить».

В пункте меню «Детали» можно посмотреть, кем и когда было создано приложение, а также когда были внесены последние изменения.

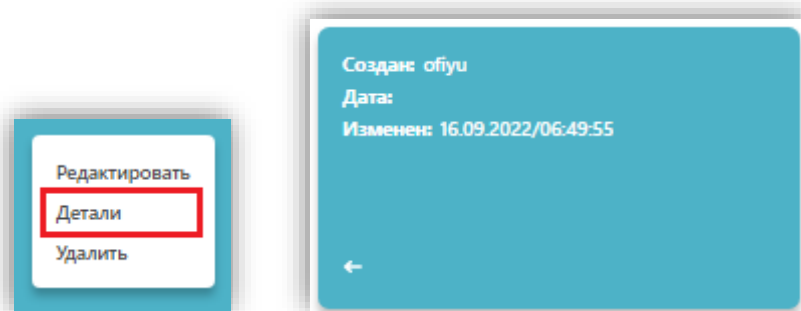


Рисунок 16. Детали приложения

- В левом нижнем углу каждого приложения расположена иконка, при наведении курсора на которую всплывает текст, введенный в поле «Описание» на этапе создания / редактирования приложения.

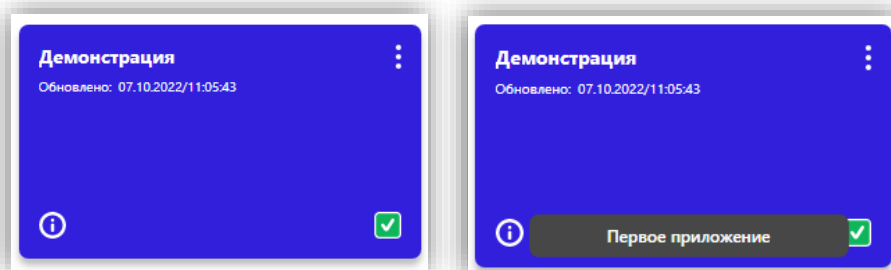


Рисунок 17. Описание приложения

5.3 Публикация приложений

При создании приложения оно доступно только создателю. Для того, чтобы приложение стало доступно другим, необходимо его опубликовать.

Примечание: при публикации приложение становится доступно только тем Пользователям, которые входят в группу доступа создателя.

При нажатии кнопки «Свойства приложения», расположенной в правом верхнем углу каждого созданного приложения, и переходе в режим редактирования становится доступным чек-бокс для установки/снятия статуса приложения «Опубликовано».

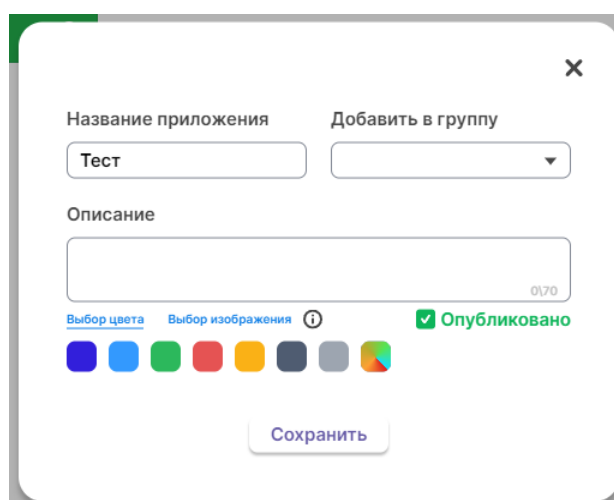


Рисунок 18. Публикация приложения

После нажатия кнопки «Сохранить» в правом нижнем углу этого приложения появится иконка в виде галочки в зеленом квадрате в случае установления статуса «Опубликовано» или иконка с изображением красного замка в случае отключения статуса «Опубликовано».

Созданное приложение имеет статус «Не опубликовано» и доступно только пользователю, который создал его. При установлении статуса «Опубликовано» приложение становится доступным для всей группы, в которую входит пользователь, создавший приложение.

5.4 Создание дашбордов

Дашборды (листы) — составная часть аналитического приложения, основные свойства которой:

1. Произвольное количество графиков из числа доступных в библиотеке графиков
2. Произвольное расположение графиков относительно друг друга
3. Кросс-фильтрация для сценарного анализа
4. Бесконечная прокрутка листа вниз

При выборе и нажатии левой кнопкой мыши по необходимому приложению пользователь переходит к отображению всех дашбордов, относящихся к этому приложению.

Для создания нового дашборда необходимо:

- нажать кнопку «Создать», расположенную в верхней части экрана
- во всплывшем окне «Создание нового дашборда» ввести «Название дашборда», добавить «Описание» и выбрать цвет, после чего нажать кнопку «Создать»

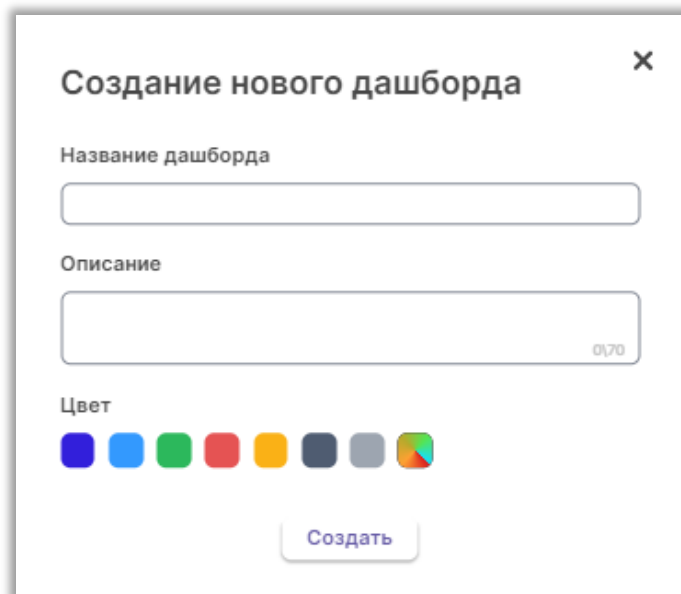


Рисунок 19. Форма создания дашборда

Количество создаваемых дашбордов не ограничено.

- Для редактирования / удаления дашборда необходимо нажать кнопку «Свойства дашборда», расположенную в правом верхнем углу каждого созданного дашборда, и выбрать редактировать / удалить.

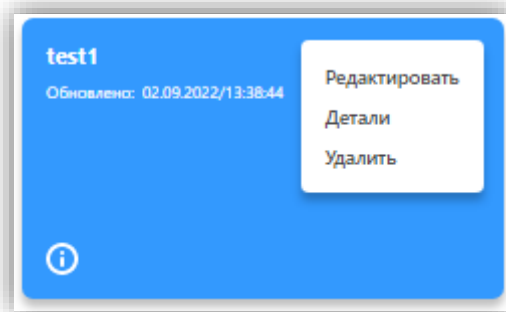


Рисунок 20. Свойства дашборда

- При редактировании дашборда появится окно, аналогичное окну «Создание нового дашборда», в котором необходимо внести все корректировки и нажать кнопку «Сохранить».

При выборе необходимого дашборда пользователь переходит к отображению области построения дашборда.

5.5 Наполнение дашбордов графиками

- Готовый дашборд представляет из себя интерактивную информационную панель, состоящую из совокупности графиков.
- График — соединение шаблона графического отображения данных и конкретного источника (объекта) данных.
- У дашборда есть два состояния: режим просмотра и режим редактирования.
- В режиме просмотра на дашборде нельзя удалять графики, изменять их расположение, добавлять новые.
- В режиме редактирования дашборд можно изменять.
- По двойному клику на графике пользователь сможет перейти к редактированию выбранного графика сразу с дашборда

Для добавления графика на основное поле дашборда необходимо:

- выбрать необходимый дашборд – произойдет переход к области построения дашборда
- нажать кнопку с иконкой карандаша в правом верхнем углу (переход в режим редактирования) – в левом верхнем углу появится кнопка «Создать»

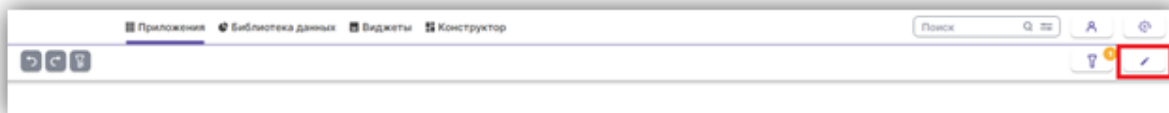


Рисунок 21. Переключение дашборда в режим редактирования

- нажать кнопку «Создать», расположенную в левой верхней части экрана – в правой части экрана появится панель «Настройки дашборда»

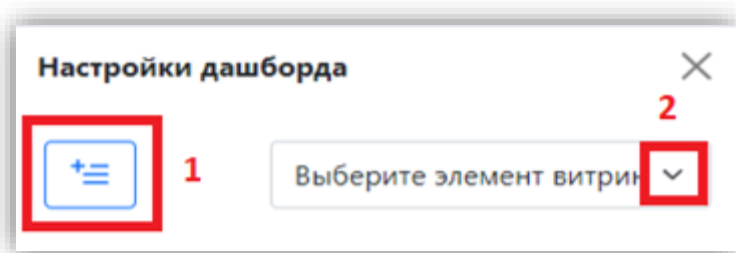
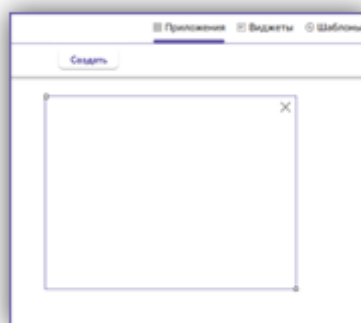


Рисунок 22. Настройки дашборда

- нажать кнопку «Добавить» (1) – на область построения дашборда будет добавлен пустой контейнер
- нажать на контейнер, в который необходимо поместить график, и в выпадающем списке «Выберите элемент витрины данных» (2) выбрать ранее созданный график из библиотеки. В контейнере отобразится выбранный график



- контейнер можно разместить в любом месте области построения дашборда в соответствии с логикой и макетом дашборда. Также можно установить необходимый масштаб контейнера, растягивая/сжимая контейнер за левый верхний или правый нижний угол контейнера
- чтобы удалить график вместе с контейнером необходимо нажать кнопку с иконкой «крестика» в правом верхнем углу контейнера в режиме редактирования
- В область построения дашборда можно добавлять неограниченное количество контейнеров с графиками в соответствии с логикой сценарного анализа

5.6 Кросс-фильтрация

Кросс-фильтрация – функциональность, реализующая интерактивность и сценарный анализ при работе с конкретным дашбордом. Позволяет при отборе значений на одном из графиков перестроить информацию на всех остальных графиках в соответствии со сделанным отбором.

Важно, что система запоминает определенное число шагов «вперед – назад» при совершении отборов – 10 шагов. Таким образом, обеспечивается быстрый возврат к некоторому состоянию работы с дашбордом, которое позволяет продолжить работу с дашбордом (информацией) в другом направлении для ускорения получения ответов на вопросы в ходе сценарного анализа.

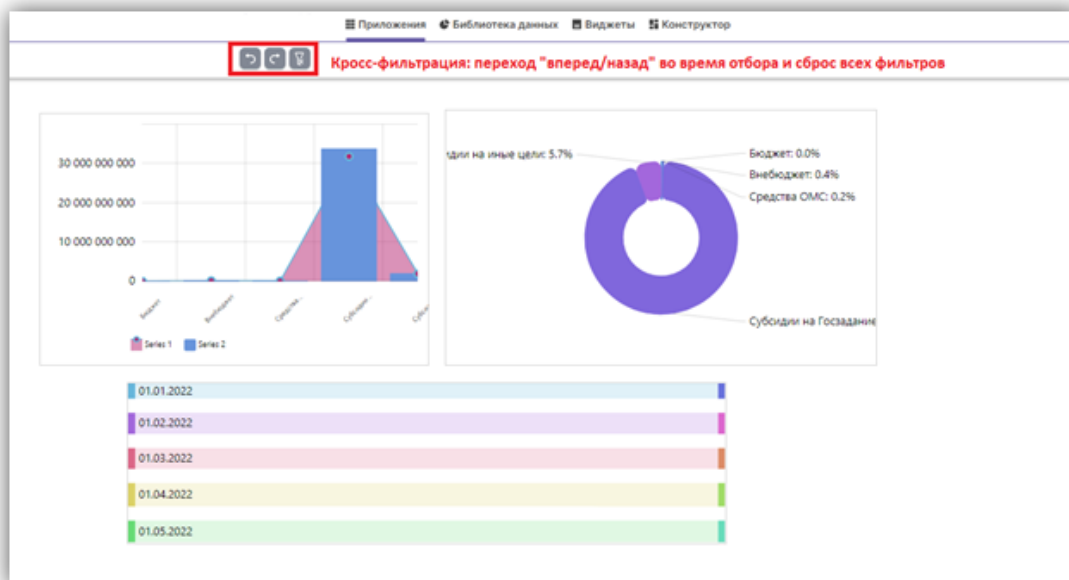


Рисунок 24. Навигация между отборами

5.7 Глобальные фильтры

Глобальные фильтры – функциональность, реализующая фильтрацию наборов данных сразу на всех дашбордах, входящих в приложение.

Например, если необходимо отфильтровать все графики дашбордов в приложении по периоду, то это можно сделать с помощью глобального фильтра, который создается на основе общего поля для всех источников данных графиков.

Настроенный глобальный фильтр отображается на второстепенной панели.



Рисунок 25. Просмотр глобального фильтра

При нажатии на кнопку Глобального фильтра открывается окно со всеми предустановленными для данного приложения фильтрами. Пользователь может выбирать необходимые значения по предоставленному полю.

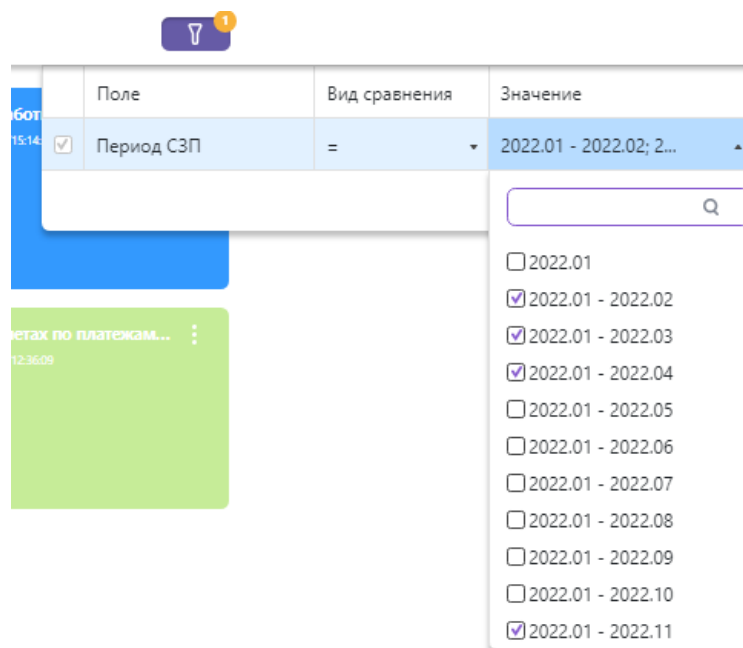


Рисунок 26. Выбор значений глобального фильтра

В качестве источников для дашборда могут использоваться как один источник, так и несколько. Важно, чтобы разные источники имели общее поле с одинаковыми значениями, по которому можно осуществить фильтрацию данных.

Функционал «Глобальные фильтры» настраивается Дата-инженером в Конструкторе.

6. Графики

6.1 Создание группы графиков

График, строится на основе выбранного шаблона визуализации данных.

Для того, чтобы создать новую группу, необходимо нажать на кнопку «Создать группу», ввести ее название и сохранить. После этого она отобразится в дереве.

Слева на экране отображаются все созданные группы графиков. По умолчанию всегда есть папка «Главная», в которую попадают все создаваемые графиков.

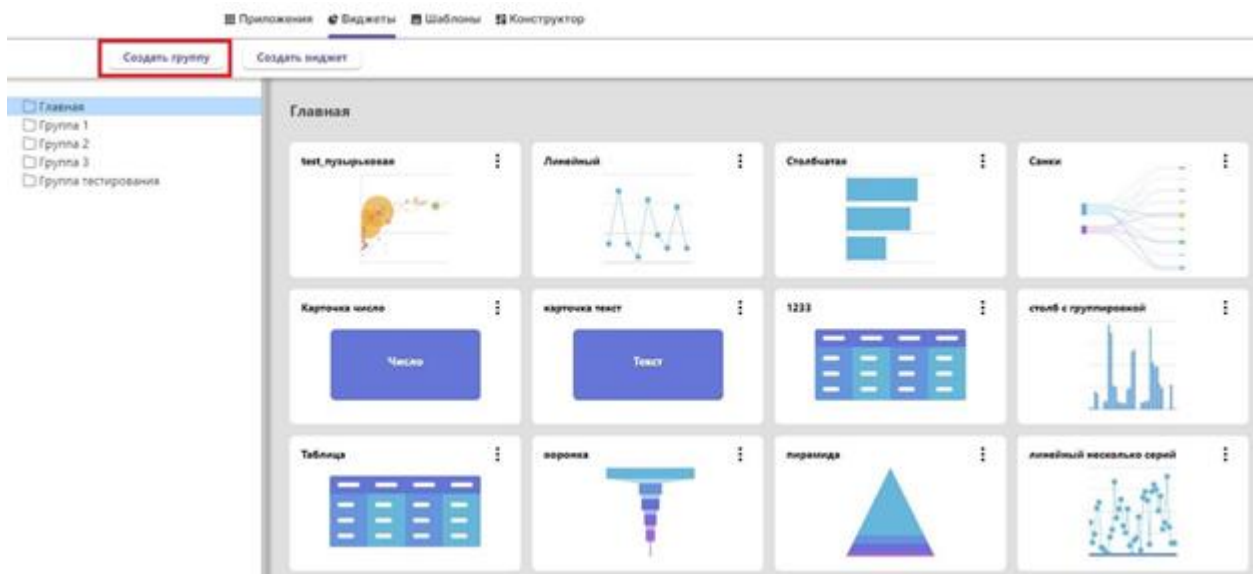


Рисунок 27. Создание группы

График можно перемещать между папками несколькими способами:

1. Захватить нужный график мышкой и перетащить в нужную папку на дереве
2. Нажать на «Свойства» графика, выбрать «Редактировать», выбрать необходимую группу

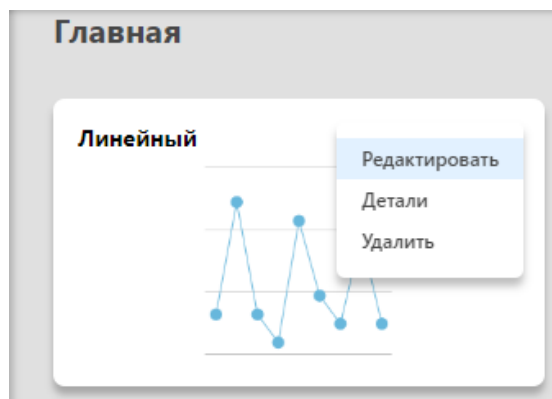


Рисунок 28. Свойства графика

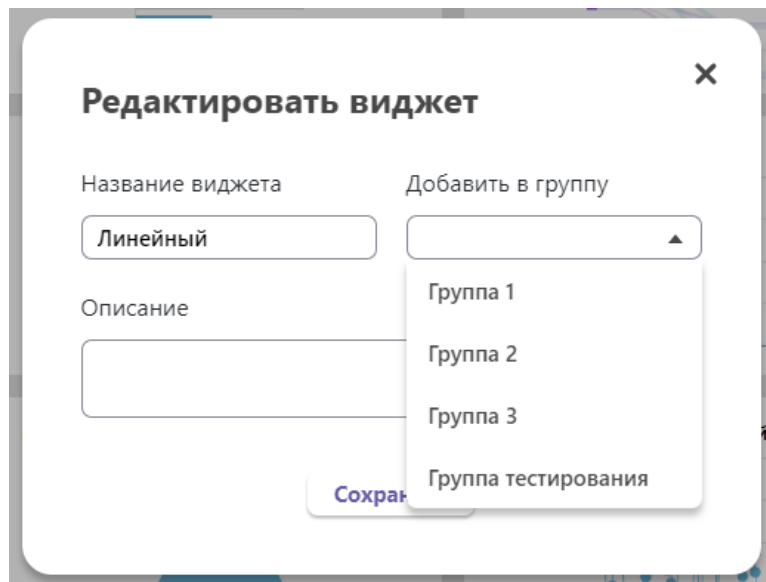


Рисунок 29. Выбор группы

8.1 Создание графика

Чтобы создать график необходимо нажать в разделе «Виджеты» на кнопку «Создать виджет».

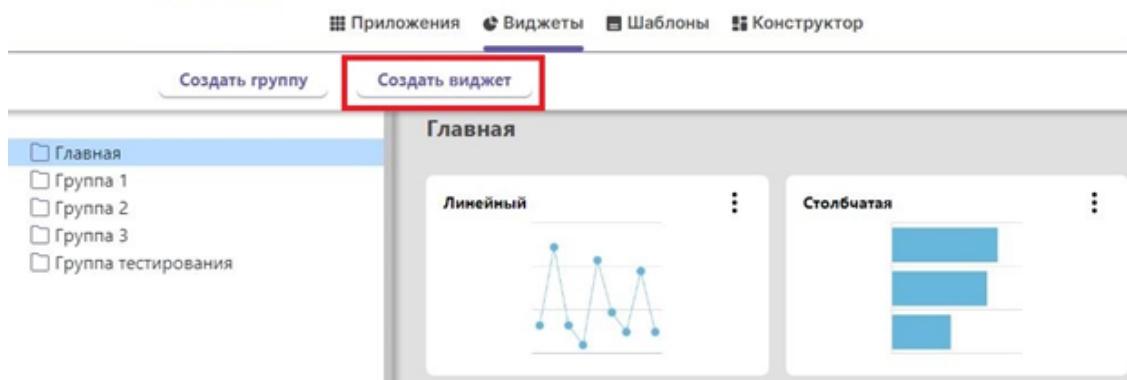


Рисунок 30. Создание виджета

На экране создания графика справа находится панель с настройками. На самом экране будет отображаться создаваемый график.



Рисунок 31. Экран создания виджетов

Для создания графика необходимо сделать несколько шагов:

1. Выбрать шаблон визуализации
2. Выбрать источник данных, необходимые колонки для построения, настроить для выбранных колонок группировку, сортировку и агрегацию
3. Выбрать, что является измерением и мерой
4. Настроить фильтрацию по колонкам (если необходимо)
5. Применить дополнительные настройки для выбранного шаблона, в том числе в Редакторе диаграмм
6. Сохранить график

Для переключения между шагами используется навигация сбоку в виде закладок. Закладки подсвечиваются разными цветами.

Значения цветов:

1. Фиолетовая полоска – шаг, на котором сейчас находится пользователь
2. Зеленая полоска – шаг пройден, выбраны все необходимые настройки
3. Серая полоска – шаг не пройден, требуется выбрать настройки
4. Закладка, полностью залитая серым – шаг недоступен (либо недоступен для конкретного Шаблона графика, либо настройки заполняются строго после другого шага)



Рисунок 32. Навигация между шагами создания графика

Шаг 1. Выбор шаблона визуализации

На первом шаге необходимо выбрать тип шаблона, который будет использоваться для создания графика. Для этого необходимо нажать на требуемый тип (чтобы увидеть название необходимо навести курсор на нужную иконку).

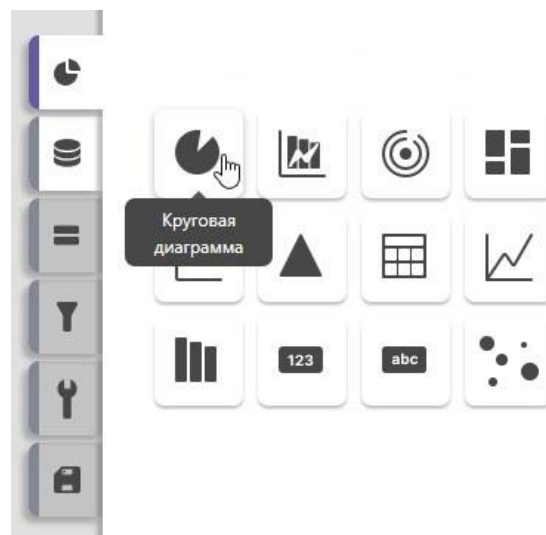


Рисунок 33. Выбор типа шаблона

После нажатия на иконку Пользователь переходит в список Шаблонов выбранного типа и нажимает на нужный Шаблон.

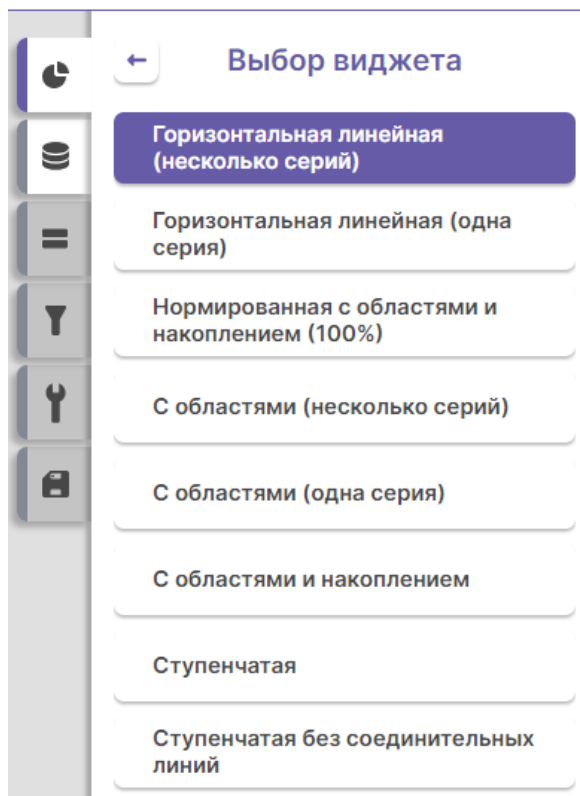
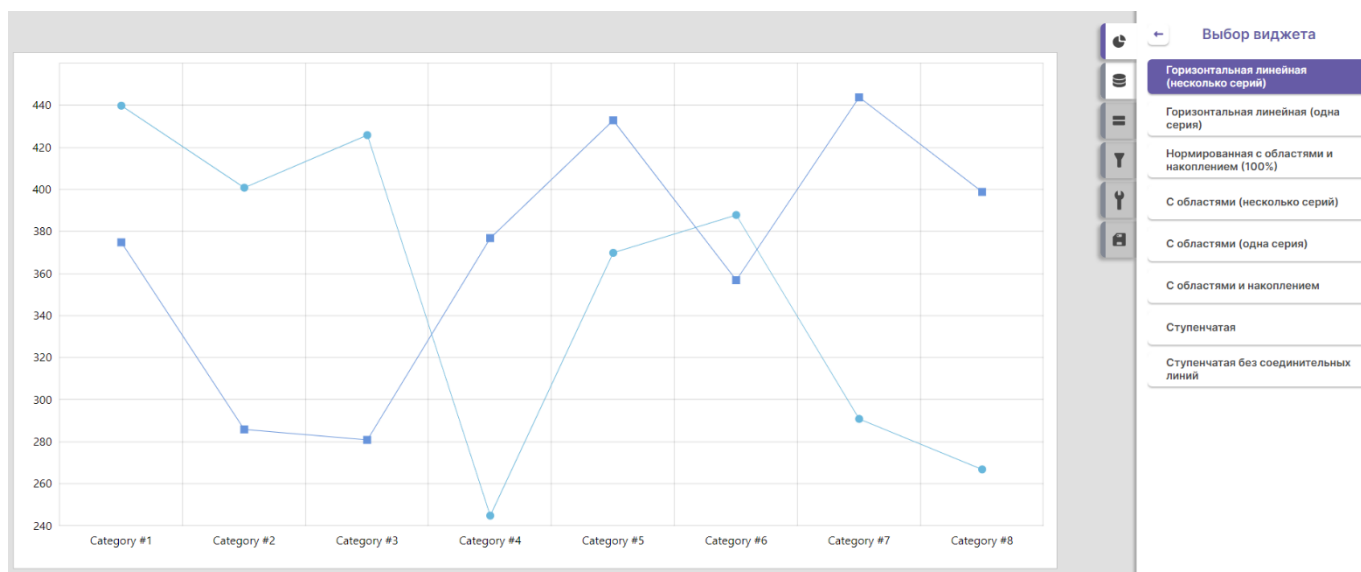


Рисунок 34. Выбор Шаблона

Как только Пользователь делает выбор на основном экране отображается выбранный Шаблон на тестовых данных. Только после выполнения Шага 3 Пользователь сможет увидеть график на реальных данных.

Рисунок 35. Отображение шаблона на тестовых данных



Далее Пользователю необходимо переключиться на Шаг 2 с помощью бокового меню с закладками.

Шаг 2. Выбор источника данных, колонок, настройка группировки, сортировки и агрегации

Выбор источника можно осуществить двумя способами:

1. Выбор доступной таблицы из выпадающего списка

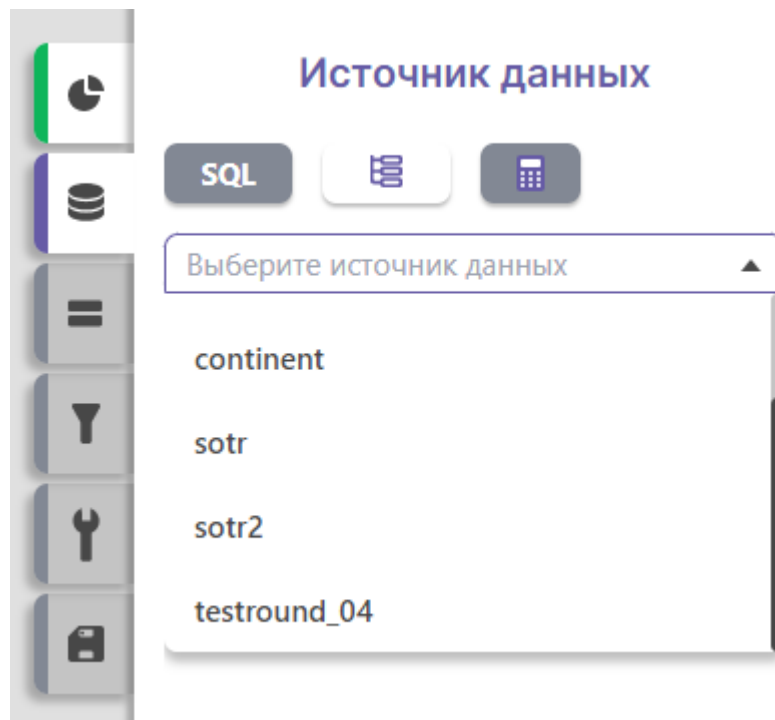


Рисунок 36. Выпадающий список с доступными таблицами

2. Воспользоваться деревом источников

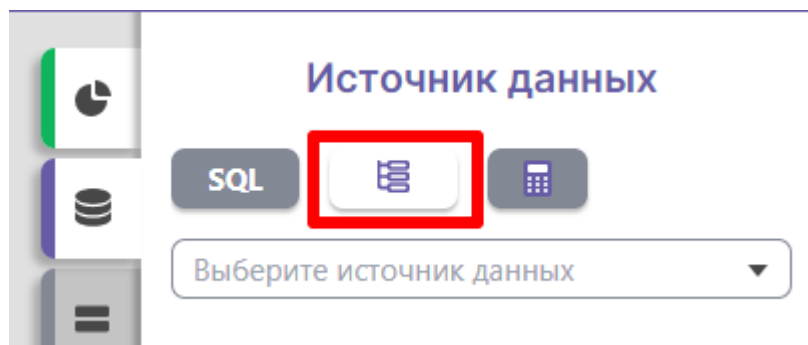


Рисунок 36. Вызов дерева источников

В дереве пользователю необходимо сначала выбрать ETL-хранилище или Аналитическое хранилище.

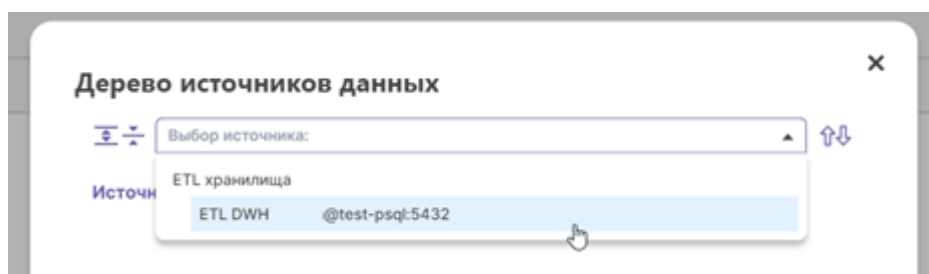


Рисунок 37. Выбор хранилища

После выбора станет доступно дерево всех таблиц, которые хранятся в выбранном хранилище. Для выбора необходимо кликнуть на таблицу и нажать на кнопку «Применить». В дереве отображаются только опубликованные таблицы. Чтобы таблица была опубликована, необходимо обратиться к пользователю с ролью «Дата инженер», который имеет доступ к разделу Конструктор.

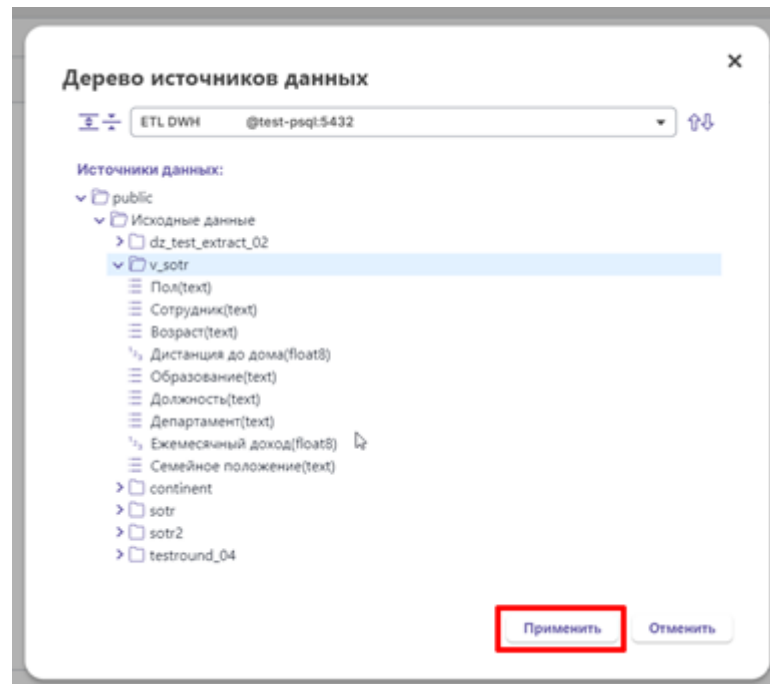


Рисунок 38. Выбор таблицы с данными

После выбора источника появляется список всех содержащихся в нем колонок. Наименование колонки, которое отображается – это бизнес-наименование, которое задается дата-инженером в разделе Конструктор. Если колонке не задано бизнес-наименование, то при построении графика будет использоваться исходное наименование из источника. Кроме наименования для каждой колонки указан тип данных.

Для того, чтобы выбрать колонки, на основе которых должен построиться график, необходимо нажать на колонку. Тогда она подсветится синим цветом и отобразится в Выбранных колонках.

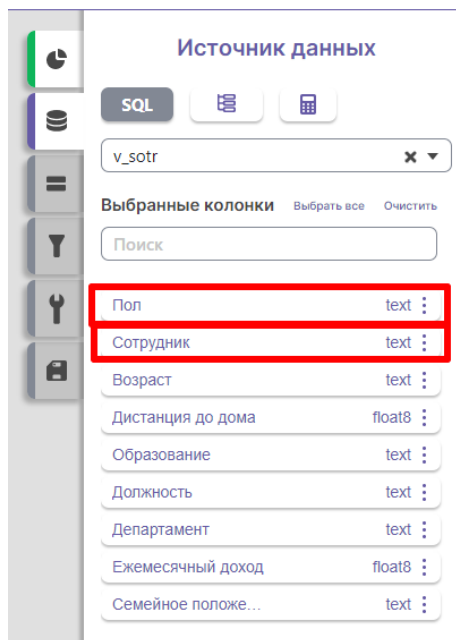


Рисунок 38. Выбор колонок

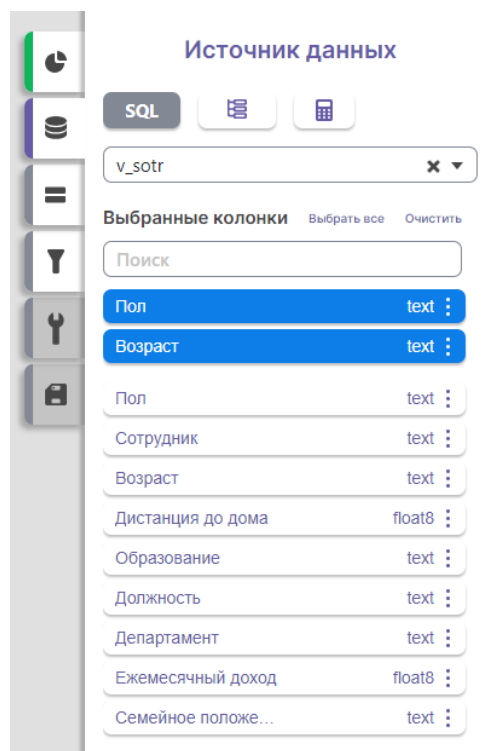


Рисунок 38. Отображение выбранных колонок

Для того, чтобы настроить группировку, сортировку и агрегацию, необходимо нажать на три точки для каждой выбранной колонки.

Функции агрегации данных:

- Сумма

- Количество
- Мин – минимальное значение
- Макс – максимальное значение
- Среднее

Остальные функции:

- Группировка
- Сортировка
- Отменить – отмена группировки, сортировки и агрегации

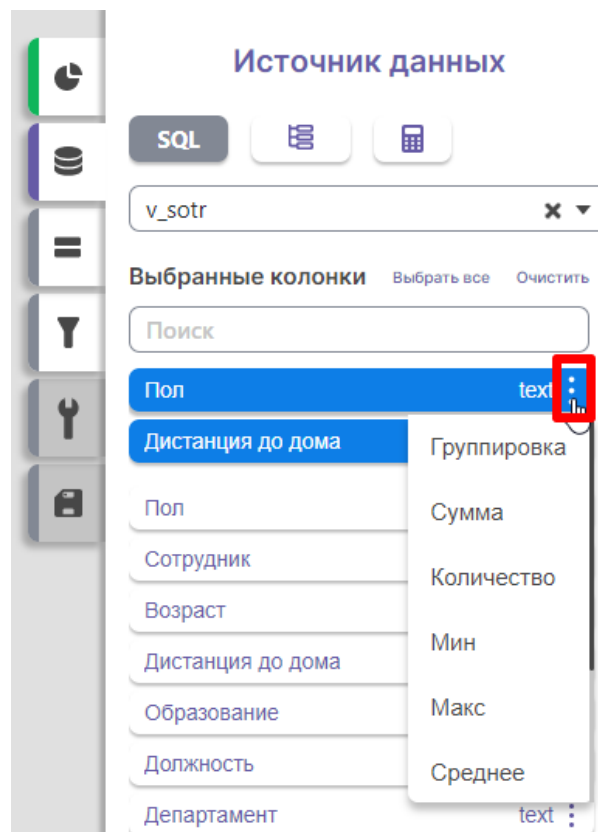


Рисунок 39. Выбор группировки, сортировки или агрегации

Шаг 3. Выбор измерения и меры

После настройки колонок, на которых будет строиться график, требуется выбрать измерение и меру. Для разных визуальных Шаблонов могут быть свои настройки:

- Для пузырьковой диаграммы – измерение, мера, ось x, ось y
- Для Sankey – откуда, куда и мера
- Для карточек с текстом и числом – раздел недоступен

- Для таблицы - раздел недоступен

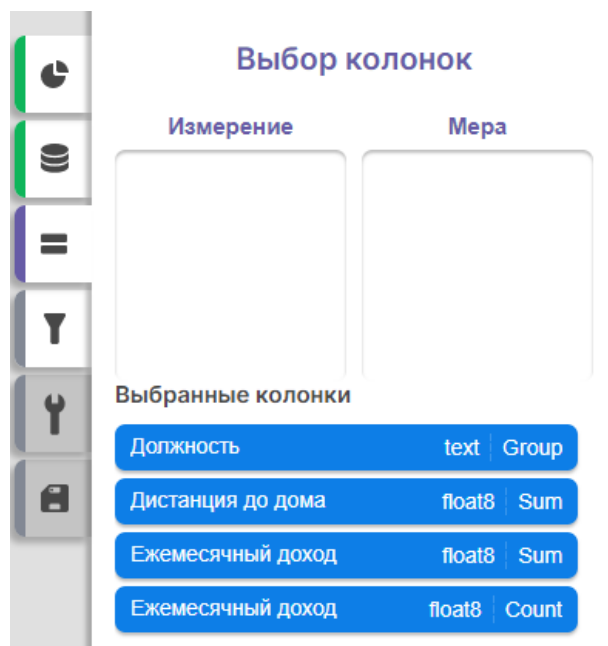


Рисунок 40. Выбор измерения и меры

Внизу видны все выбранные на предыдущем шаге колонки, которые с помощью drag-and-drop необходимо перетаскивать в нужные блоки.

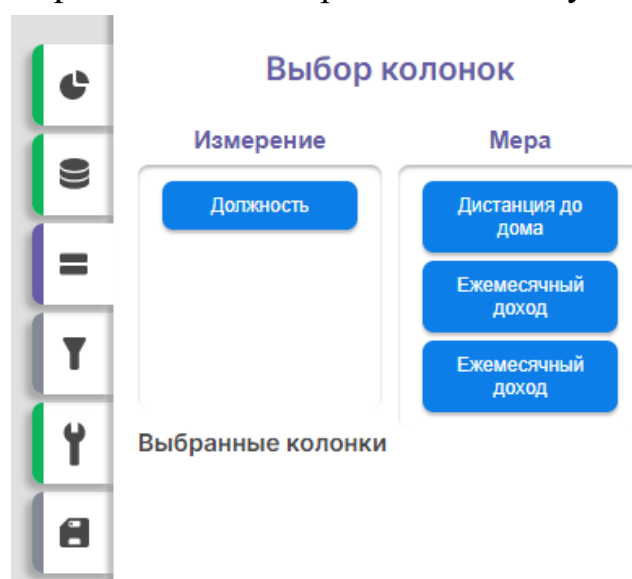


Рисунок 41. Результат перетаскивания колонок

После выбора меры и измерения (или других настроек) Пользователь сможет увидеть график на реальных данных.

Шаг 4. Настройка фильтрации по колонкам

На 4 шаге Пользователь может настроить фильтрацию для графика.

Для этого необходимо:

- выбрать столбец (из выпадающего списка) – необходимую колонку, для которой будет настраиваться фильтрация

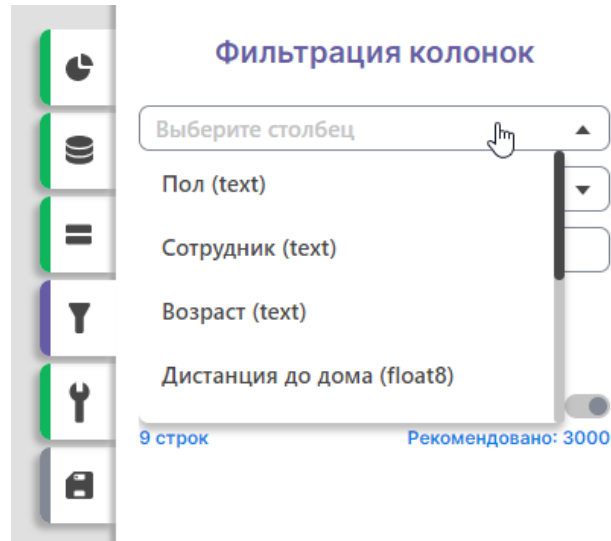


Рисунок 42. Выбор столбца

- выбрать тип фильтра (из выпадающего списка)
 - <
 - >
 - =
 - != (не равно)
 - В списке
 - Не в списке

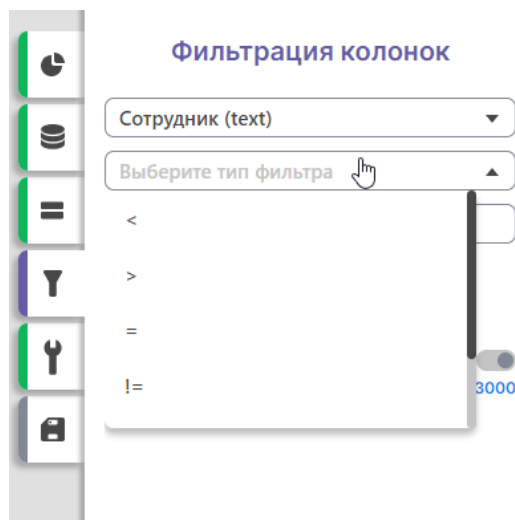


Рисунок 43. Выбор типа фильтра

- ввести значение, по которому будет производиться фильтрация:
 - для знаков < , > , = , != необходимо задавать значения вручную

- для типа фильтра «в списке», «не в списке» предлагается выбрать значения из списка, куда подгружаются все значения, содержащиеся в колонке

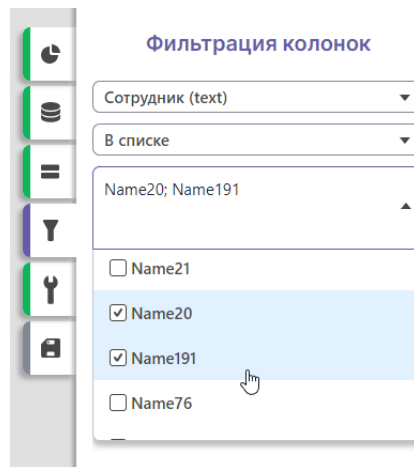


Рисунок 44. Выбор значений

После того, как заполнены все три поля необходимо нажать на кнопку «Применить». График перестроится согласно сформированному запросу.

Шаг 5. Дополнительные настройки

На 5 шаге есть возможность перейти в Редактор графиков с дополнительными настройками.

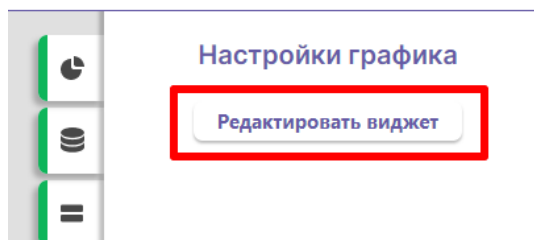


Рисунок 45. Переход в Редактор **виджетов**

Шаг 6. Сохранения графика

На финальном 6 шаге Пользователь должен сохранить график.

При создании нового графика есть возможность задать наименование и сохранить график.

При редактировании созданного ранее графика реализована возможность:

- Изменить название графика
- Сохранить текущий графика со всеми изменениями

- Сохранить скорректированный график, как новый с новым наименованием

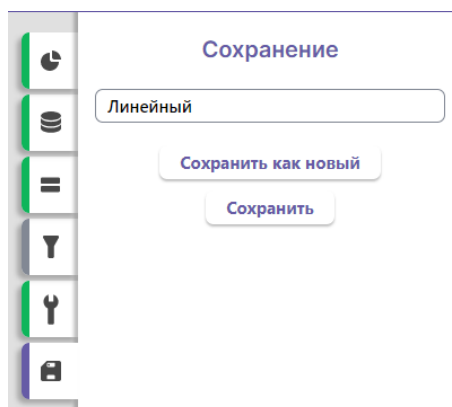


Рисунок 46. Сохранение графика

8.2 Редактор графиков

Редактор графиков представляет собой встроенный в систему инструмент, позволяющий осуществить редактирование визуального представления диаграмм.

Переход к редактору диаграмм осуществляется через меню «Настройки графика» при создании или редактировании графика.

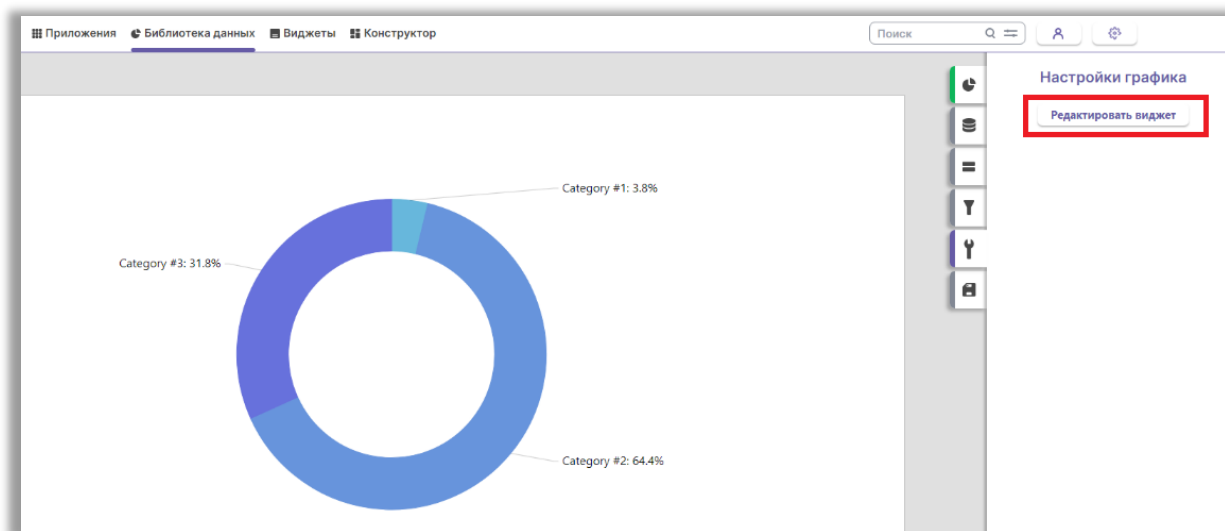
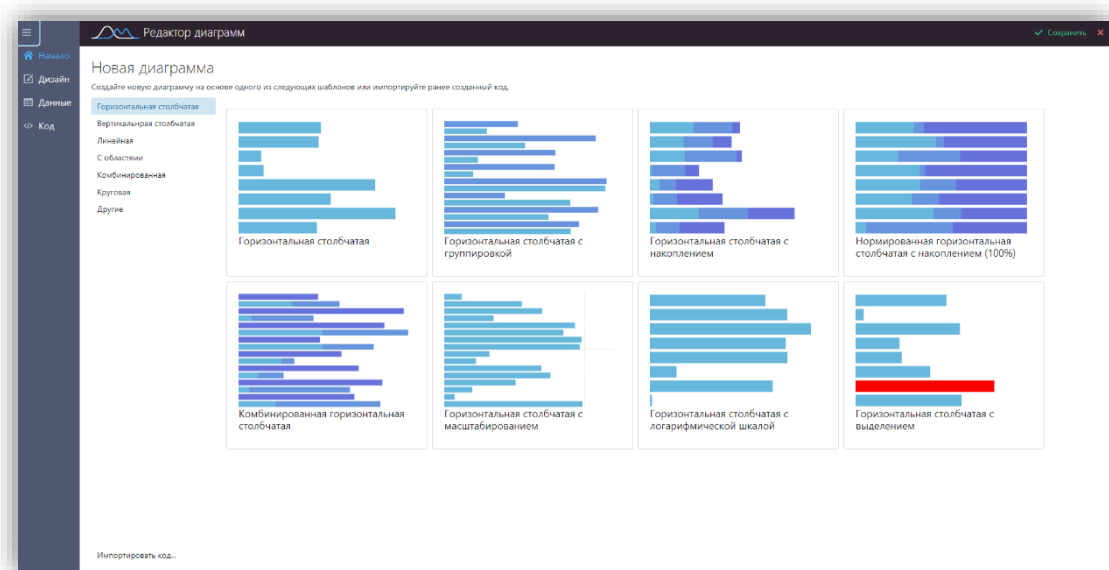


Рисунок 48. Переход в Редактор графиков

Общий вид стартового окна работы с редактором графиков выглядит следующим образом:



Отсуюнок 49. Стартовая страница редактора графиков

Основное меню в левом верхнем углу представлено следующими вкладками:

- **«Начало»** - содержит шаблоны диаграмм;
- **«Дизайн»** - настройки внешнего вида диаграмм;
- **«Данные»** - данные, используемые в качестве источника для диаграмм;
- **«Код»** - конфигурация диаграмм в виде кода.

На вкладке **«Дизайн»** содержатся настройки внешнего вида диаграмм. Редактор имеет сложную иерархичную структуру и разбит на основные блоки. Данная структура настроек аналогична для элементов диаграммы:

- Главные
- Данные
- Компоновка
- Внешний вид
- Состояние
- Интерактивность
- Доступность
- Другие

Настройки графиков

Настройки Осей X и Y

Задать/изменить заголовок оси X (ось категорий):

Главные – ОСИ X – ▼ – Выбрать существующую ось из списка – Главные – Заголовок

Задать/изменить заголовок оси Y (ось значений):

Главные – ОСИ Y – ▼ – Выбрать существующую ось из списка – Главные – Заголовок

Настройки подписи оси X (аналогично для Оси Y):

Главные – ОСИ X – ▼ – Выбрать существующую ось из списка – Внешний вид – ⚡Еще – Размер шрифта / Толщина шрифта / Шрифт / Размер шрифта / Толщина шрифта

Настройка поворота подписи оси X (аналогично для Оси Y):

Главные – ОСИ X – ▼ – Выбрать существующую ось из списка – Внешний вид – ⚡Еще – Средство визуализации – ⚡Еще – Подписи – Компоновка – Поворот

Настройка начального и конечного значения Оси X для Радара

Главные – ОСИ X – ▼ – Выбрать существующую ось из списка – Главные – Диапазоны осей – ▼ – Выбрать нужный элемент данных – Главные – Настроить начальное и конечное значение

Заливка сегментов радиуса для Радара

Главные – ОСИ X – ▼ – Выбрать существующую ось из списка – Главные – Диапазоны осей – ▼ – Выбрать нужный элемент данных – Главные – Заполнение оси – Внешний вид – ⚡Еще – Заливка

Настройка цвета контура для Радара

Главные – ОСИ X – ▼ – Выбрать существующую ось из списка – Главные – Диапазоны осей – ▼ – Выбрать нужный элемент данных – Главные – Заполнение оси – Внешний вид – ⚡Еще – Цвет контура

Настройка Оси Y в %

Необходимо найти в коде графика следующий код и добавить то, что выделено красным:

```

"yAxes": [{
  "type": "ValueAxis",
  "min": 0,
  "renderer": {
    "maxLabelPosition": 0.98
  },
  "numberFormatter": {
    "type": "NumberFormatter",
    "forceCreate": true,
    "numberFormat": "#.%"
  }
}],

```

Добавление новой серии

Редактор позволяет осуществить добавление сколько угодно серий различного вида, в том числе и для создания комбинированных диаграмм.

Комбинированная диаграмма – диаграмма, сочетающая в себе несколько видов диаграмм (например, линейная+столбчатая).

Создание визуального представления новой серии на диаграмме подразумевает под собой наличие данных для этой серии (необходимо заранее проверить, есть ли данные в источнике).

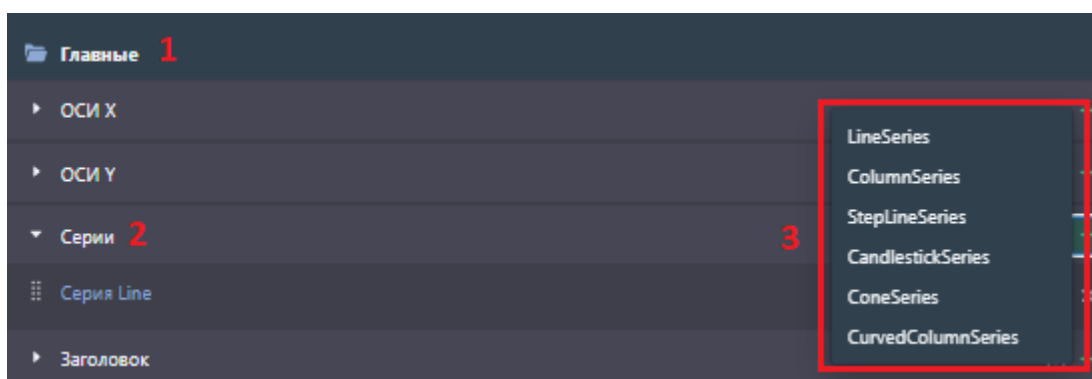


Рисунок 50. Добавление серий

Добавление новой серии Линейной (предварительно необходимо добавить данные в источник)

Главное – Серии (Наборы данных) – «+» – Серия Линейной (LineSeries) –

Данные – Поля данных – «+» – Поле категории X/Y – установить поле категории

X/Y (выбор из списка) – Поле значения X/Y – установить Поле значения X/Y (выбор из списка)

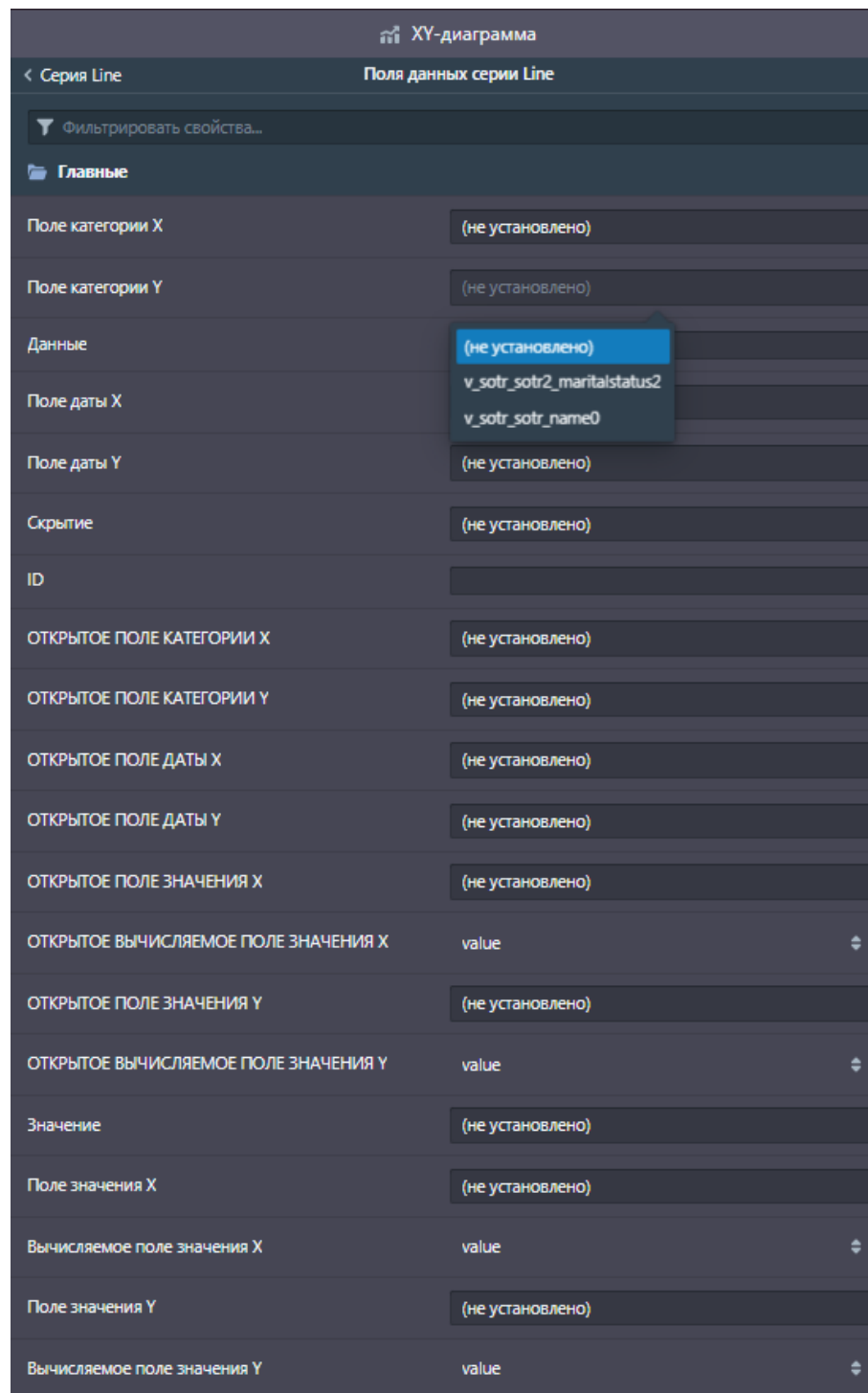


Рисунок 51. Добавление новой серии

Добавление новой серии Столчатой (предварительно необходимо добавить данные в источник)

Главные – Серии (Наборы данных) – «+» – Серия Столчатой (ColumnSeries) – Данные – Поля данных – «+» – Поле категории X/Y – установить поле категории X/Y (выбор из списка) – Поле значения X/Y – установить Поле значения X/Y (выбор из списка)

Добавление новой серии Ступенчатой (предварительно необходимо добавить данные в источник)

Главные – Серии (Наборы данных) – «+» – Серия Ступенчатой (StepLineSeries) – Данные – Поля данных – «+» – Поле категории X/Y – установить поле категории X/Y (выбор из списка) – Поле значения X/Y – установить Поле значения X/Y (выбор из списка)

Добавление новой серии Японских свечей (предварительно необходимо добавить данные в источник)

Главные – Серии (Наборы данных) – «+» – Серия Японских свечей (CandlestickSeries) – Данные – Поля данных – + – Поле категории X/Y – установить поле категории X/Y (выбор из списка) – Поле значения X/Y – установить Поле значения X/Y (выбор из списка)

Добавление новой серии Конуса (предварительно необходимо добавить данные в источник)

Главные – Серии (Наборы данных) – «+» – Серия Конус (ConeSeries) – Данные – Поля данных – «+» – Поле категории X/Y – установить поле категории X/Y (выбор из списка) – Поле значения X/Y – установить Поле значения X/Y (выбор из списка)

Добавление новой серии Закругленной серии (предварительно необходимо добавить данные в источник)

Главные – Серии (Наборы данных) – «+» – Серия закругленных столбцов (CurvedSeries) – Данные – Поля данных – «+» – Поле категории X/Y – установить поле категории X/Y (выбор из списка) – Поле значения X/Y – установить Поле значения X/Y (выбор из списка)

Настройки серий

Вкл/выкл накопление

Главные – Серии (Наборы данных) – ▼ – Выбрать нужную серию из списка –

Главные – Stacked

Заливка серии

для столбчатой диаграммы

Главные – Серии (Наборы данных) – ▼ – Выбрать нужную серию из списка –

Главные – Столбцы – Настройки – Внешний вид – Заливка – Цвет заливки серии столбчатой диаграммы (выбор из палитры)

для линейной диаграммы

Главные – Серии (Наборы данных) – ▼ – Выбрать нужную серию из списка – Внешний вид – ⇓Еще – Цвет контура – Цвет линии линейной диаграммы (выбор из палитры)

Настройка градиентной заливки (для столбчатой диаграммы)

Главные – Серии (Наборы данных) – ▼ – Выбрать нужную серию из списка – Главные – Столбцы – Внешний вид – ⇓Еще – Модификатор заливки – «+» – Линейный/радиальный градиент – Установка значений для каждого параметра в формате массива [#, #.#, #]

Настройка ширины столбцов

Главные – Серии (Наборы данных) – ▼ – Выбрать нужную серию из списка – Главные – Столбцы – Компоновка – ⇓Еще – Ширина

Настройка внешнего вида (добавление тени, размытие и т.д.)

Главные – Серии (Наборы данных) – ▼ – Выбрать нужную серию из списка – Главные – Столбцы – Внешний вид – ⇓Еще – Фильтры – Выбрать нужный фильтр

Настройка градиентной заливки (области)

Главные – Серии (Наборы данных) – ▼ – Выбрать нужную серию из списка – Главные – Части – Внешний вид – ⇓Еще – Модификатор заливки – «+» – Линейный/радиальный градиент – Установка значений для каждого параметра в формате массива [#, #.#, #]

Настройка градиентной заливки (для круговой диаграммы)

Главные – Серии (Наборы данных) – ▼ – Выбрать нужную серию из списка – Главные – Срезы – Внешний вид – ⇓Еще – Модификатор заливки – «+» – Линейный/радиальный градиент – Установка значений для каждого параметра в формате массива [#, #.#, #]

Заливка сегментов круговой диаграммы

Главные – Серии (Наборы данных) – ▼ – Выбрать нужную серию из списка – Внешний вид – ⇓Еще – Цвет – «+» – Главные – ⇓Еще – Список цветов – Установка значений для каждого сегмента в порядке категорий #e81717, #74c40c, #f5f107 (выбрать цвета самостоятельно и прописать их коды)

Настройки маркера линейной диаграммы

Цвет заливки маркера (точки) линейной диаграммы

Главное – Серии (Наборы данных) – ▼ – Выбрать нужную серию из списка – Внешний вид – ⇓Еще – Заливка (выбор из палитры)

Добавить маркеры (точки) линейной диаграммы

Главное – Серии (Наборы данных) – ▼ – Выбрать нужную серию из списка –

Главное – Маркеры – «+» – Выбрать тип маркера

Удалить маркеры (точки) линейной диаграммы

Главное – Серии (Наборы данных) – ▼ – Выбрать нужную серию из списка –

Главное – Маркеры – ▼ – ✕Удалить

Цвет заливки маркера (точки) линейной диаграммы

Главное – Серии (Наборы данных) – ▼ – Выбрать нужную серию из списка –

Главное – Маркеры – ▼ – Выбрать маркеры из списка – Внешний вид – Заливка

Настройки метки

Добавление метки. Работает с линейной, с областями. В скобках {} указывается наименование серии

Главное – Серии (Наборы данных) – ▼ – Выбрать нужную серию из списка –

Главное – Маркеры – «+» – LabelBullet – Главное – Подпись – «+» – Главное –

Текст – {value1}

Цвет текста метки

Главное – Серии (Наборы данных) – ▼ – Выбрать нужную серию из списка –

Главное – Маркеры – ▼ – Выбрать из списка LabelBullet – LabelBullet – Главное –

– Подпись – Внешний вид – Заливка (выбор из палитры)

Шрифт текста метки

Главное – Серии (Наборы данных) – ▼ – Выбрать нужную серию из списка –

Главное – Маркеры – ▼ – Выбрать из списка LabelBullet – LabelBullet – Главное –

– Подпись – Внешний вид – ⇓Еще – Шрифт

Размер шрифта текста метки

Главное – Серии (Наборы данных) – ▼ – Выбрать нужную серию из списка –

Главное – Маркеры – ▼ – Выбрать из списка LabelBullet – LabelBullet – Главное –

– Подпись – Внешний вид – ⇓Еще – Размер шрифта

Толщина шрифта текста метки

Главные – Серии (Наборы данных) – ▼ – Выбрать нужную серию из списка –
Главные – Маркеры – ▼ – Выбрать из списка LabelBullet – LableBullet – Главные
– Подпись – Внешний вид – ⇓Еще – Толщина шрифта

Формат числа метки (на сайте указан формат "#,###.#####", но не применяется)

Главные – Серии (Наборы данных) – ▼ – Выбрать нужную серию из списка –
Главные – Маркеры – ▼ – Выбрать из списка LabelBullet – LableBullet – Главные
– Подпись – Внешний вид – ⇓Еще – Форматирование числа – «+»

Настройка подсказок

Подсказка (для круговой диаграммы)

Главные – Серии (Наборы данных) – ▼ – Выбрать нужную серию из списка –
срезы – Интерактивность – ввести текст подсказки

- тип: {category}
- доля: {value.percent.formatNumber("#.#")}%
- количество объектов: {value.value}

Подсказка (для столбчатой)

Главные – Серии (Наборы данных) – ▼ – Выбрать нужную серию из списка –
срезы – Интерактивность – ввести текст подсказки

- месяц: {name}\n
- день: {categoryY}\n
- выручка: {valueX} руб.

Настройка подписей

Подпись - абсолютные значения (круговая)

Главные – Серии (Наборы данных) – ▼ – Выбрать нужную серию из списка –
Подписи – Главные – Текст – Установка формата подписи в виде

- {category:}
- {value}

Подпись (столбчатая диаграмма)

Главные – Серии (Наборы данных) – ▼ – Выбрать нужную серию из списка –
Метки – «+»

Затем зайти в код графика, найти следующий код и добавить то, что выделено красным:

```
{
  "values": [{
    "type": "LabelBullet",
    "label": {
      "type": "Label",
      "text": "{valueX}",
      "verticalCenter": "bottom",
      "truncate": false
    },
    "paddingLeft": 40,
    "paddingTop": 10,
    "horizontalCenter": "middle"
  }],
  "template": {
    "type": "Bullet"
  }
},
```

Перенос текста подписи оси

Зайти в код графика, найти следующий код и добавить то, что выделено красным:

```
"yAxes": [{
  "type": "CategoryAxis",
  "dataFields": {
    "category": ["zp_sep_org3"]
  },
  "renderer": {
    "labels": {
```

```
"template": {  
  "type": "AxisLabel",  
  "textAlign": "right",  
  "maxWidth": 600,  
  "wrap": true,  
  "fontSize": 16
```

Настройки заголовка графика

Заголовок графика (допустимо добавление нескольких заголовков)

Главные – Заголовки – «+»– Главные – Текст

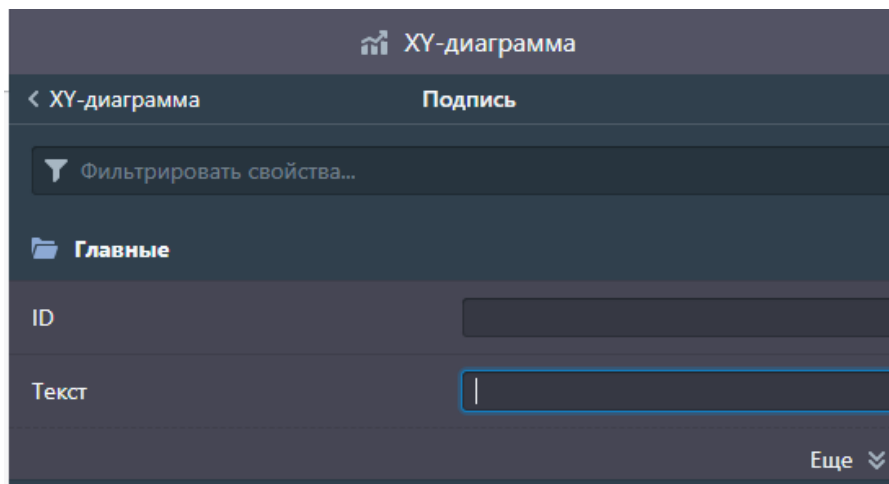


Рисунок 52. Ввод текста заголовка

Цвет текста заголовка

Главные – Заголовки – «+» – Внешний вид – ⚡Еще – Заливка

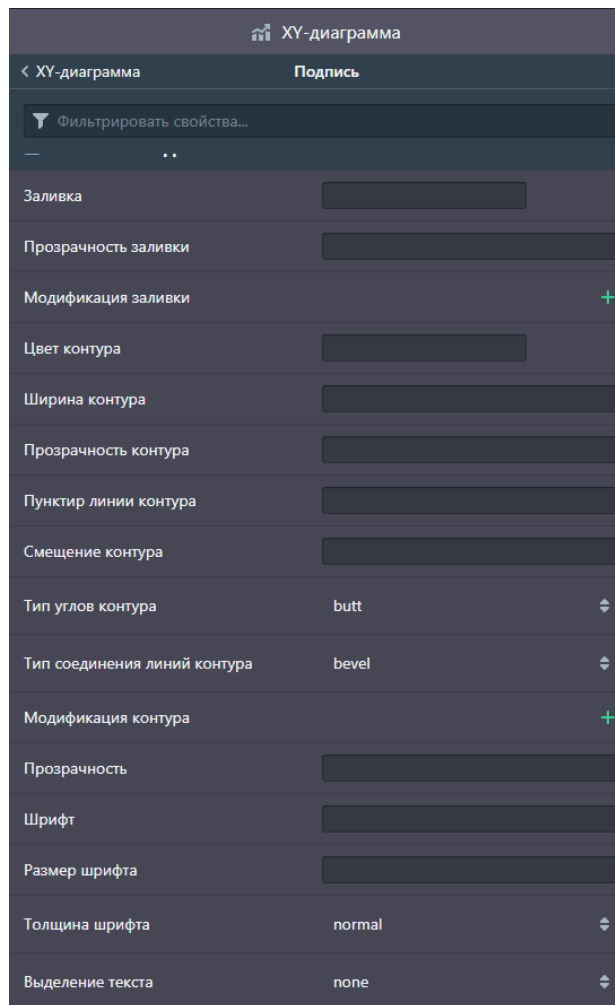


Рисунок 53. Настройки внешнего вида заголовка

Цвет контура текста заголовка

Главные – Заголовки – «+» – Внешний вид – ⚡Еще – Цвет контура

Шрифт заголовка

Главные – Заголовки – «+» – Внешний вид – ⚡Еще – Шрифт

Размер шрифта заголовка

Главные – Заголовки – «+» – Внешний вид – ⚡Еще – Размер шрифта

Толщина шрифта заголовка

Главные – Заголовки – «+» – Внешний вид – ⚡Еще – Толщина шрифта

Оформление текста заголовка (выбор: нет/подчеркнутый снизу/перечеркнутый...)

Главные – Заголовки – «+» – Внешний вид – ⚡Еще – Выделение текста

Выравнивание заголовка (слева/по центру...)

Главные – Заголовки – «+» – Компоновка – ⚡Еще – Выравнивание

Горизонтальное выравнивание текста заголовка (слева/по центру...)

Главные – Заголовки – «+» – Компоновка – ⇓Еще – Выравнивание текста

Вертикальное выравнивание текста заголовка (слева/по центру...)

Главные – Заголовки – «+» – Компоновка – ⇓Еще – Выравнивание по вертикали

Настройка прокрутки

Горизонтальная прокрутка

Главные – Горизонтальная прокрутка – «+»

Вертикальная прокрутка

Главные – Вертикальная прокрутка – «+»

Настройки легенды

Положение легенды (снизу диаграммы/справа от диаграммы...)

Главные – Легенда – «+» – Компоновка – ⇓Еще – Положение

Текстовое поле для ввода наименования шрифта легенды

Главные – Легенда – «+» – Внешний вид – ⇓Еще – Шрифт

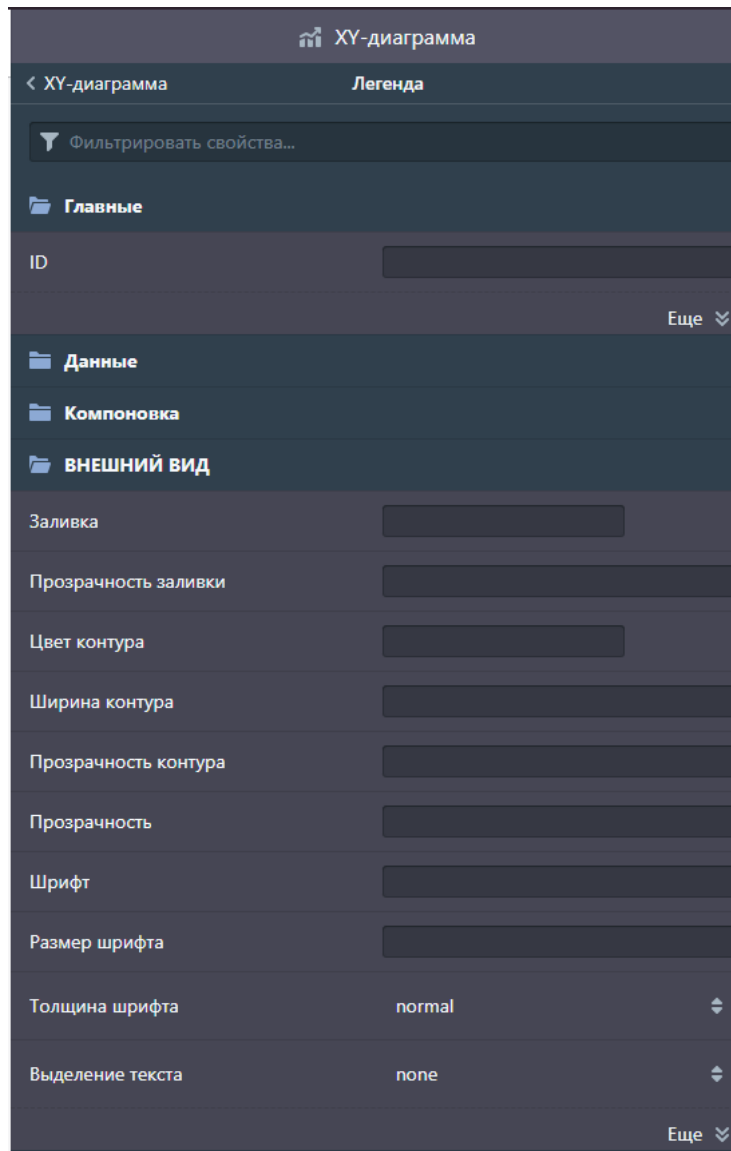


Рисунок 54. Настройки внешнего вида легенды

Размер шрифта легенды

Главные – Легенда – «+» – Внешний вид – ⚡Еще – Размер шрифта

Толщина шрифта легенды

Главные – Легенда – «+» – Внешний вид – ⚡Еще – Толщина шрифта

Оформление текста легенды (выбор: нет/подчеркнутый снизу/перечеркнутый...)

Главные – Легенда – «+» – Внешний вид – ⚡Еще – Выделение текста

Поворот текста легенды

Главные – Легенда – «+» – Главные – ⚡Еще – Подписи – Компоновка – Поворот

Легенда - для круговой отключить %

Легенда – «+» – Значение подписи – отключение подписи

Настройки положения диаграммы

Для всех типов диаграмм

Компоновка – Отступ слева

Компоновка – Отступ сверху

Компоновка – Отступ справа

Компоновка – Отступ снизу

Компоновка – Выравнивание (горизонтальное выравнивание диаграммы)

Компоновка – Выравнивание по вертикали (вертикальное выравнивание диаграммы)

Компоновка – Ширина (настройка ширины диаграммы, %)

Компоновка – Высота (настройка высоты диаграммы, %)

Компоновка – Горизонтальное смещение

Компоновка – Вертикальное смещение

Компоновка – Z индекс (смещение диаграммы по оси Z)

Компоновка – Поворот (поворот диаграммы вокруг своей оси, по умолчанию значение «0»)

Компоновка – Масштаб (по умолчанию значение «1»)

Для Радара

Компоновка – ↯Еще – Внутренний радиус

Компоновка – ↯Еще – Радиус

Для Sankey

Зайти в код графика, найти следующий код и добавить то, что выделено красным:

```
"links": {  
  "template": {  
    "type": "SankeyLink"  
  }  
},  
"dataFields": {  
  "value": "o",  
  "toName": "o",
```

```
"fromName": "o"  
},  
"paddingRight": 200  
}
```

7. Шаблоны

7.1 Назначение

Раздел «Шаблоны» – каталог шаблонов элементов визуального отображения информации.

Шаблон — это Тип визуального представления информации. Например: кольцевая диаграмма, таблица, карточка с текстом, столбчатый график.

В системе доступны следующие типы визуальных шаблонов:

1. Круговая диаграмма
2. Комбинированный график
3. Радар
4. Sankey
5. Пирамида/воронка
6. Таблица
7. Линейный график
8. Столбчатые график
9. Карточка с числом
10. Карточка с текстом
11. Пузырьковая
12. Дерево

Для некоторых типов визуальных шаблонов есть подтипы.

1) Круговая диаграмма

- a. Круговая - Pie Chart (пай-чарт) - визуализация, представленная в виде круга.

Он разделен на сектора, размер которых соответствует значению каждой категории данных.

Благодаря этому сразу понятно, что играет большую, а что меньшую роль в составе целого.

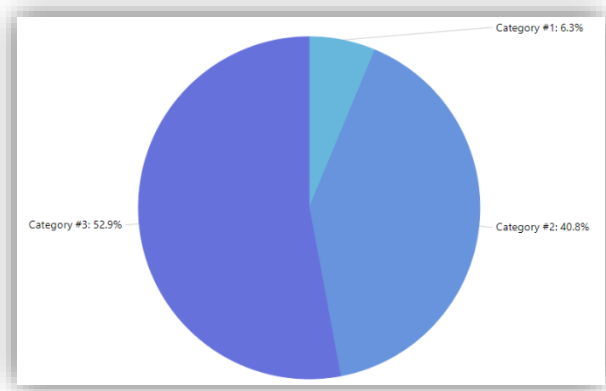


Рисунок 55. Круговая диаграмма

- в. Кольцевая - разновидность круговой диаграммы. Представляет собой тот же разделенный на части круг, только с пустотой посередине, то есть в форме бублика.

Все правила разработки и оформления одинаковы для обоих вариантов.

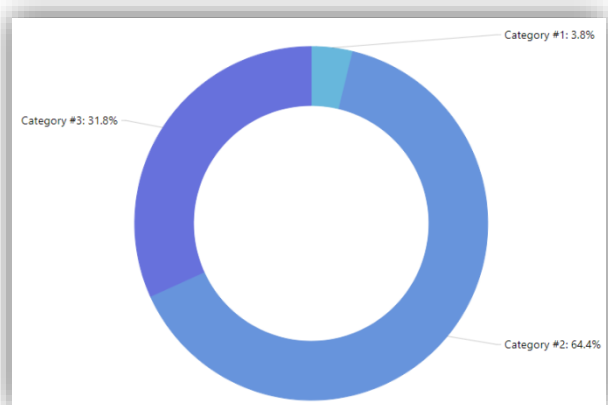


Рисунок 56. Кольцевая диаграмма

- с. Кольцевая со сглаженными углами – разновидность кольцевой диаграммы с измененным визуальным представлением – сглаженными углами секторов.

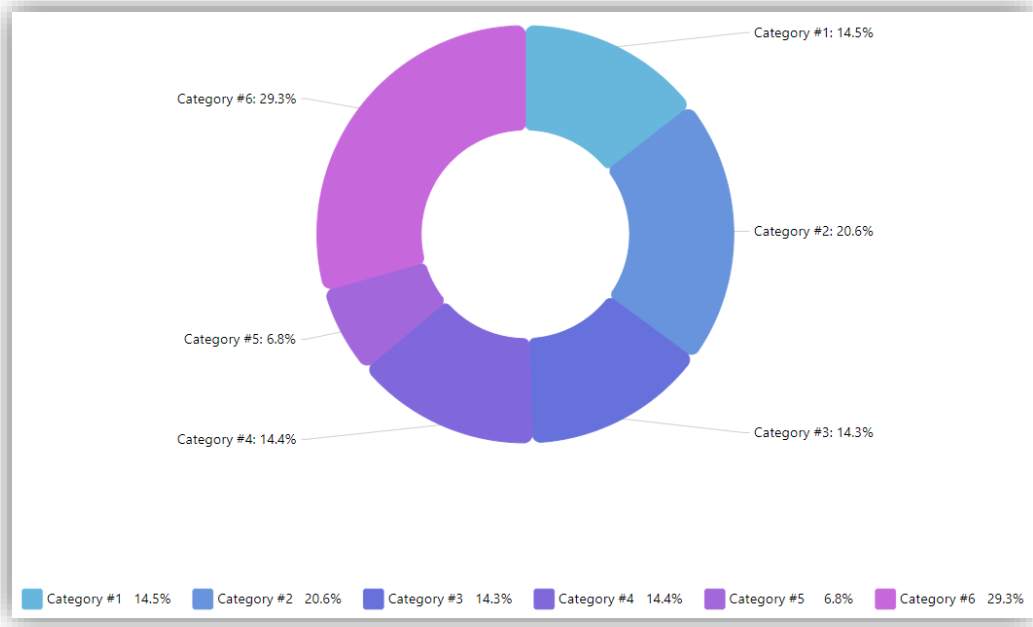


Рисунок 57. Кольцевая со сглаженными углами

2) Комбинированный график подходит для сравнения двух наборов значений мер, которые обычно трудно сравнивать из-за различий в масштабе. Комбинированная диаграмма может отображаться только вертикально.

а. Столбчатая / линейная

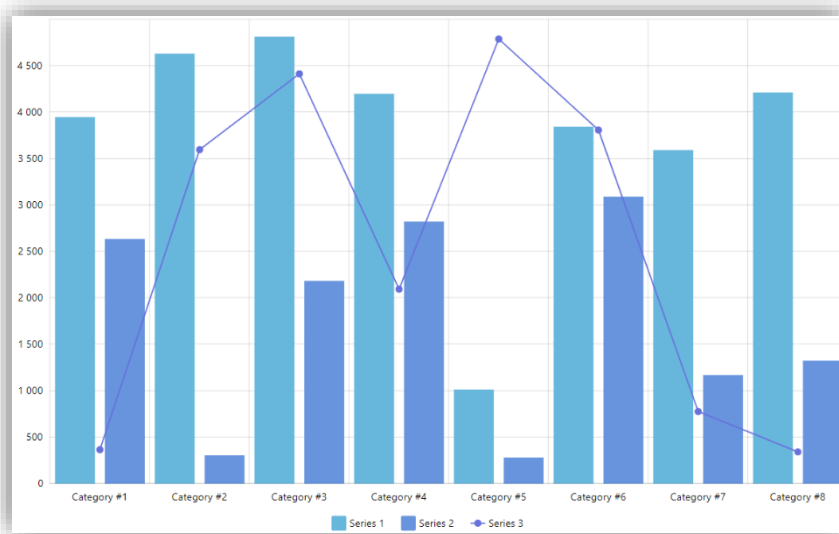


Рисунок 58. Столбчатая и линейная

б. Столбчатая / с областями

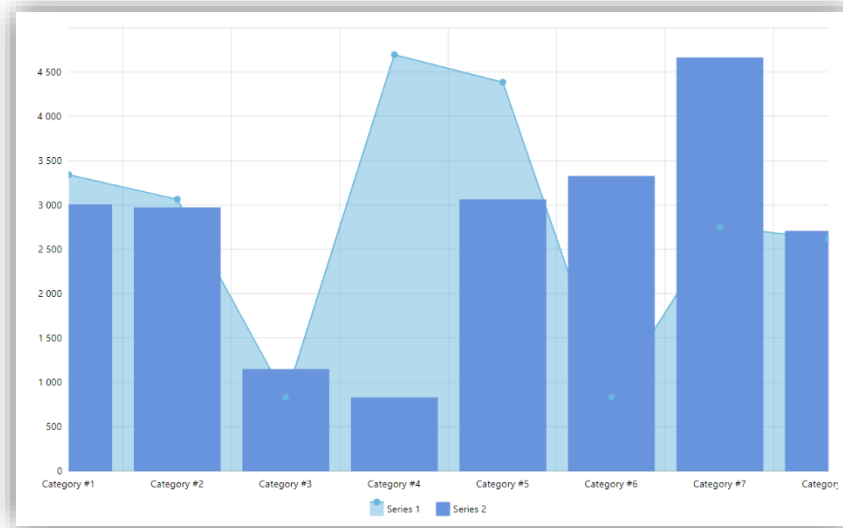


Рисунок 59. Столбчатая с областями

3) Диаграмма «Радар» представляет собой диаграмму с двумя измерениями, на радиальных осях которой отображаются одна или несколько групп значений, для сопоставления которых используются несколько переменных.

Диаграммы Радар можно использовать для визуализации и сравнения показателей с заданными стандартными значениями или результатами, показанными какой-либо группой.

Для диаграмм Радар необходимо использовать два измерения и одну меру. Ось Y проходит от центра к периметру, ось X представляет собой периметр диаграммы. Каждое значение представляет расстояние до центра диаграммы и отображается на осях, выходящих из центра.

Центр диаграммы представляет минимальное значение, а периметр — максимальное.

Диаграмма Радар поддерживает отображение нескольких последовательностей, каждая из которых связана с отдельными значениями.

Данные используются в диаграмме только один раз.

Добавление значений способствует повышению степени детализации диаграммы.

Диаграмма Радар состоит из сегментов (осей), разделенных одинаковыми интервалами и выходящих из одной точки.

Каждому значению данных второго измерения соответствует сегмент. Значение каждой меры представлено в виде узла на соответствующей оси.

Значения, относящиеся к одной последовательности, соединены линией. Таким образом, диаграмма имеет форму звезды.

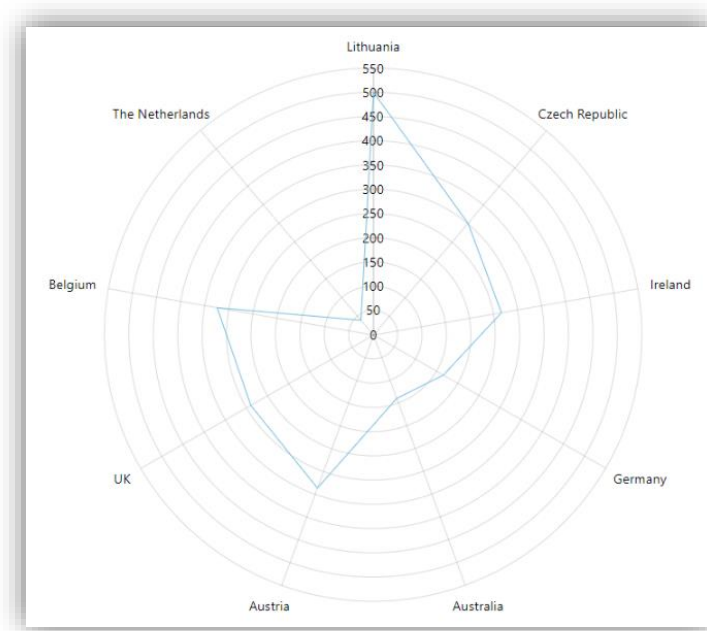


Рисунок 60. Радар

4) Потокковая диаграмма Санкей - Sankey - одна из диаграмм для иллюстрации бизнес-процессов.

Санкей отражает ключевые шаги, интенсивность протекания на каждом участке.

Линии могут соединяться или разветвляться.

Ширина линий напрямую связана с долей потока: чем больше параметр, тем толще линия.

Связываемые вещи называются узлами, а соединения — линками.

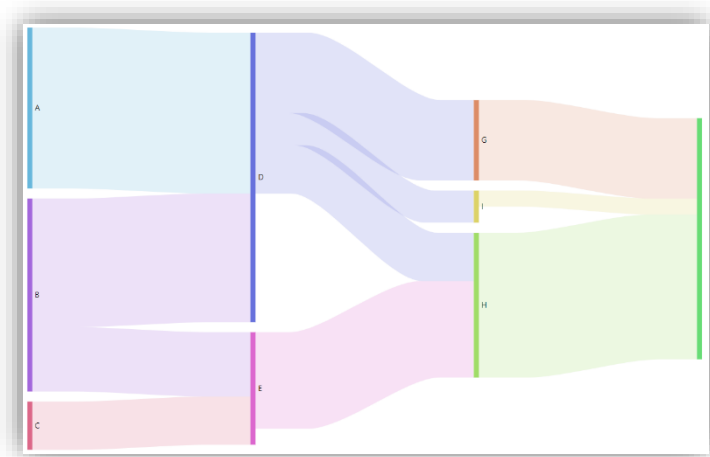


Рисунок 61. Диаграмма Sankey

5) Пирамида / Воронка

а. Пирамида

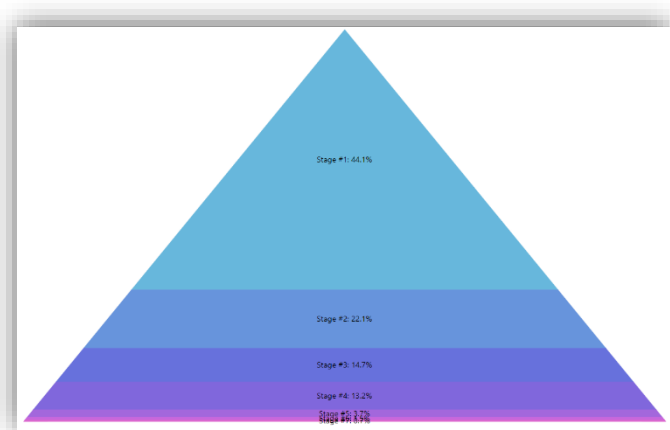


Рисунок 62. Пирамида

б. Диаграмма Воронка (Funnel chart) позволяет пользователям добавлять диаграмму последовательности для представления связанных этапов процесса.

Объем данных на каждом последующем этапе уменьшается.

Последующий этап должен содержать подмножество данных предыдущего этапа.

Постепенное уменьшение приводит к отображению диаграммы в виде сужающейся воронки.

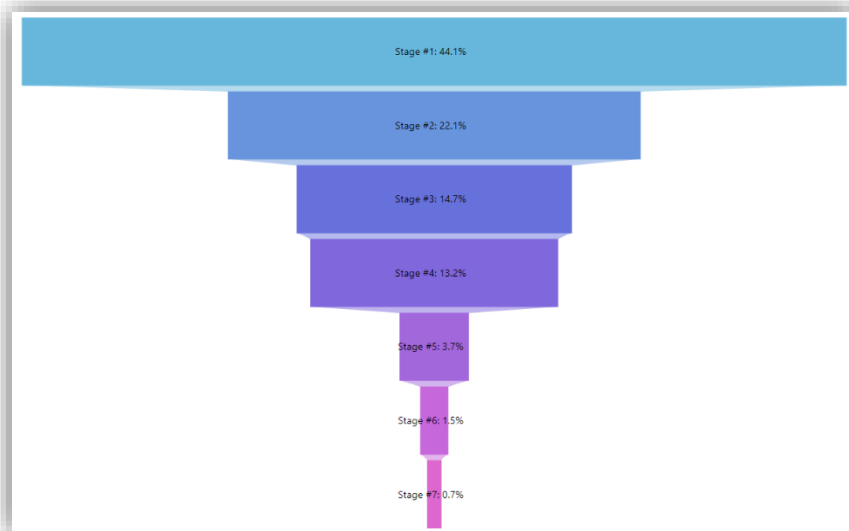


Рисунок 63. Пирамида

б) Таблица - очень важный и иногда незаменимый вариант визуализации. Их рекомендуется использовать в случаях, когда мы показываем единичные или текстовые значения, не взаимосвязанные показатели.

По расположению на дашборде таблицы желательно размещать на третьем уровне информации.

Не стоит использовать сложные таблицы с большим количеством строк или столбцов: все должно читаться без усилий.

Группы	demo_dz_data_ostatoknacionesyaca6
Организация 05 (1)	181 060 344.66
Организация 06 (1)	460 846 033.64
Организация 07 (1)	6 217 417.97
Организация 08 (1)	171 593 726.04
Организация 09 (1)	1 310 101 005.29
Организация 10 (1)	39 131 506.46
Организация 11 (1)	151 104 138.79
Организация 12 (1)	199 834 187.08
Организация 13 (1)	137 038 868.22
Организация 14 (1)	623 556.64
Организация 15 (1)	5 309 862.7
Организация 16 (1)	976 329 822.98
Организация 17 (1)	1 856 808 657.47
Организация 18 (1)	9 881 893.76
Организация 19 (1)	1 932 875 269.16
Итого:	8 992 195 401.76

Рисунок 64. Таблица

7) Линейный график — это основной инструмент отображения динамических изменений, очень популярный благодаря своей понятности и универсальности.

Используйте линейный график, когда хотите показать изменения данных во времени.

а. Горизонтальная линейная (несколько серий)

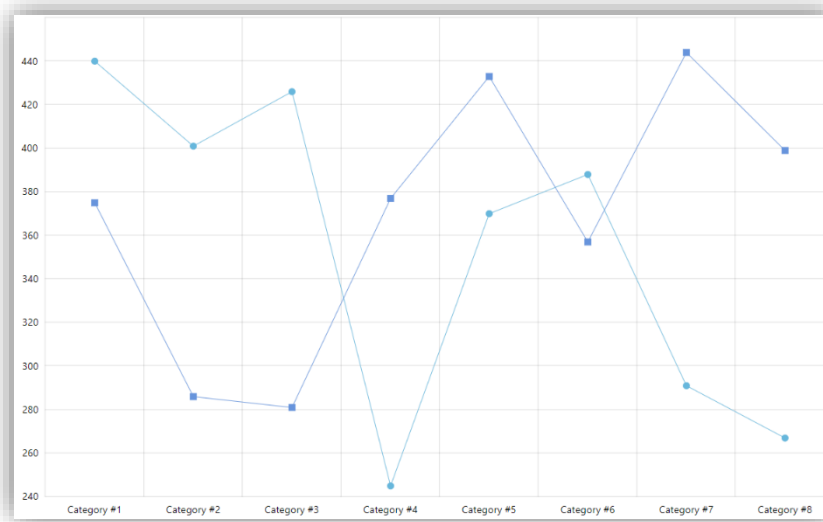


Рисунок 65. Горизонтальная линейная

б. Горизонтальная линейная (одна серия)

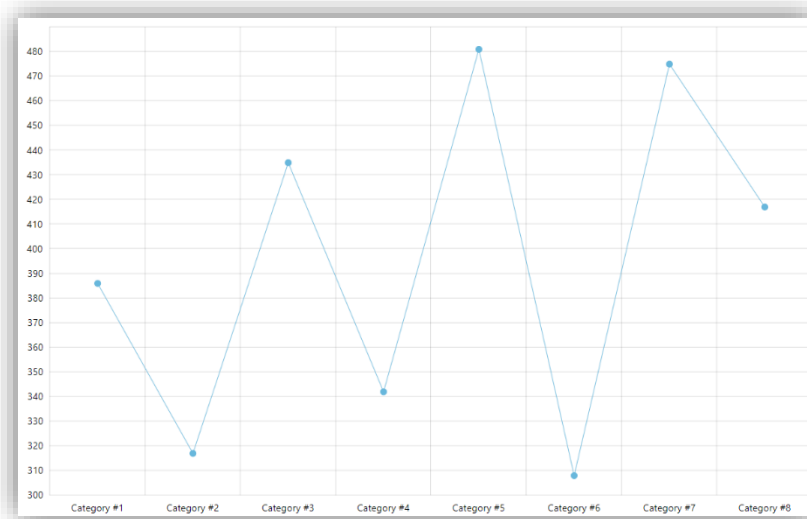


Рисунок 66. Горизонтальная линейная (одна серия)

с. Нормированная с областями и накоплением (100%) - данный тип графика стоит использовать, если необходимо подчеркнуть различие между данными.

Такой тип диаграммы не подходит для большого количества данных, потому что визуально их будет сложно воспринимать. Используйте полупрозрачные цвета для лучшего восприятия.

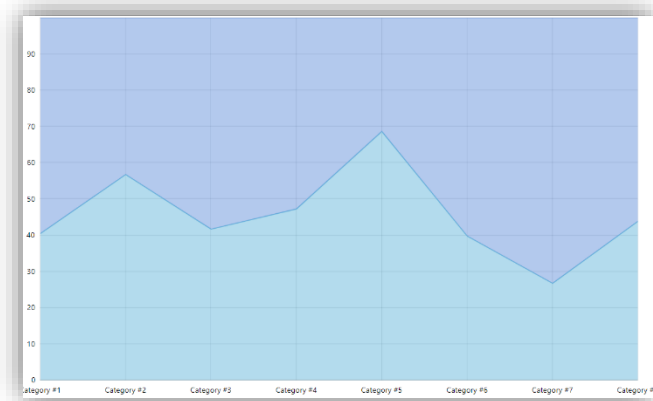


Рисунок 67. Нормированная с областями и накоплением (100%)

д. С областями (несколько серий)

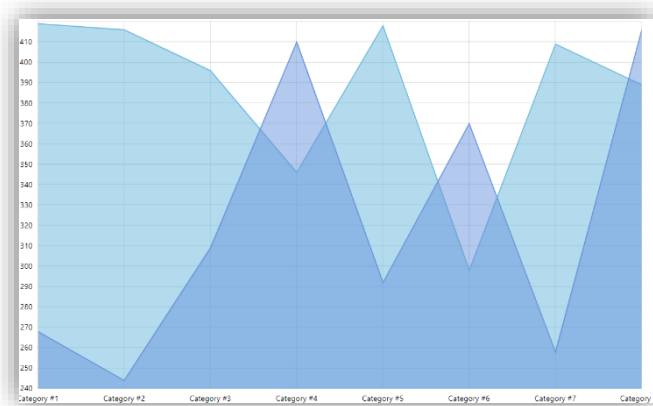


Рисунок 68. С областями (несколько серий)

е. С областями (одна серия)

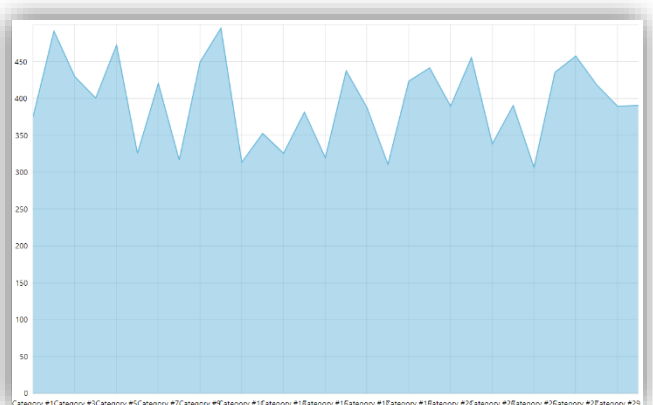


Рисунок 69. С областями (одна серия)

f. С областями и накоплением

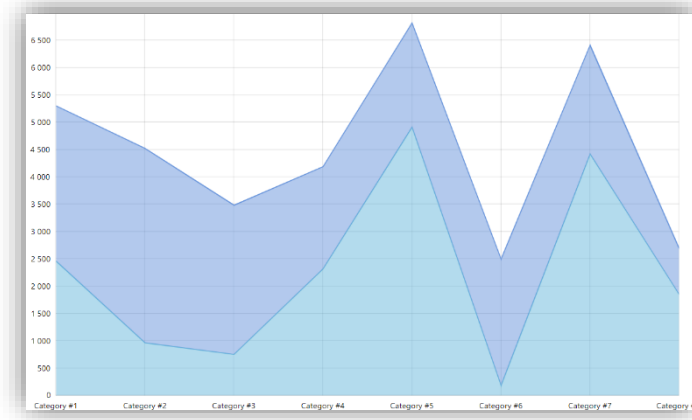


Рисунок 70. С областями и накоплением

г. Ступенчатая

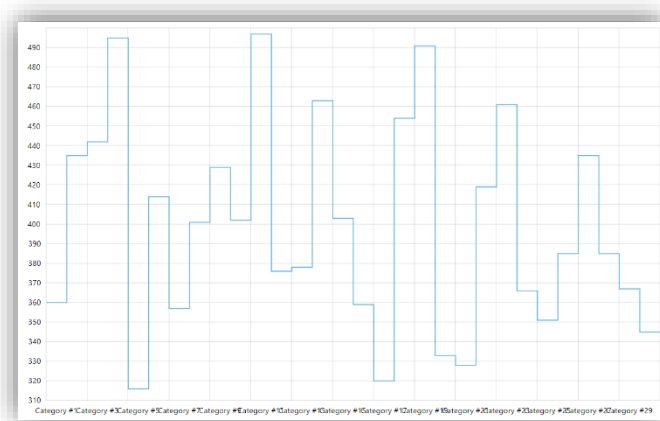


Рисунок 71. Ступенчатая

а. Ступенчатая без соединительных линий

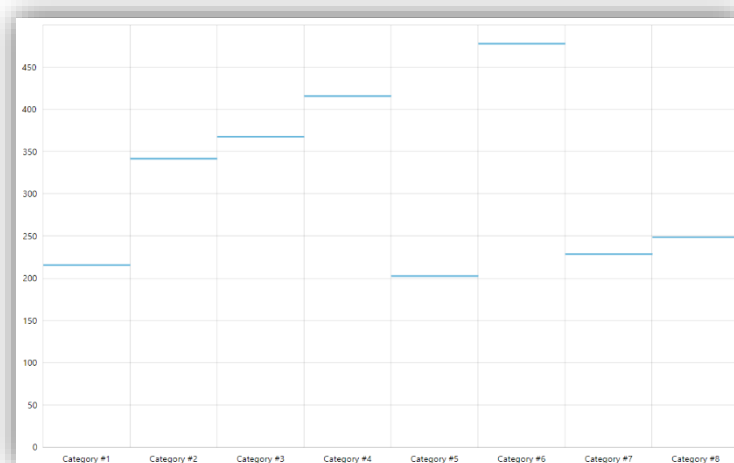


Рисунок 72. Ступенчатая без соединительных линий

8) Столбчатый график - самый распространенный тип визуализаций, ведь при всем доступном многообразии визуализаций, наверное, нет ничего более наглядного, понятного и легко читаемого.

Столбчатые диаграммы прекрасно подходят для отображения рейтинга, но их можно использовать и для того, чтобы показать динамику, и для других видов анализа.

а. Вертикальная столбчатая

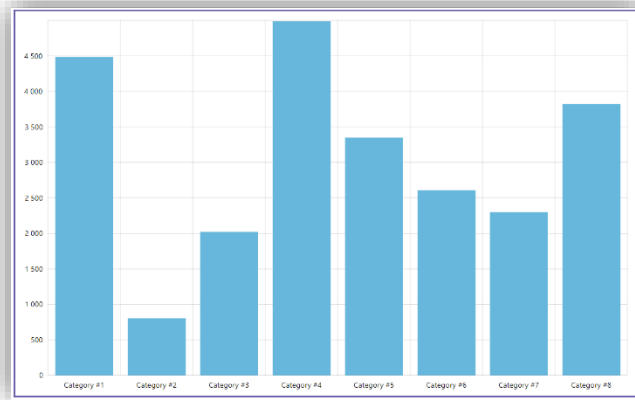


Рисунок 73. Вертикальная столбчатая

б. Вертикальная столбчатая с группировкой

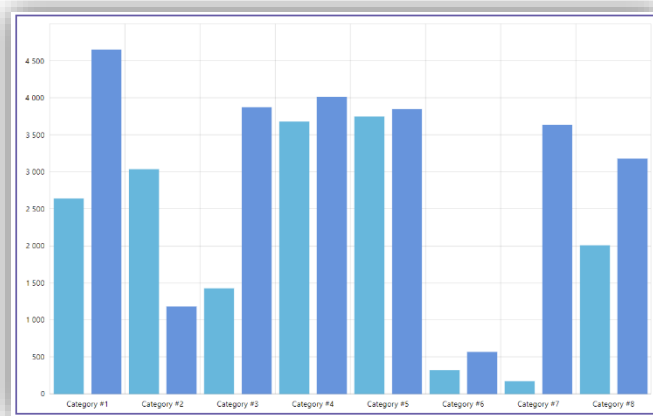


Рисунок 74. Вертикальная столбчатая с группировкой

с. Вертикальная столбчатая с накоплением

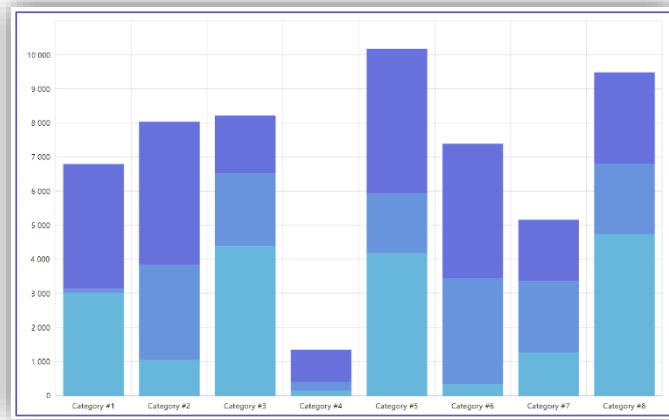


Рисунок 75. Вертикальная столбчатая с накоплением

д. Горизонтальная столбчатая

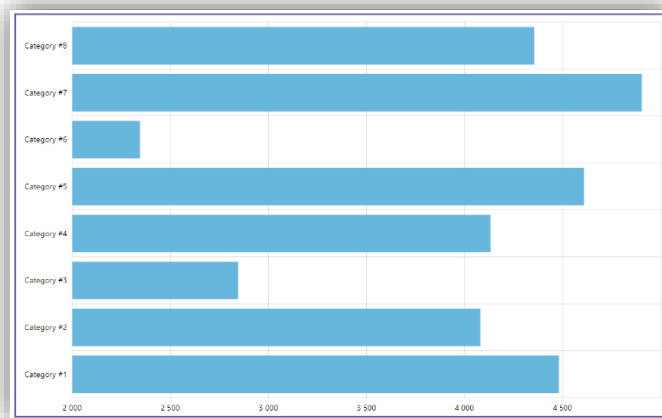


Рисунок 76. Горизонтальная столбчатая

е. Горизонтальная столбчатая с группировкой

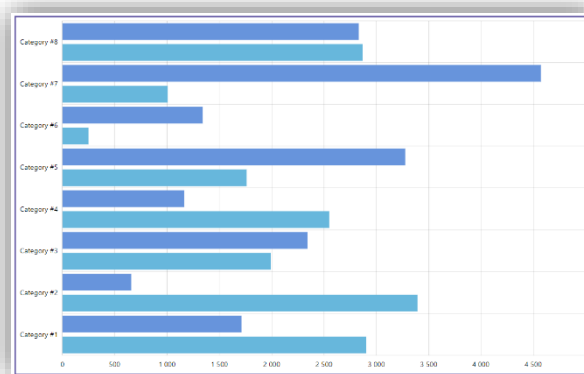


Рисунок 77. Горизонтальная столбчатая с группировкой

f. Горизонтальная столбчатая с накоплением

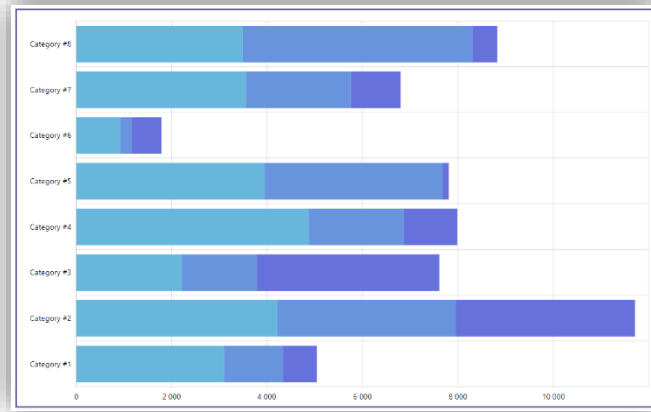


Рисунок 78. Горизонтальная столбчатая с накоплением

g. Комбинированная вертикальная столбчатая

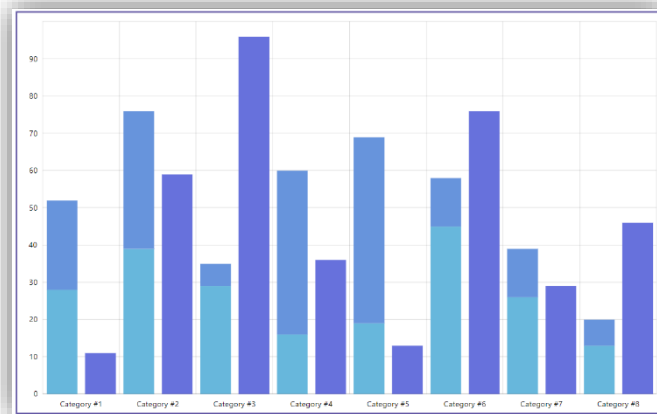


Рисунок 79. Комбинированная вертикальная столбчатая

h. Комбинированная горизонтальная столбчатая

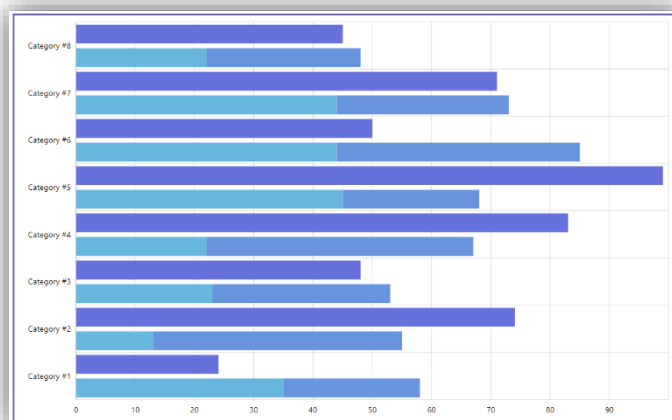


Рисунок 80. Комбинированная горизонтальная столбчатая

і. Нормированная вертикальная столбчатая с накоплением (100%)

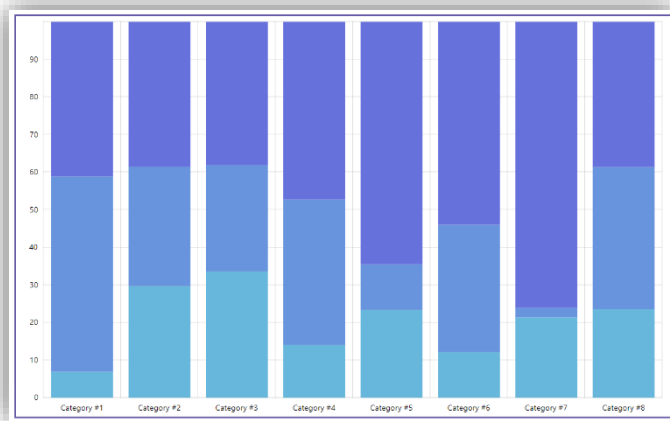


Рисунок 81. Нормированная вертикальная столбчатая с накоплением (100%)

ј. Нормированная горизонтальная столбчатая с накоплением (100%)

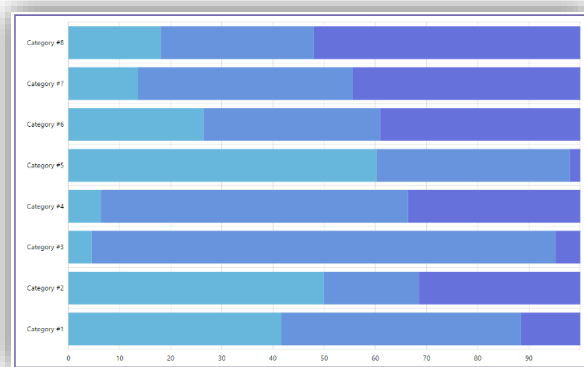


Рисунок 82. Нормированная горизонтальная столбчатая с накоплением (100%)

9) Карточка с числом - карточки с ключевыми значениями необходимо располагать в верхней части экрана – так они обращают на себя внимание в первую очередь. Оптимально располагать от 3 до 6 карточек.

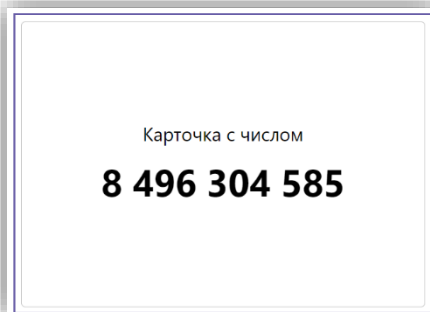


Рисунок 83. Карточка с числом

- 10) Карточка с текстом – как правило, используются для выделения заголовков.

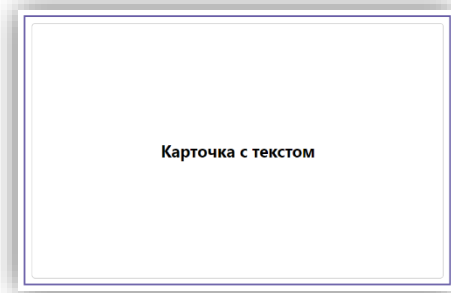


Рисунок 84. Карточка с текстом

- 11) Пузырьковая

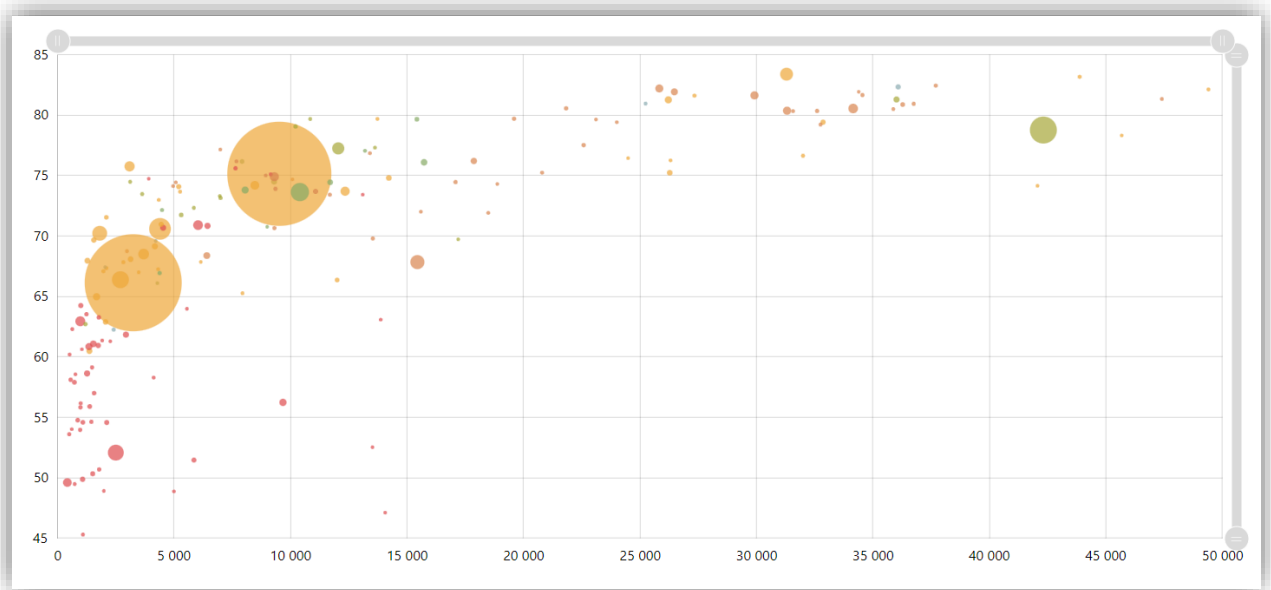


Рисунок 85. Пузырьковая диаграмма

12) Дерево



Рисунок 86. Диаграмма дерево

8. Конструктор

8.1 Начало работы в конструкторе после создания нового домена

Для начала работы в новом домене необходимо зарегистрировать в системе ETL-хранилище, в которое будут в дальнейшем выгружаться данные для трансформации и формирования моделей данных.



Рисунок 87. Процесс настройки хранилищ и сервисов для начал работы в Конструкторе

Для этого необходимо сначала в разделе «Модели данных» в подразделе «Источники данных» добавить IP-адрес ETL-хранилища, имя базы данных (можно создать новое, можно выбрать из существующих, если это не нарушает безопасность), выбрать коннектор `connector-psql`, имя схемы данных (указывается `public`), указать тип авторизации (нет авторизации, логин и пароль, ключ).

После этого, необходимо добавить аналитическое хранилище, указывая аналогичные настройки: IP-адрес, имя базы данных, выбрать коннектор `connector-druid`, имя схемы данных, указать тип авторизации (нет авторизации, логин и пароль, ключ). Однако надо учитывать, что у Apache Druid в качестве аналитического хранилища имя базы и имя схемы нужно указывать «druid».

Далее необходимо зайти в подраздел «ETL» и добавить новое ETL хранилище с точно такими же настройками, какие были указаны в подразделе «Источники данных», с тем же именем базы и схемы. После сохранения всех настроек необходимо нажать на кнопку «Регистрация хранилища» и посмотреть в консоли статус операции. Если операция успешна, то можно переходить к следующему шагу.

Для регистрации аналитического хранилища достаточно лишь в подразделе «ETL DWH» добавить новое хранилище с теми же настройками, который были указаны в подразделе «Источники данных». **Нажимать кнопку «Регистрировать хранилище» не нужно.**

Подраздел «Источники данных» необходим для того, чтобы работать непосредственно с данными. В данный раздел добавляются как внешние источники данных, так и внутренние хранилища. Инженер данных может просматривать схему источников, экспортировать данные в форматах csv, excel, json в данном подразделе.

В подразделах «ETL» и «ETL DWH» необходимо указывать доступы к хранилищам для того, чтобы сообщить системе, что именно из источников данных будет ETL-хранилищем, а что будет Аналитическим хранилищем. Таким образом система будет разделять внешние источники данных и внутренние хранилища.

В разделе «ETL» выбрать пункт «ETL-сервисы».

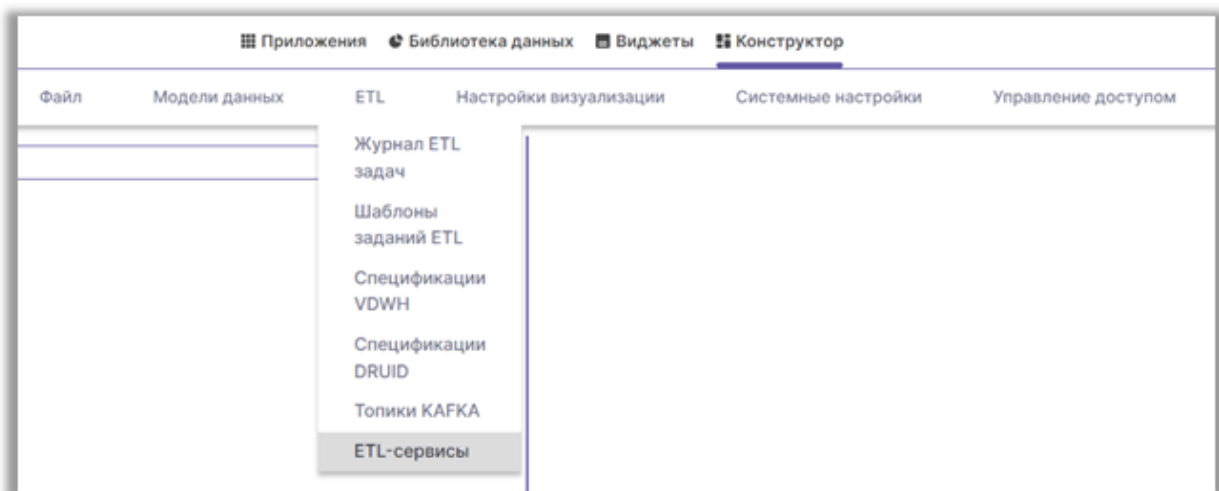


Рисунок 88. Раздел ETL-сервисов

Для добавления нового сервиса необходимо нажать кнопку «Новый», в открывшемся окне заполнить обязательные поля:

- ввести наименование
- выбрать нужный сервис из списка предустановленных коннекторов/сервисов: «kafka-broker» - позволяет забрать данные из источника и передать данные во внутреннюю шину (Kafka), «broker-

druid» - позволяет записать данные в аналитическое хранилище (Druid)

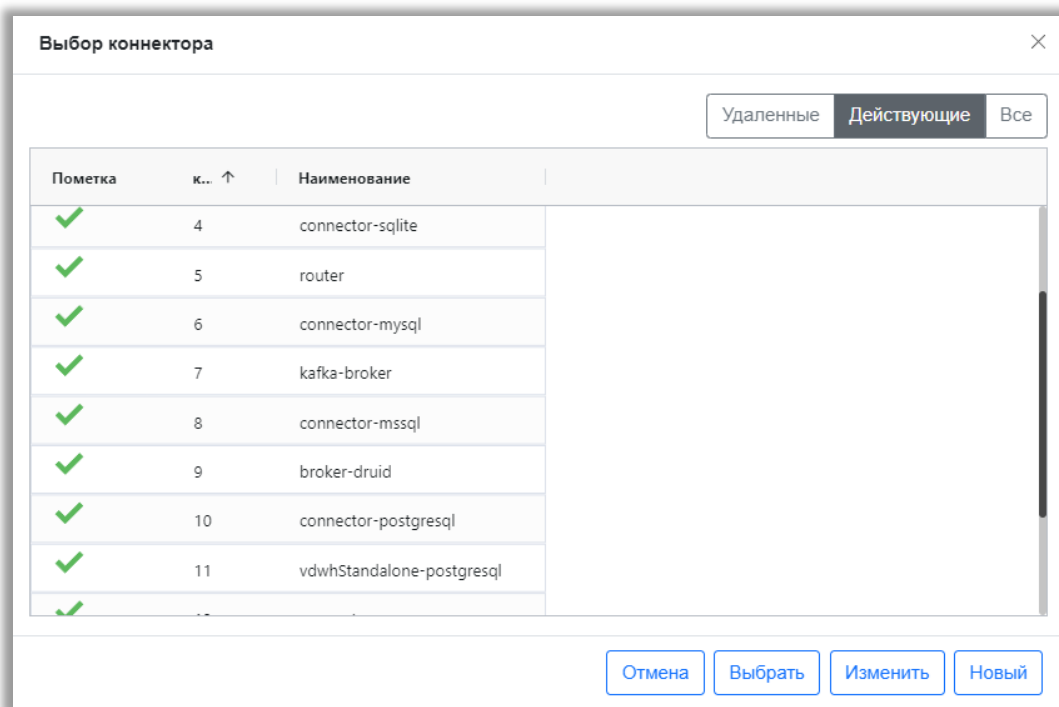


Рисунок 89. Выбор коннекторов/сервисов

- ввести http-адрес: указать сервер и порт, на котором запущены Kafka и Druid
- в поле «Вид авторизации» выбрать «Нет авторизации», затем нажать кнопку «Сохранить»

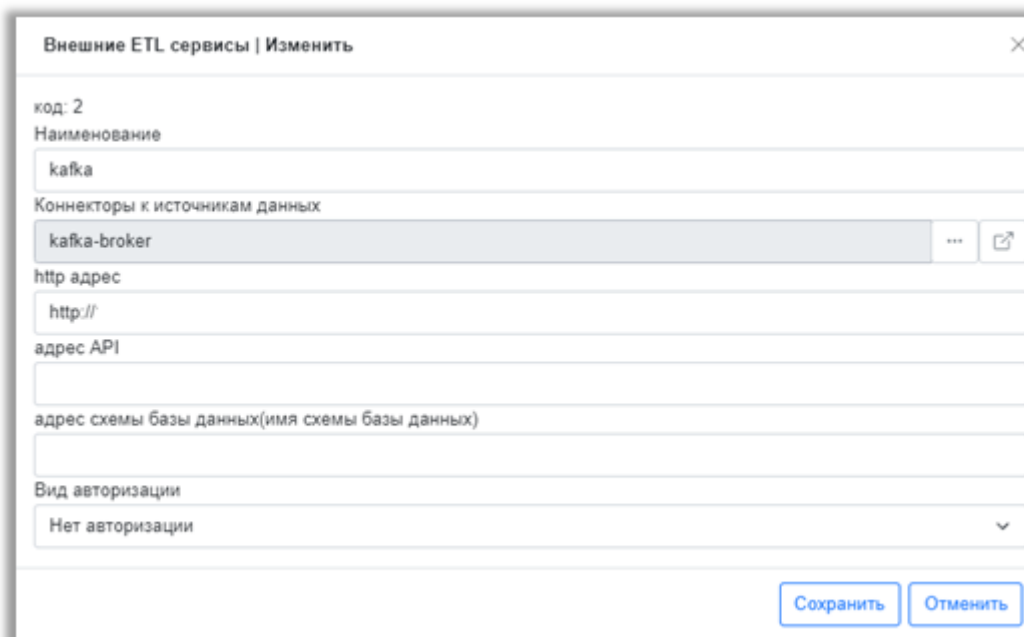


Рисунок 90. Добавление сервиса для передачи данных в шину kafka

Внешние ETL сервисы | Изменить

код: 3
 Наименование

Коннекторы к источникам данных

http адрес

адрес API

адрес схемы базы данных(имя схемы базы данных)

Вид авторизации

Рисунок 91. Добавление сервиса для отправки данных в Аналитическое хранилище Druid

Для редактирования созданной записи необходимо выбрать нужный элемент и нажать кнопку «Изменить», для удаления – нажать кнопку «Пометить на удаление».

Внешние ETL сервисы

Пометка	код	Наименование	http адрес	Коннектор
✓	3	DRUID		broker-druid
✓	4	kafka		kafka-broker

Внешние ETL сервисы

Пометка	код	Наименование	http адрес	Коннектор
✓	3	DRUID	http://	broker-druid
✓	4	kafka	http://	kafka-broker

Рисунок 92. Изменение и пометка на удаление настроенных сервисов

Помеченный на удаление элемент экранной формы будет перемещен на вкладку «Удаленные», при необходимости удаленный элемент можно восстановить, для этого потребуется выбрать нужный элемент и нажать кнопку «Снять пометку на удаление», после чего элемент переместится на вкладку «Действующие».

Внешние ETL сервисы

Новый Изменить Снять пометку на удаление Копировать

Удаленные Действующие Все

Пометка	код	Наименование	http адрес	Коннектор
✗	3	DRUID	http://	broker-druid
✗	4	kafka	http://	kafka-broker

Рисунок 93. Удаленные сервисы

8.2 ETL-процесс

Чтобы доставить данные до аналитического хранилища необходимо выполнить следующие действия:

- Извлечение данных из внешних источников в ETL-хранилище
 - Получить схему данных внешнего источника
 - Создать спецификации VДWH на таблицы, которые необходимо извлечь. В рамках создания данной обеспечивается возможность изменить исходный тип данных
 - Создать шаблон задачи на извлечение таблицы в ETL-хранилище. Для каждой таблицы требуется создать свой шаблон задачи
 - Создать ETL-задачу на извлечение таблицы. Для каждой таблицы необходимо создавать отдельную ETL-задачу. В случае если после запуска задача успешно выполнена, то извлекаемая таблица появится в ETL-хранилище
- Трансформация данных (если требуется)

- Создать необходимые представления в конструкторе запросов на основе извлеченных таблиц
- Отправка данных в шину данных (Apache Kafka)
 - Описать топик Kafka, в котором указывается как будет называться топик и какую таблицу/представление Дата-инженер передает в шину данных для дальнейшей доставки в аналитическое хранилище
 - Создать шаблон задачи на отправку таблицы/представления в шину
 - Создать ETL-задачу на отправку таблицы в шину. Запуск задачи
- Отправка данных в Аналитическое хранилище (Apache Druid)
 - Описать спецификацию таблицы для Druid, то есть указать измерения, меры, timestamp и т.д.
 - Создать шаблон задачи на отправку таблицы в Druid
 - Создать ETL-задачу на отправку таблицы в Druid. В случае если после запуска задача успешно выполнена, то таблица появится в Аналитическом хранилище

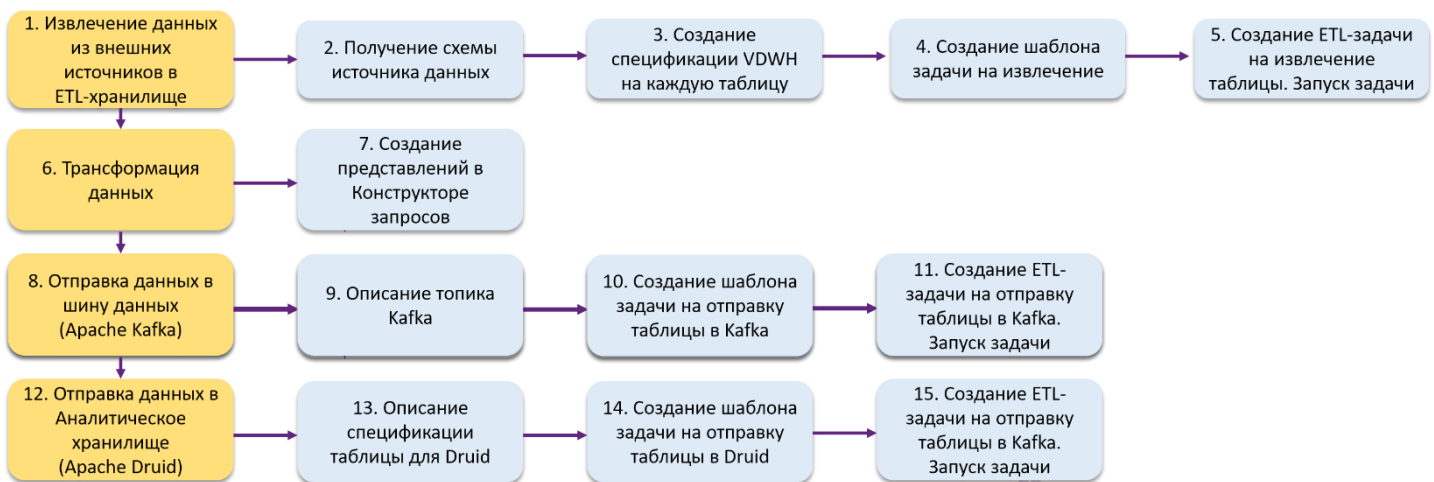


Рисунок 94. ETL-процесс

В системе используются спецификации для описания настроек извлекаемых таблиц.

В системе есть несколько спецификаций:

- Спецификация DWH. Используется для извлечения данных в ETL хранилище. В подразделе «Спецификация DWH» создается новая спецификация, в которой указывается конкретная таблица из внешнего источника для извлечения и просматриваются/корректируются типы данных
- Топики Kafka. Используется для отправки данных в шину данных (kafka). В спецификации указывается только наименование указываемой таблицы
- Спецификация Druid. Используется для отправки данных в аналитическое хранилище. В спецификации указывается таблица, время гранулирования данных, timestamp, измерения, метрики и типы данных

Спецификации используются в шаблонах задач. Этот процесс необходим для того, чтобы сделать шаблон любой задача: на извлечение, на отправку в кафку, на отправку в druid. Для каждой задачи ETL нужно сначала сделать шаблон, в котором указывается таблица, для которой необходимо выполнить операцию, и параметры передачи данных (размер пакета, количество попыток и тип).

После создания шаблонов задач в подразделе «Задачи ETL» создается непосредственно задача с ссылкой на шаблон, которой можно назначить расписание.

8.2.1 Подключение к внешнему источнику данных, получение схемы данных

Для подключения нового источника необходимо в разделе «Модели данных» в подразделе «Источники данных» добавить IP-адрес внешнего источника, имя базы данных, имя схемы данных, указать тип авторизации (нет авторизации, логин и пароль, ключ).

После добавления нового источника необходимо обновить его схему данных, чтобы увидеть содержимое.

Для просмотра таблиц, содержащихся в источнике, необходимо его выбрать и нажать «Просмотр объектов источника». По каждой таблице можно посмотреть состав полей, для этого необходимо выбрать таблицу и нажать на кнопку «Поля источника данных».

В системе есть возможность экспортировать таблицы в форматах: excel, csv, json. Чтобы выгрузить тестовое количество строк необходимо нажать «Экспорт данных» и выбрать формат.

Если необходимо просто посмотреть таблицу, не выгружая ее, то можно нажать «Просмотр источника».

Когда пользователь подключается к источнику, он работает с его схемой. Данные в этот момент не грузятся в ETL хранилище. Для этого нужно поставить задачу на извлечение данных.

8.2.2. Коннекторы и сервисы

В системе используются следующие коннекторы для подключения к различным сервисам и источникам:

- postgresql
- MYSQL
- MSSQL
- H2
- rest Api
- gml
- odata
- couchdb
- mongodb
- apache druid
- csv
- excel
- json

Сервисы, используемые в системе для отправки и получения данных:

- Сервис отправки данных в PostgreSQL vDWH – необходим для отправки данных в ETL хранилище на этапе извлечения данных из внешнего источника PostgreSQL
- Сервис отправки данных в H2 vDWH
- Сервис отправки данных в Kafka – используется для отправки данных из ETL хранилища в шину данных (kafka) для дальнейшей передачи данных в аналитическое хранилище (druid)
- Сервис публикации данных в Druid по HTTP
- Сервис приемки данных из Kafka
- Сервис публикации данных в Druid из Kafka – используется для передачи данных из шины данных (Kafka) в аналитическое хранилище (Druid)

8.2.3 Извлечение таблицы данных в ETL хранилище: создание спецификации, шаблона и ETL-задачи

Перед созданием задачи на извлечение необходимо добавить новую спецификацию DWH в разделе «ETL» подразделе «Спецификации vDWH».

- заполнить все поля
- выбрать ETL хранилище (куда нужно извлечь таблицу)
- выбрать таблицу для извлечения
- после выбора нужной таблицы автоматически заполнится табличная часть формы, для настройки маппинга (соответствия) типов полей между внешним источником и ETL хранилищем
- для того, чтобы вместе с данными извлеклись и их типы, то необходимо обязательно установить признак «Использовать карту типов»

Спецификации vDWH | Новый

код:
Наименование

ETL хранилище

Таблица данных
flights

Использовать карту типов при создании таблицы

Настройки колонок мапинга полей

Добавить строку

Имя поля	Тип в источнике	Тип в vDWH
flight_id	serial	serial
flight_no	bpchar	bpchar
scheduled_departure	timestamptz	timestamptz
scheduled_arrival	timestamptz	timestamptz

Записать Отменить

Рисунок 95. Форма спецификации vDWH

Далее необходимо создать шаблон на извлечение данных (раздел «ETL» подраздел «Шаблоны заданий ETL» кнопка «Новый»):

- Необходимо заполнить все поля, кроме Спецификация kafka и спецификация druid.
- В поле «Вид ETL» выбираем сервис отправки данных в ETL-хранилище (PostgreSQL vDWH)

Выбор типа ETL операции

Удаленные Действующие Все

Пометка	Порядок	Наименование
✓	1	Сервис отправки данных в PostgreSQL vDWH
✓	2	Сервис отправки данных в H2 vDWH
✓	3	Сервис отправки данных в Kafka

Рисунок 96. Выбор сервиса отправки данных в ETL-хранилище

Заполнение табличной части «Основные свойства»:

- Источник данных – внешний источник, откуда извлекаем данные

- Таблица данных – выбираем нужную нам таблицу (должна совпадать с таблицей, описанной в спецификациях vdwh)
- ETL хранилище – куда хотим положить данные
- Kafka и druid оставляем пустыми
- Настройка спецификации VDWH – выбираем созданную ранее спецификацию

Свойства шаблона

Основные свойства

Источник данных

Таблица данных

ETL хранилище

Настройка топика KAFKA

Настройка спецификации DRUID

Настройка спецификации VDWH

Рисунок 97. Свойства Шаблона задачи

Заполнение табличной части «Дополнительные свойства»:

- Имя создаваемой таблицы – новое имя таблицы, не должно совпадать с существующим, так как может перезаписать данные
- Кол-во строк (0-все) – 0
- Размер пакета – рекомендованный размер 50 000 записей
- Пропустить от начала – 0
- Тип
- Количество попыток
- Интервал между попытками в секундах

Дополнительные свойства
Имя создаваемой таблицы <input type="text"/>
Кол-во строк (0-все) <input type="text"/>
Размер пакета <input type="text"/>
Пропустить от начала <input type="text"/>
<input type="checkbox"/> Добавлять к имеющимся записям
Кол-во попыток при ошибке <input type="text"/>
Интервал между попытками в секундах <input type="text"/>

Рисунок 98. Дополнительные свойства в шаблоне задачи

Далее необходимо создать саму задачу в разделе «ETL» подразделе «Журнал заданий ETL»:

- нажать на кнопку «Новый»
- выбрать только что созданный шаблон
- заполнить наименование задачи
- В списке задач выбрать нужную задачу и нажать кнопку «Запустить», затем «Обновить данные по задаче» и кнопку «Обновить»

На данном этапе должен завершиться процесс извлечения данных. Увидеть статус задачи можно в таблице Журнала задач. Статус может быть «Завершена», «Есть ошибки». Статус ошибки можно посмотреть, если перейти в саму задачу.

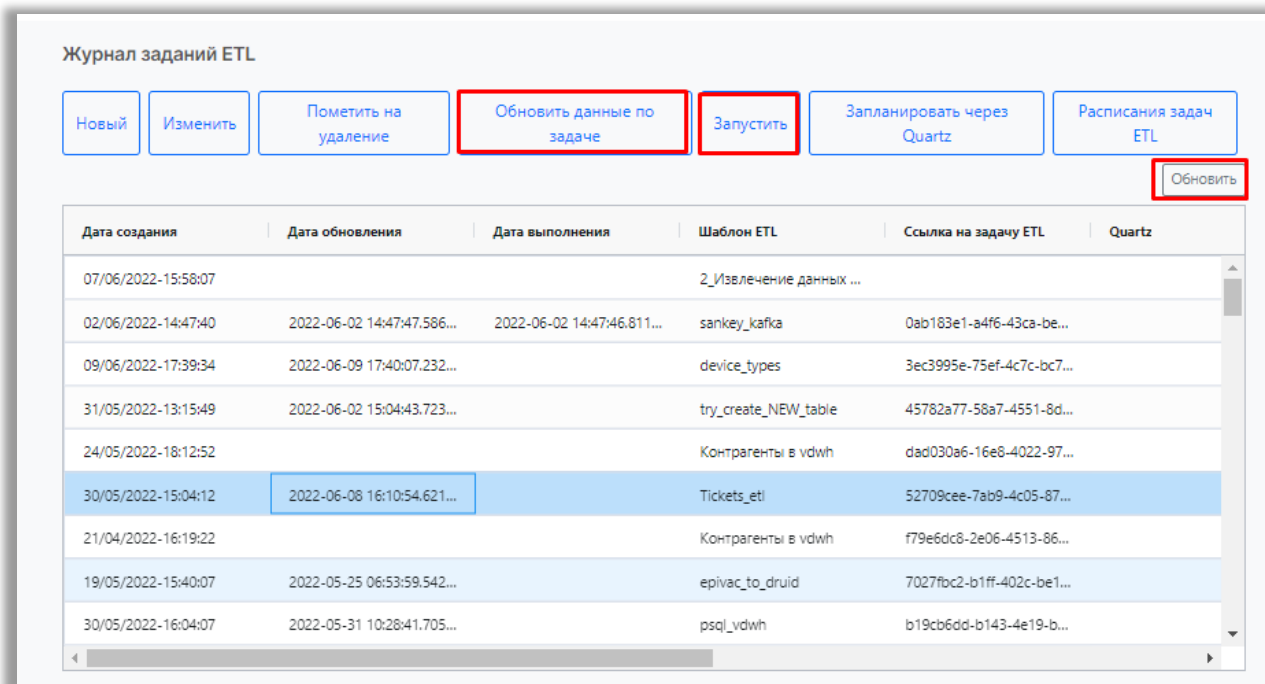


Рисунок 99. Журнал ETL-задач

8.2.4 Трансформация (создание представлений, объединение, группировка, сортировка, условия)

Цель данного этапа ETL-процесса - подготовка данных к размещению в аналитическое хранилище данных и приведение их к виду более удобному для последующего анализа. В процессе преобразования данных будет использоваться инструмент «Конструктор запросов», который способен реализовать различные методы обработки и очистки данных.

Одним из методов обработки данных является объединение двух или нескольких таблиц в одну при помощи SQL запроса. Результатом такого объединения будет создание представления. Представление – это виртуальная таблица, содержимое которой определяется запросом. В отличие от обычных таблиц реляционных баз данных, представление не является самостоятельной частью набора данных, хранящегося в базе. Содержимое представления динамически вычисляется на основании данных, находящихся в реальных таблицах. Изменение данных в реальной таблице базы данных немедленно отражается в содержимом всех представлений, построенных на основании этой таблицы.

Самым простым способом создания объединения является использование оператора JOIN.

В инструменте «Конструктор запросов» используются следующие модификаторы JOIN:

- INNER JOIN – внутреннее объединение таблиц.
- LEFT JOIN – левое внешнее объединение таблиц.
- RIGHT JOIN – правое внешнее объединение таблиц.
- CROSS JOIN – перекрестное объединение таблиц.

Конструктор запросов позволяет составить текст запроса на языке запросов исключительно визуальными средствами.

Пользователям, не обладающим достаточным опытом написания SQL-запросов, приходится запоминать многочисленные операторы и области их применения.

Эффективно решает эту проблему конструктор запросов. Этот инструмент облегчает работу с запросами и освобождает пользователя от дополнительной нагрузки.

Конструктор запроса имеет достаточно простой, интуитивно понятный интерфейс.

Результатом работы конструктора будет являться синтаксически правильный текст запроса. Таким образом, пользователь может составить работоспособный запрос, даже не владея синтаксисом языка запросов - необходимые синтаксические конструкции конструктор сгенерирует автоматически. Готовый текст запроса может быть сразу же вставлен в рабочую область модели представления или скопирован в буфер обмена.

Для работы с конструктором запросов необходимо в разделе «Модели данных» выбрать пункт «Модели представлений».



Рисунок 100. Журнал ETL-задач

Для добавления новой модели необходимо нажать кнопку «Новый», в открывшемся окне заполнить обязательные поля:

- ввести наименование;
- выбрать ETL хранилище;
- текст SQL запроса будет вставлен автоматически после использования инструмента «Конструктор запросов»;
- нажать кнопку «Query Builder» для непосредственной работы с инструментом.

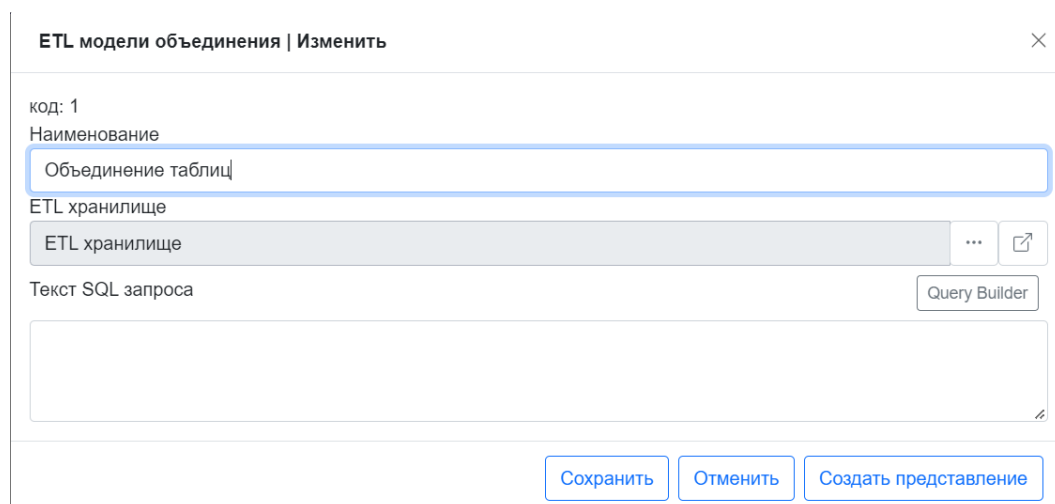


Рисунок 101. Вызов конструктора запросов

После вызова конструктора запросов открывается форма данного инструмента. Она поделена на несколько закладок, каждая из которых имеет индивидуальное назначение и функциональные возможности:

- «Таблицы и поля». Это стартовая вкладка, где отражена вся информация о задействованных полях и таблицах в запросе. В левой

части находится дерево источников, в центре расположены выбранные таблицы, а в правой – конечные поля. В дереве источников необходимо выбрать с помощью кнопок > или >> таблицы, которые будут участвовать в объединении, затем с помощью кнопок > или >> выбрать поля, которые попадут в представление.

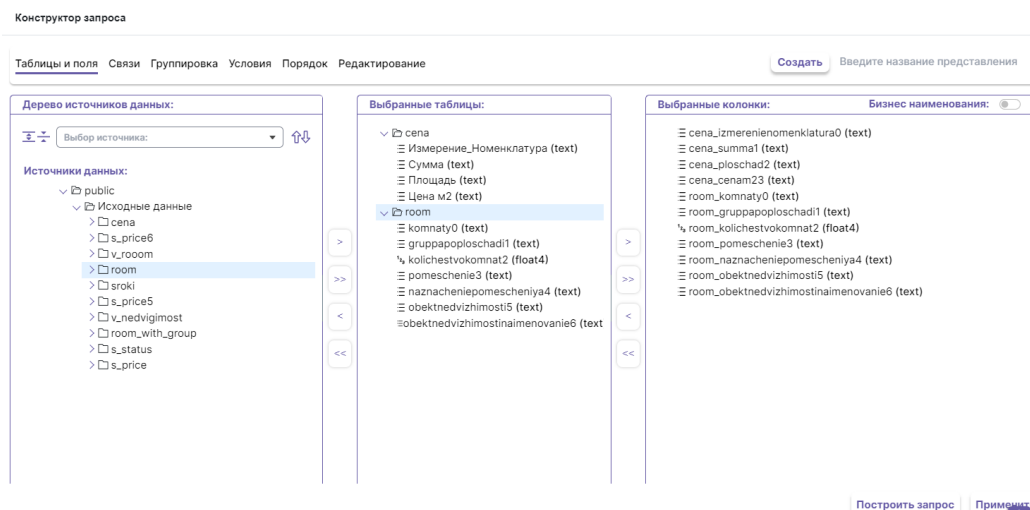


Рисунок 102. Выбор таблиц и полей

в. «Связи». В этом разделе содержится информация о выбранных таблицах, а также поля, по которым они соединены. После нажатия на кнопку «+», в появившейся строке необходимо выбрать объединяемые таблицы и указать тип их соединения (left join, right join и т.д.), затем выбрать поля, по которым будут соединены таблицы.

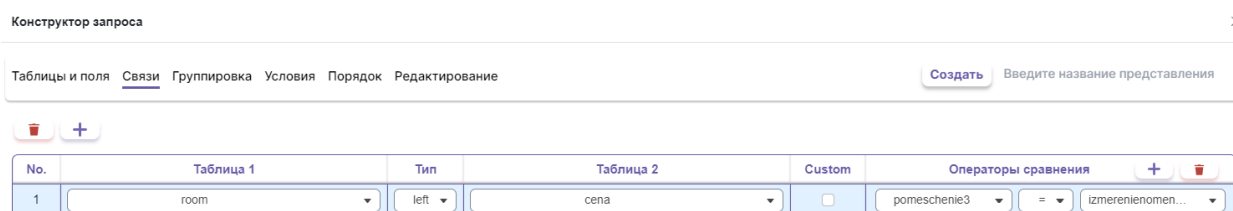


Рисунок 103. Раздел для объединения таблиц

При необходимости можно добавить дополнительные условия на вкладках «Группировка» и «Условия».

с. «Группировка». В этом разделе строки группируются по определенным полям. Для того, чтобы сгруппировать поля с

текстовыми значениями необходимо переместить нужные поля с помощью кнопок > или >> в область «Сгруппированные колонки». Для агрегирования числовых значений необходимо переместить нужные колонки в область «Агрегированные колонки» или в области «Сгруппированные колонки» правой кнопкой мыши вызвать контекстное меню, в котором выбрать требуемую функцию.

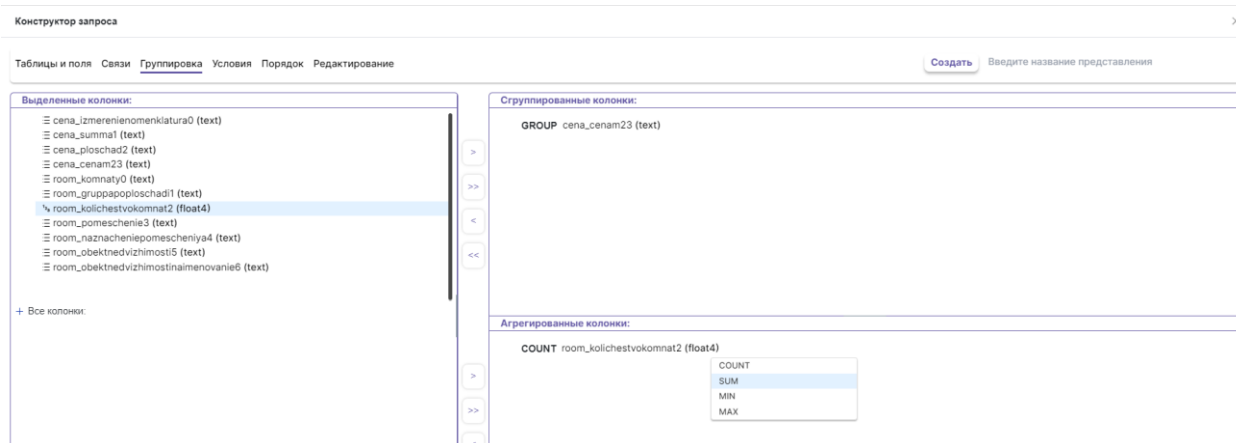


Рисунок 104. Настройка группировки

д. «Условия». В этом разделе выставляются дополнительные условия на данные, которые оказываются в результате выборки.

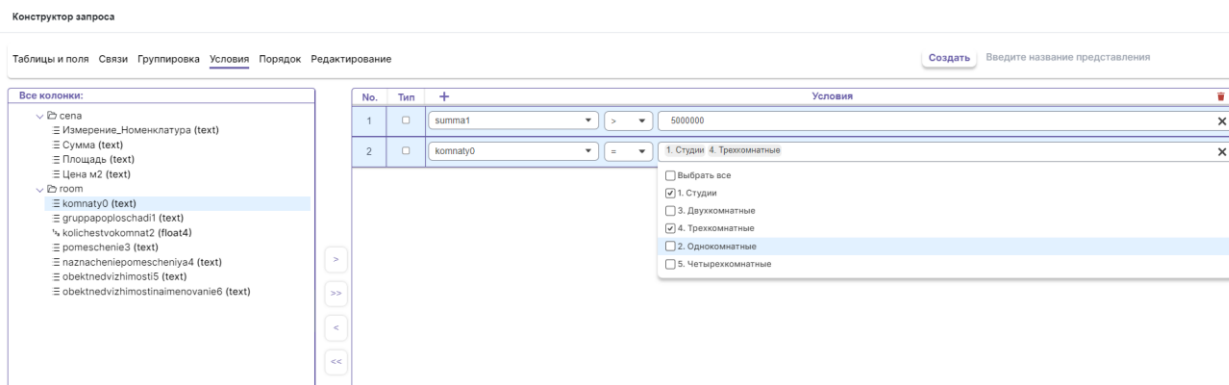
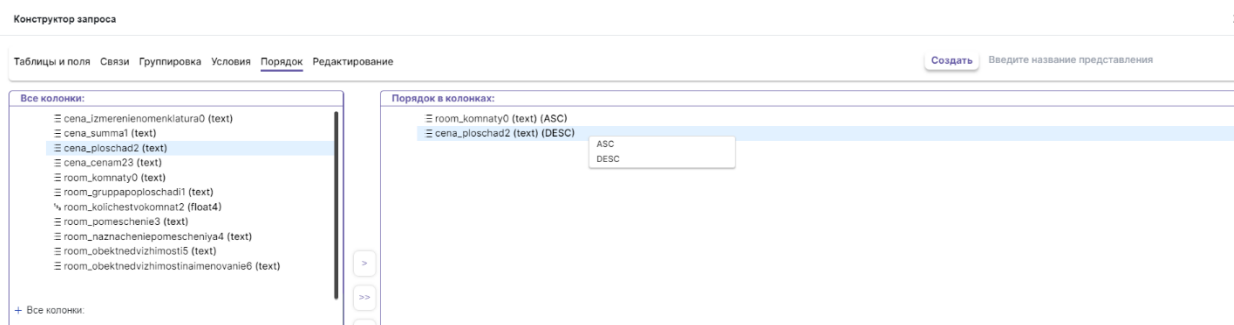


Рисунок 105. Настройка условий

е. «Порядок». В этом разделе сортируются итоговые записи. По умолчанию выбранные колонки сортируются по возрастанию -



«ASC», при необходимости можно поменять порядок с помощью вызова контекстного меню. Команда «Desc» отсортирует значения по убыванию.

Рисунок 106. Настройка порядка

- f. «Редактирование». В данном разделе можно просмотреть итоговый запрос в формате SQL и JSON, а также просмотреть результат запроса (кнопка «View»). После ввода всех необходимых условий можно просмотреть сформированный запрос, нажав кнопку «Построить запрос».

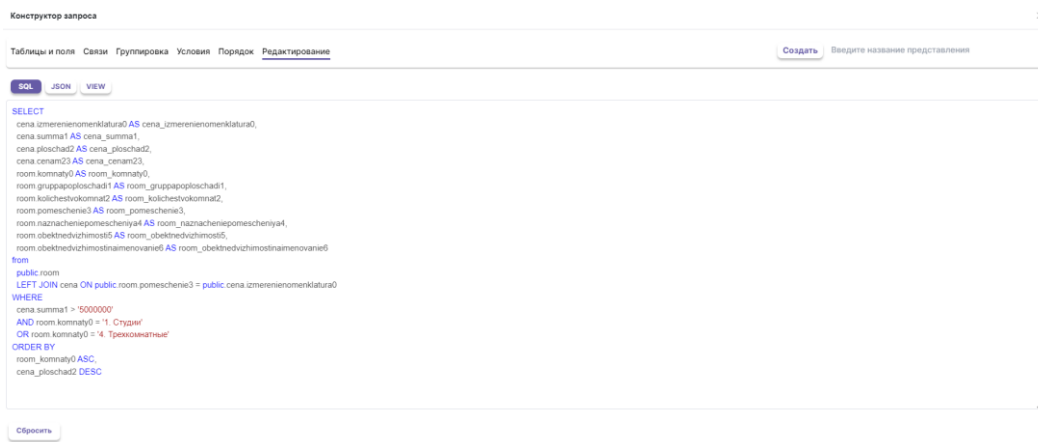


Рисунок 107. Возможность редактирования финального запроса

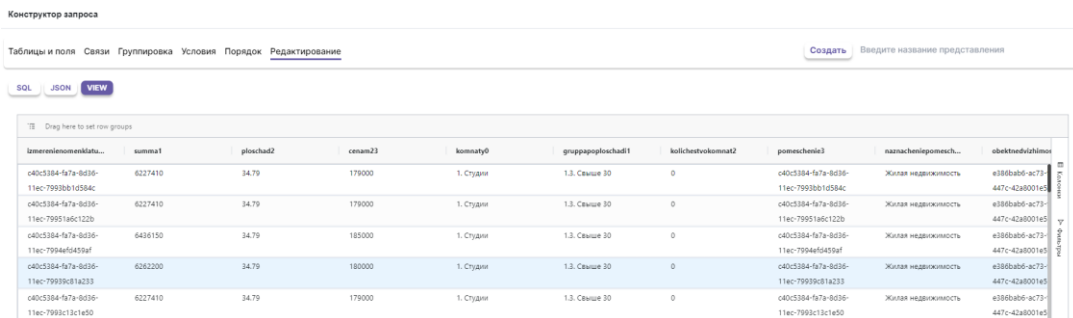


Рисунок 108. Просмотр таблицы по результатам запроса

При необходимости можно корректировать текст запроса, установив курсор внутри существующего текста запроса.

Далее необходимо задать имя представления в верхней правой части окна, предварительно нажав кнопку «Создать».

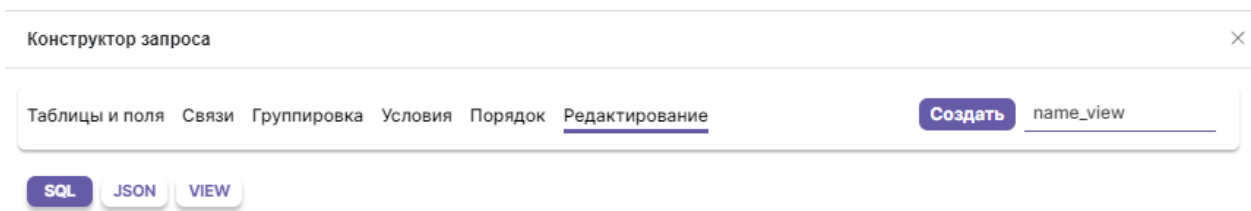


Рисунок 109. Добавление наименования для представления

Затем нажать кнопку «Применить», после чего «Конструктор запросов» закроется, а поле «Текст SQL запроса», в создаваемой модели объединения, заполнится автоматически сгенерированным SQL-запросом. Далее нажать кнопку «Сохранить» и «Создать представление».

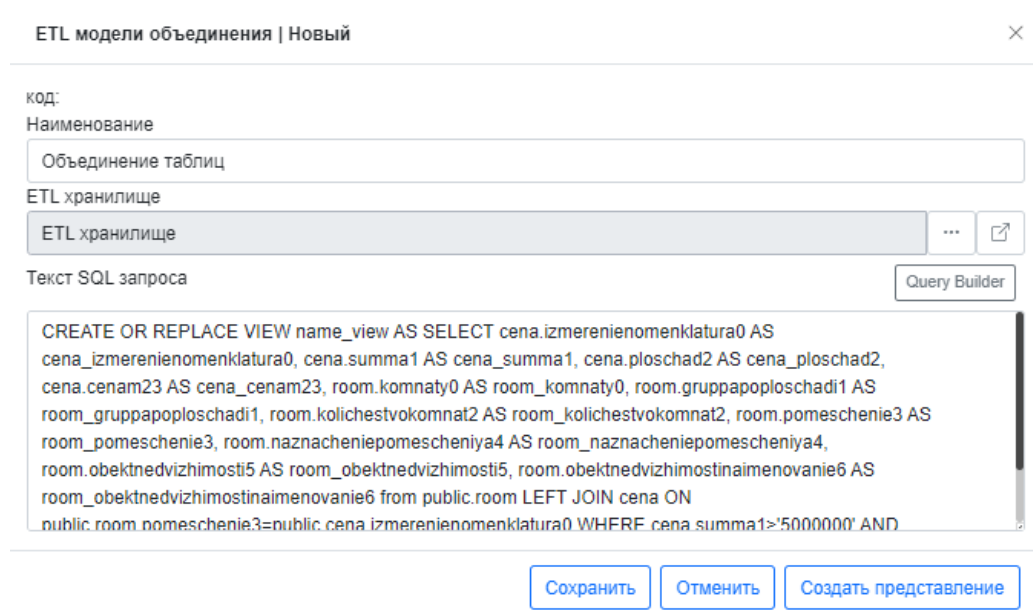


Рисунок 110. Отображение финального текста запроса

8.2.5 Отправка данных в шину данных: создание топика, шаблона и ETL-задачи

Загрузка данных в аналитическое хранилище является финальным этапом ETL-процесса. Процесс загрузки заключается в переносе готовых таблиц/представлений (преобразованных и извлеченных данных) через шину данных из внутреннего ETL-хранилища в Аналитическое хранилище для последующего анализа или визуализации.

Задачу по отправке данных в Kafka необходимо начать с описания спецификаций топика Kafka. Спецификация шины данных Kafka определяет правила взаимодействия для получения или отправки данных через Apache.Kafka.

В разделе «ETL» необходимо выбрать пункт «Топики Kafka».

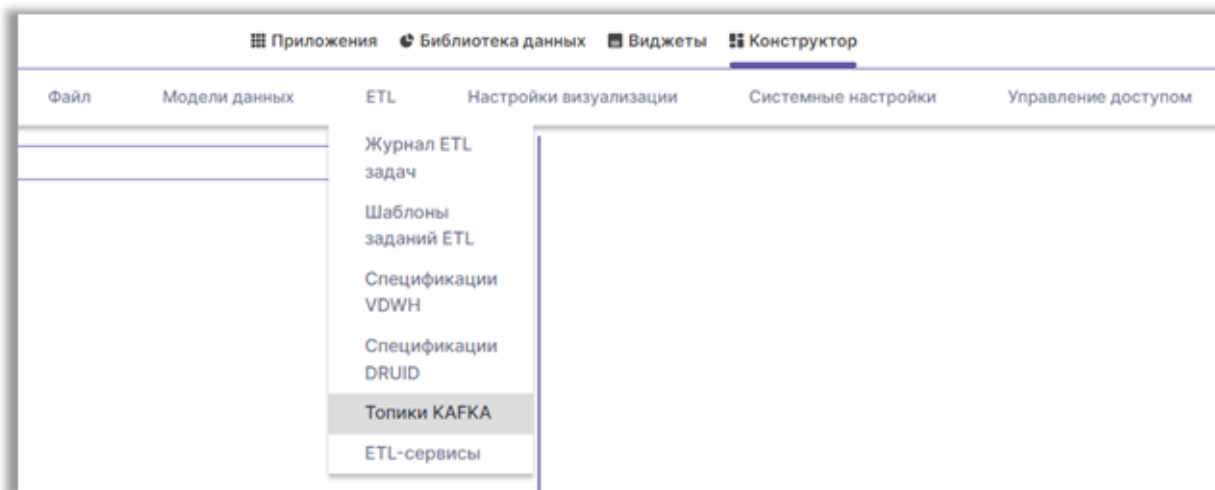


Рисунок 111. Раздел для заполнения топиков Kafka

Для добавления новых спецификаций необходимо нажать кнопку «Новый», в открывшемся окне заполнить обязательные поля:

- ввести имя топика Kafka;
- выбрать ETL-сервис;
- выбрать из предлагаемого списка режим получения данных:
 - «Получение данных из таблицы» - для данного режима необходимо в поле «Таблица данных» выбрать нужную таблицу из списка объектов источника данных;
 - «Получение данных из запроса» - для данного режима потребуется с помощью инструмента «Query Builder» или самостоятельно сформировать текст SQL-запроса;

- «Получение данных от внешнего провайдера» - в заданном режиме данные в шину попадают напрямую из внешнего источника.

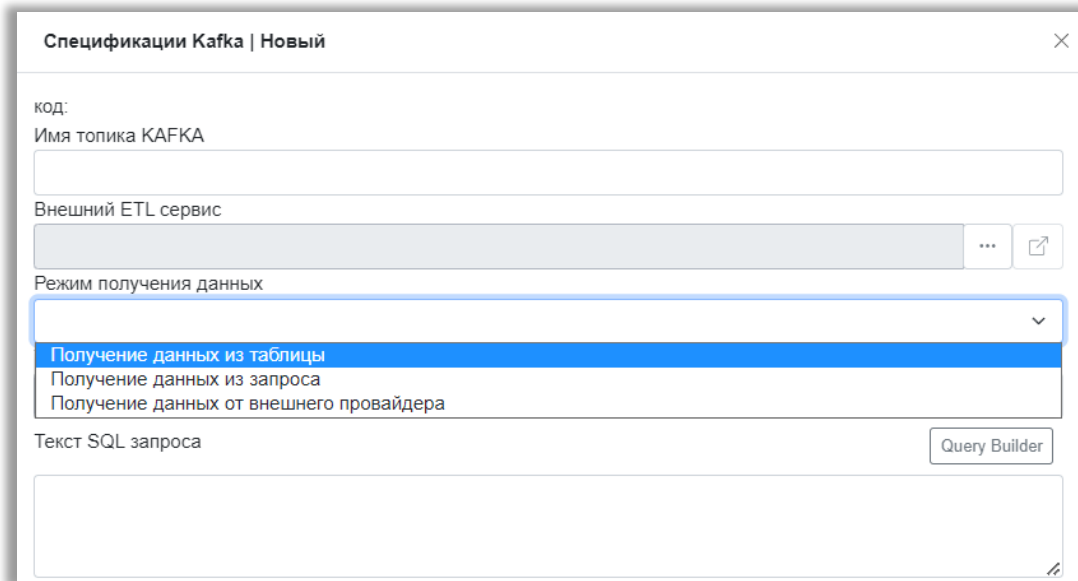


Рисунок 111. Заполнение топика

При пустом поле «Режим получения данных» по умолчанию применяется режим «Получение данных из таблицы».

После заполнения всех обязательных полей нажать кнопку «Сохранить».

Следующим шагом выполнения задачи по отправке данных в Kafka является создание шаблона заданий ETL.

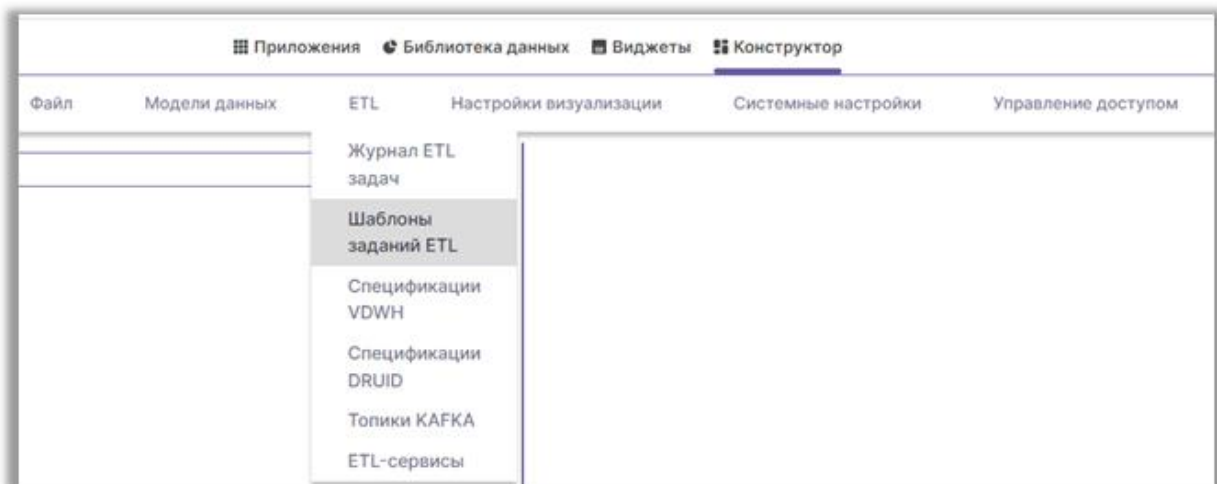


Рисунок 112. Раздел для создание шаблонов задач

В разделе «ETL» необходимо выбрать пункт «Шаблоны заданий ETL».

Для добавления нового шаблона необходимо нажать кнопку «Новый», в открывшемся окне заполнить обязательные поля:

- ввести наименование;
- выбрать вид ETL «Сервис отправки данных в Kafka»;

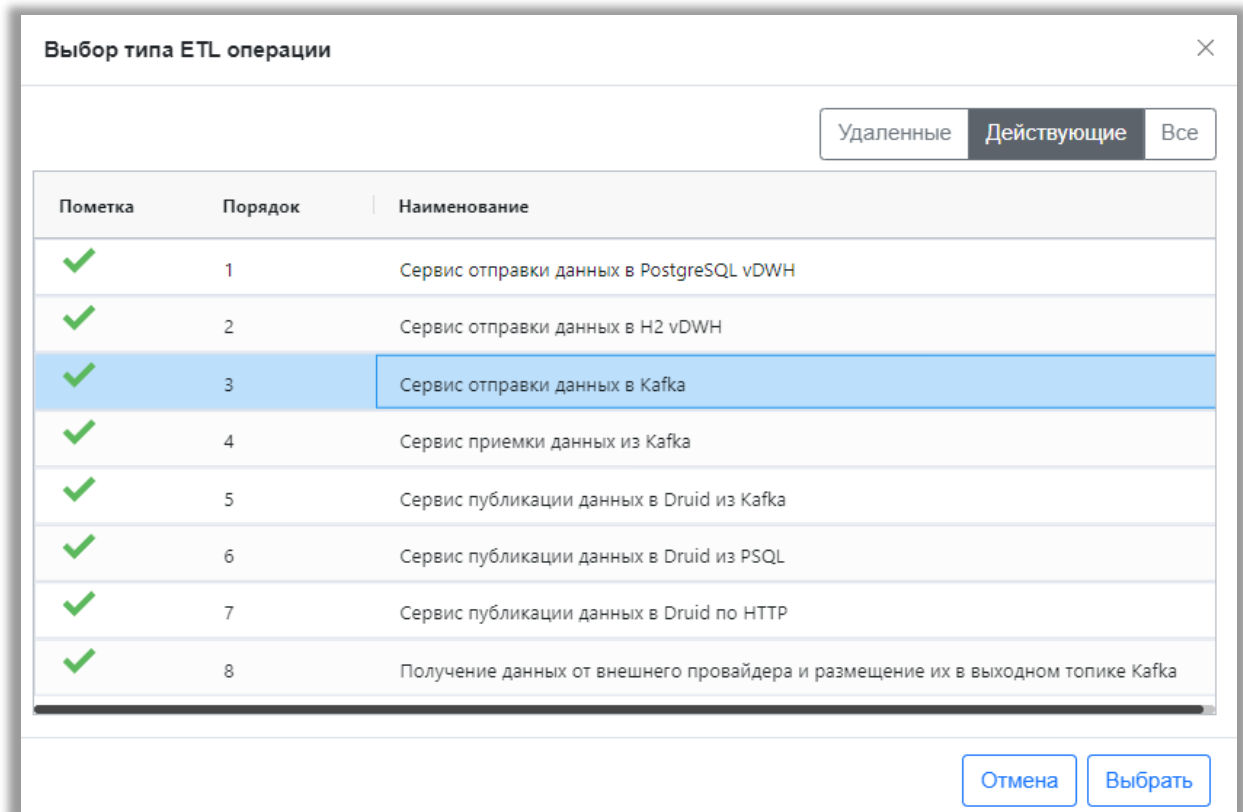


Рисунок 113. Выбор сервиса отправки данных для шаблона задачи

- выбрать источник данных, из которого планируется забирать данные;
- в поле «Таблица данных» выбрать таблицу из списка объектов источника данных;
- выбрать топик Kafka;
- установить количество загружаемых строк, при установленном параметре «0» будут загружены все строки;
- установить размер передаваемого пакета, рекомендованный размер не более 200 тыс. строк;
- заполнить поле «Пропустить от начала», если указать значение «0», то будут загружены строки начиная с первой, если необходимо

загружать данные с определенной строки, указывается соответствующий номер строки;

- указать количество попыток при ошибке, если оставить поле пустым, то по умолчанию будет проставлено «3» попытки;
- указать интервал между попытками, если оставить поле пустым, то по умолчанию будет проставлено «60» секунд, затем нажать кнопку «Сохранить».

Следующим шагом выполнения задачи по отправке данных в Kafka является создание журнала заданий ETL на основании созданного шаблона ETL.

В разделе «ETL» необходимо выбрать пункт «Журнал заданий ETL».

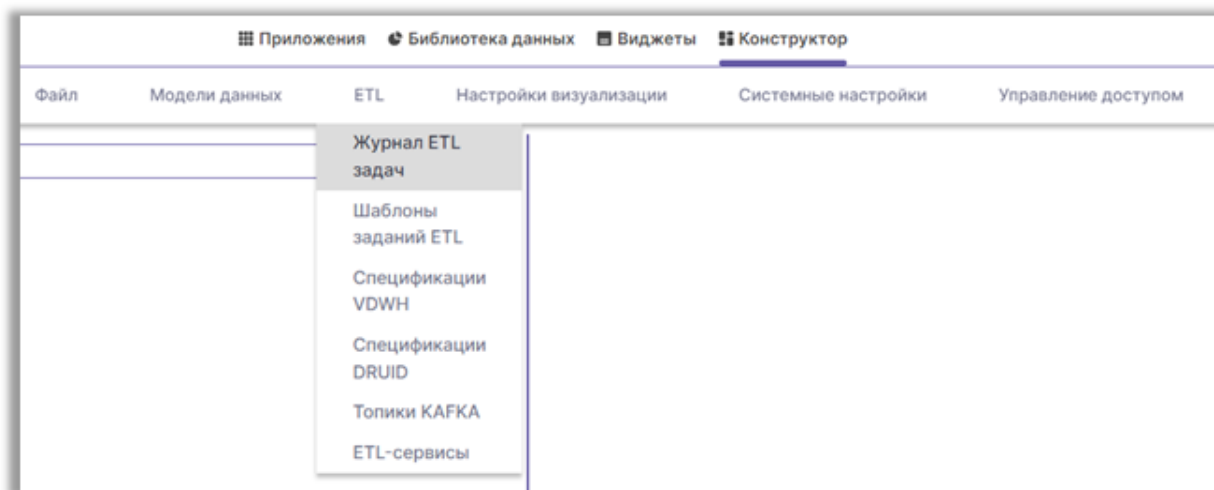


Рисунок 114. Раздел для создания ETL-задач

Для добавления новой задачи в журнал необходимо нажать кнопку «Новый», в открывшемся окне заполнить обязательные поля:

- ввести наименование;
- выбрать шаблон задачи ETL, затем нажать кнопку «Сохранить».

Задача может быть запущена немедленно или по расписанию, для немедленного запуска задачи на отправку данных в Kafka необходимо в списке задач выделить нужную и нажать кнопку «Запустить». Для получения статуса о ходе выполнения задачи необходимо нажать на кнопку «Обновить данные по задаче».

На данном этапе должен завершиться процесс передачи данных в Kafka.

8.2.6 Отправка данных в аналитическое хранилище: создание топики, шаблона и ETL-задачи

Следующим шагом является доставка данных из шины в аналитическое хранилище.

Задачу на доставку данных в аналитическое хранилище необходимо начать с описания спецификаций Druid. Спецификация Druid определяет правила создания источника данных в базе данных Apache Druid (измерения, метрики, расписания, гранулирование и т.д.).

В разделе «ETL» необходимо выбрать пункт «Спецификации Druid».

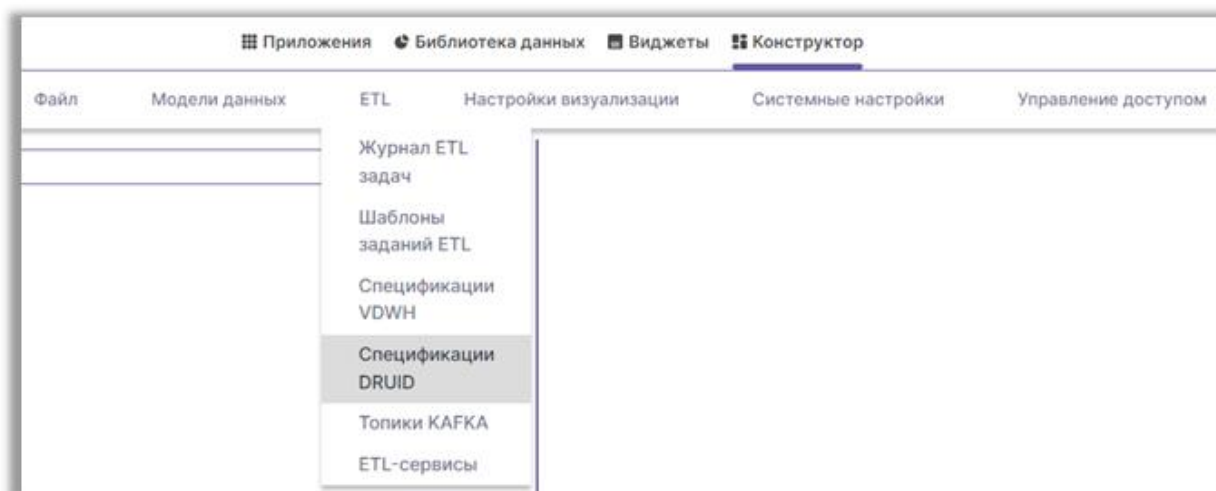


Рисунок 115. Раздел для заполнения спецификации Druid

Для добавления новых спецификаций необходимо нажать кнопку «Новый», в открывшемся окне заполнить обязательные поля:

- ввести наименование спецификации;
- выбрать ETL-сервис;
- в поле «Таблица данных» выбрать таблицу из списка объектов источника данных, после заполнения данного поля область «Настройка источника данных» заполняется автоматически служебными данными таблицы (поля данных и исходные типы данных), также для всех полей устанавливается тип данных, который будет передаваться в Druid, по умолчанию это строковый тип данных;

Настройка источника данных

Ручная настройка timestamp

Настройка timestamp

Имя поля timestamp Формат timestamp Значение по умолчанию

Настройки измерений

[Добавить строку](#)

Имя поля	Исходный тип поля	Тип поля
No Rows To Show		

Настройки метрик

[Добавить строку](#)

Имя поля	Исходный тип поля	Имя поля DRUID	Вид колонки
No Rows To Show			

Рисунок 116. Форма спецификации druid

- в поле «Гранулирование сегмента» рекомендовано выбирать «День», гранулирование сегмента определяет степень детализации для создания фрагментов времени, за один временной отрезок можно создать несколько сегментов, например, если выбрать «День», то события одного и того же дня попадают в один и тот же временной фрагмент, который можно дополнительно разделить на несколько сегментов на основе других конфигураций и размера входных данных

Гранулирование сегмента

- Час
- День
- Неделя
- Месяц
- Год
- Все

Рисунок 117. Гранулирование сегмента

- в поле «Расписание службы Druid в минутах» указывается продолжительность задачи — это промежуток времени, по истечении которого задачи перестают читать и начинают публиковать свой сегмент, рекомендовано указывать 60 минут;
- параметр «Ручная настройка колонок источника данных», данный параметр позволяет влиять на настройку измерений и метрик, по умолчанию в настройках для всех полей установлен строковый тип данных

Настройка источника данных

Ручная настройка timestamp

Настройка timestamp

Имя поля timestamp: priznak0

Формат timestamp: auto

Значение по умолчанию:

Настройки измерений

Добавить строку | Удалить выбранную строку

Имя поля	Исходный тип поля	Тип поля	
org3	text	string	
inn4	text	string	
fond6	float4	string	
ssch7	float4	string	

Настройки metrics

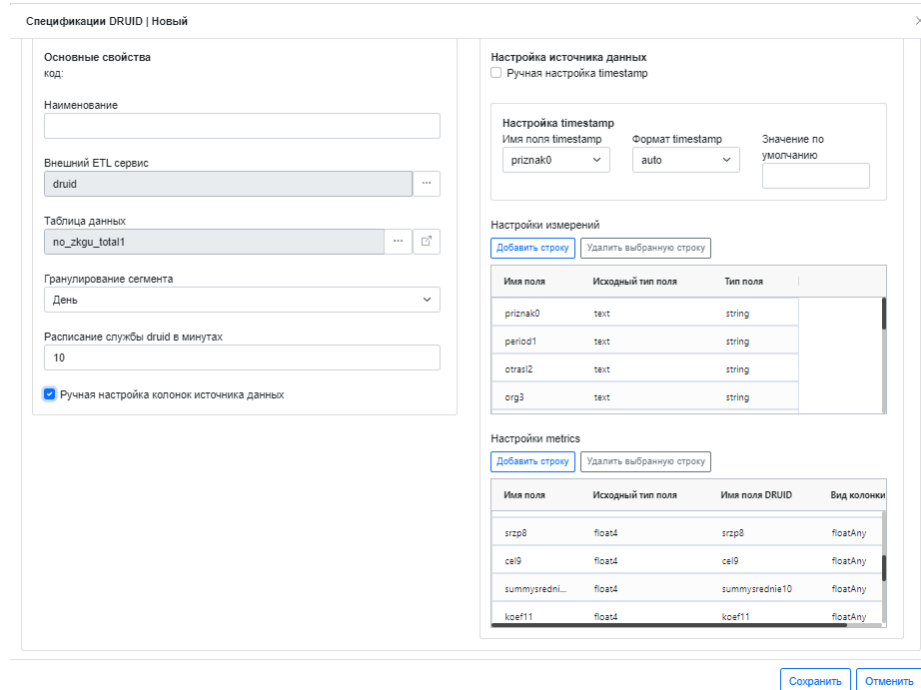
Добавить строку | Удалить выбранную строку

Имя поля	Исходный тип поля	Имя поля DRUID	Вид колонки
org3	text	org3	stringAny
inn4	text	inn4	stringAny
fond6	float4	fond6	stringAny
ssch7	float4	ssch7	stringAny

Сохранить | Отменить

Рисунок 118. Настройка источника данных

- a. если признак не установлен, в Druid будут передаваться настройки, определенные по умолчанию, изменить тип данных и количество полей данных в таком режиме невозможно;
- b. если признак установлен, то пользователь самостоятельно определяет перечень полей, передаваемых в измерениях и метриках,



а также производит настройки для маппинга полей данных. Рекомендовано в настройках измерений передавать поля со строковыми типами данных, а в настройках метрик – поля с числовыми типами данных.

Рисунок 119. Спецификация Druid

Поэтому поля, у которых исходный тип данных относится к числовому необходимо удалить из настроек измерений. В настройках метрик рекомендовано удалить все поля со строковыми типами данных или в «Виде колонок» выбрать «count», а также в настройках метрик для числовых полей изменить тип данных в колонке «Вид колонки» на числовой тип.

Признак «Ручная настройка timestamp», если данный признак не установлен, то timestamp создается автоматически, если признак установлен, то timestamp определяется из типов полей выбранной таблицы.

Рисунок 120. Настройка Timestamp

После произведенных настроек нажать кнопку «Сохранить».

Следующим шагом выполнения задачи на доставку данных из топика Kafka в Druid является создание шаблона заданий ETL.

В разделе «ETL» необходимо выбрать пункт «Шаблоны заданий ETL». Для добавления нового шаблона необходимо нажать кнопку «Новый», в открывшемся окне заполнить обязательные поля:

- ввести наименование
- выбрать вид ETL «Сервис публикации данных в Druid из Kafka»
- выбрать источник данных, из которого планируется забирать данные;
- в поле «Таблица данных» выбрать таблицу из списка объектов источника данных;
- выбрать топик Kafka;
- выбрать настройки спецификации Druid;
- задать имя создаваемой таблицы в Druid;
- установить количество загружаемых строк, при установленном параметре «0» будут загружены все строки;
- установить размер передаваемого пакета, рекомендованный размер не более 200 тыс. строк;
- заполнить поле «Пропустить от начала», если указать значение «0», то будут загружены строки начиная с первой, если необходимо загружать данные с определенной строки, указывается соответствующий номер строки;
- поле «Режим обновления таблицы» оставляем пустым;

- указать количество попыток при ошибке, если оставить поле пустым, то по умолчанию будет проставлено «3» попытки;
- указать интервал между попытками, если оставить поле пустым, то по умолчанию будет проставлено «60» секунд, затем нажать кнопку «Сохранить».

Шаблоны заданий ETL | Новый

Наименование
proverka-druid

Вид ETL
Сервис публикации данных в Druid из Kafka

Свойства шаблона

Основные свойства

Источник данных
ETL хранилище (PostgreSQL)

Таблица данных
proverka

ETL хранилище
ETL хранилище

Настройка топика KAFKA
proverka

Настройка спецификации DRUID
proverka-dr

Настройка спецификации VDW

Дополнительные свойства

Имя создаваемой таблицы
proverka

Кол-во строк (0-все)
0

Размер пакета
2000

Пропустить от начала
0

Режим обновления таблицы

Кол-во попыток при ошибке

Интервал между попытками в секундах

Сохранить Отменить

Рисунок 121. Шаблон задачи на отправку данных в Аналитическое хранилище

Следующим шагом выполнения задачи на доставку данных из топика Kafka в Druid является создание журнала заданий ETL на основании созданного шаблона ETL.

В разделе «ETL» необходимо выбрать пункт «Журнал заданий ETL». Для добавления новой задачи в журнал необходимо нажать кнопку «Новый», в открывшемся окне заполнить обязательные поля:

- ввести наименование
- выбрать шаблон задачи ETL, затем нажать кнопку «Сохранить»

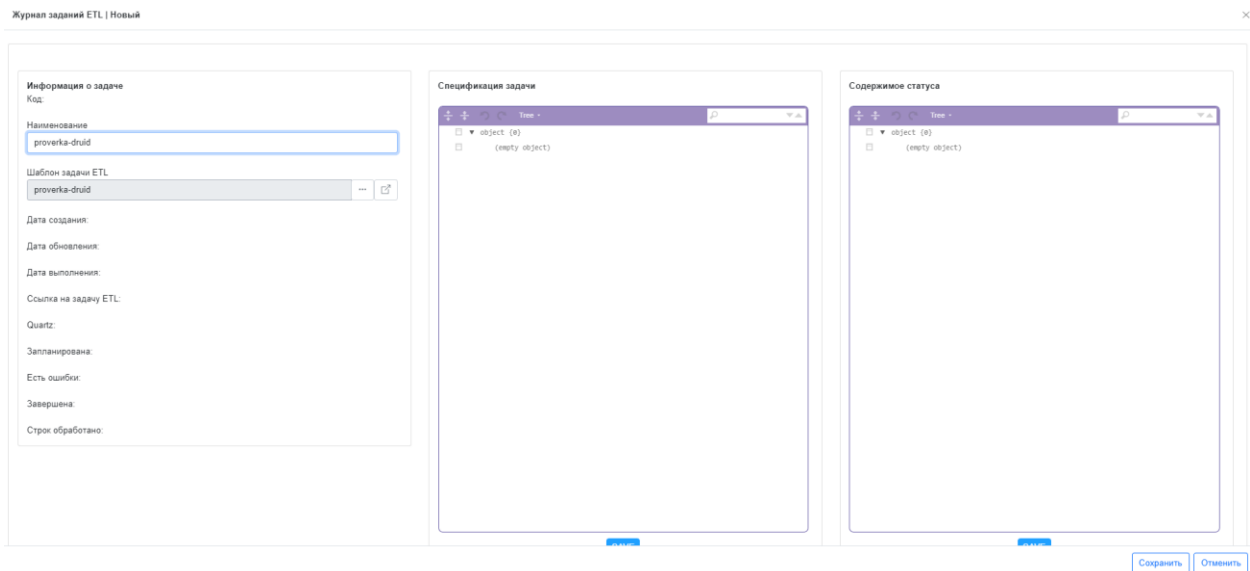


Рисунок 122. Создание задачи ETL

Задача может быть запущена немедленно или по расписанию, для немедленного запуска задачи на доставку данных из топика Kafka в Druid необходимо в списке задач выделить нужную и нажать кнопку «Запустить». Для получения статуса о ходе выполнения задачи необходимо нажать на кнопку «Обновить данные по задаче».

На данном этапе должен завершиться процесс получения данных из топика Kafka в Druid.

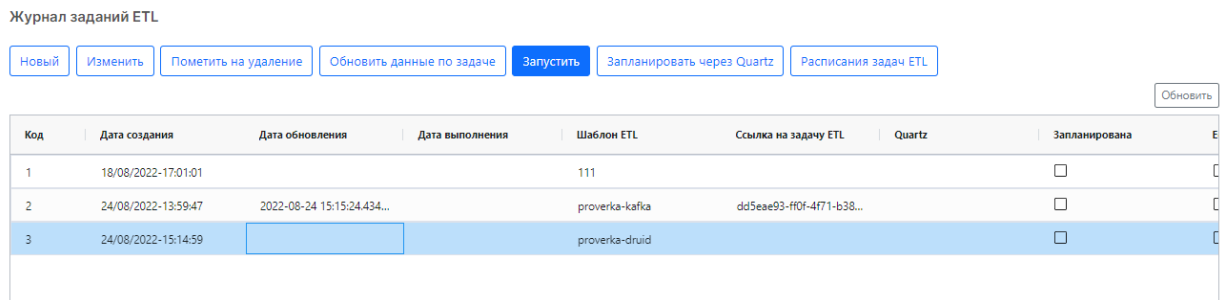


Рисунок 123. Запуск Задачи ETL

8.2.7 Импорт данных в форматах csv, json, excel. Назначение типов, перезапись и дозапись данных

В системе предусмотрена возможность извлечения данных из любых источников, в том числе импорт данных в формате csv, json, excel.

Для импорта данных необходимо перейти в меню «Конструктор», далее в «Модели данных» выбрать «ETL DWH»:

Далее в списке выбрать необходимое ETL-хранилище и нажать на кнопку «Импорт данных»:

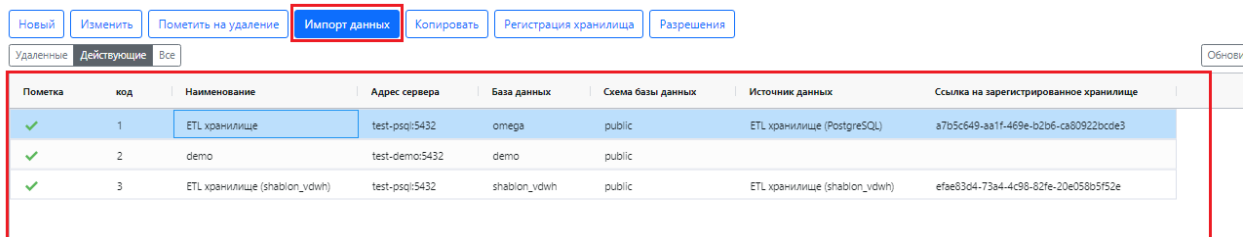


Рисунок 124. Выбор ETL-хранилища

Откроется форма импорта данных.

Далее необходимо нажать на кнопку «Выбрать файл» и выбрать загружаемый файл в формате csv, json, excel с компьютера.

Затем необходимо установить название источника и тип загружаемых данных.

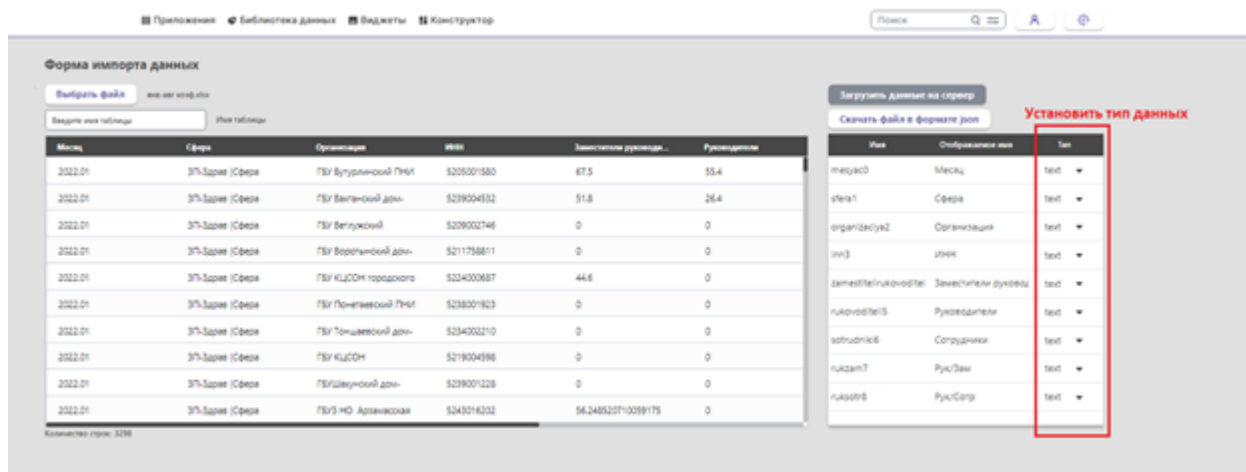


Рисунок 125. Установка типа данных для полей загружаемой таблицы

Источник можно загрузить как новый – задать новое, ранее не используемое наименование и нажать на кнопку «Загрузить данные на сервер», либо можно осуществить дозагрузку/перезапись в уже существующем в системе источнике, выбрав из числа наименований нужный источник и отметив в правой части «Добавить к имеющимся записям» или «Перезаписать имеющиеся записи», после чего нажать «Загрузить данные на сервер».

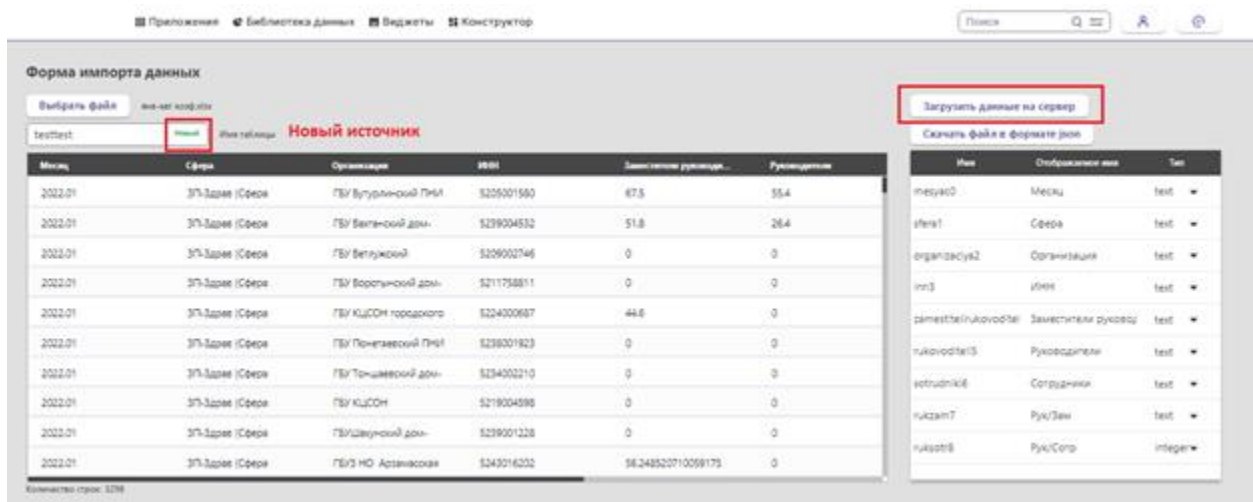


Рисунок 126. Добавление новой таблицы

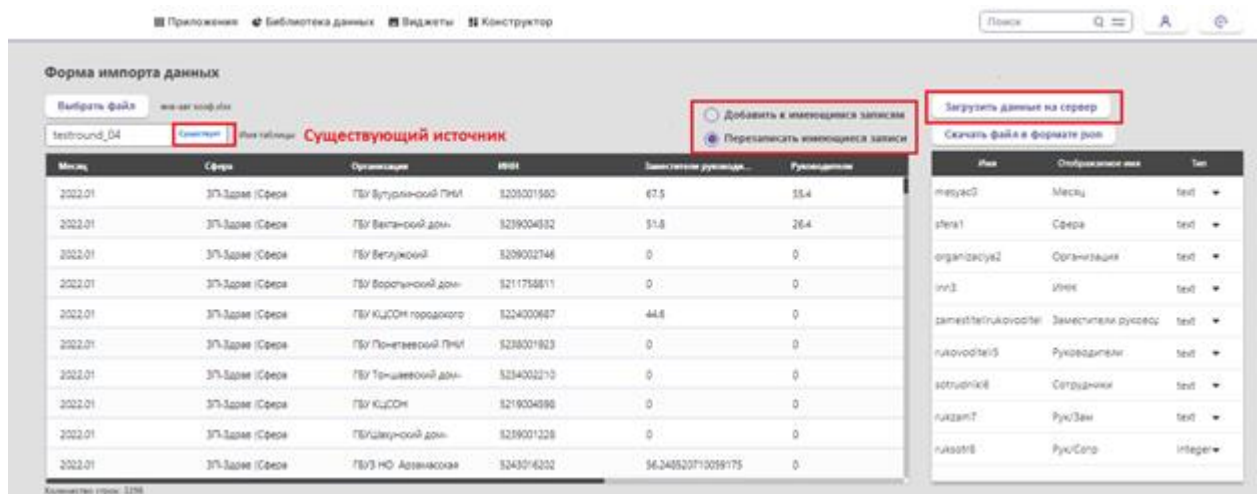


Рисунок 127. Перезапись существующей таблицы

После успешной загрузки данных отобразится сообщение: Данные загружены.

8.3 Настройка расписания

Каждая ETL-задача может быть запущена немедленно, а также в соответствии с расписанием. Управление расписанием выполняется с использованием Cron и компонента Quartz.

Quartz — это многофункциональный фреймворк планирования заданий с открытым исходным кодом, написанный на Java и предназначенный для интеграции с любым фреймворком J2EE или J2SE. Он предлагает огромную гибкость без ущерба для сложности или масштабируемости.

В системе будет использоваться Quartz для планирования выполнения задач ETL по расписанию следующим образом:

- точное время выполнения задания;
- многократное выполнение задания по сложному расписанию;
- выполнение задание по событию (триггеру);
- каскадное выполнение дерева заданий (ETL-flow).

Планировщик задач Cron – это система для автоматического запуска задач в определенное время. Cron - выражения предназначены для возможности задавать периодичность и время срабатывания задач планировщика.

Расписание задач можно добавлять из журнала заданий ETL. Для этого необходимо выделить задачу и нажать кнопку «Расписание задач ETL».

Журнал заданий ETL

Новый Изменить Пометить на удаление Обновить данные по задаче Запустить Запланировать через Quartz **Расписания задач ETL** Обновить

Код	Дата создания	Дата обновления	Дата выполнения	Шаблон ETL	Ссылка на задачу ETL	Quartz	Зап
3	09.09.2022 08:35:06	09.09.2022 08:35:27	09.09.2022 08:35:27	proverka-druid	7c964dbb-c6fd-4cef-		<input type="checkbox"/>
4	09.09.2022 08:37:44	09.09.2022 08:37:51	09.09.2022 08:37:51	proverka-kafka	5bfdcf77-16cb-4b77-		<input type="checkbox"/>

Рисунок 128. Вызов Расписания

Также можно перейти в раздел «Системные настройки» и выбрать пункт «Расписание».

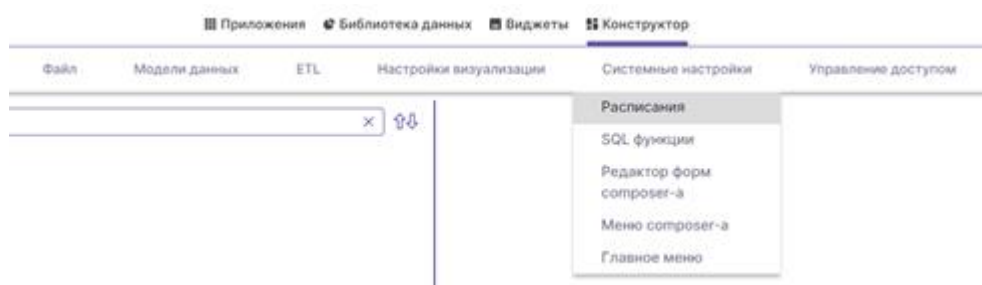


Рисунок 129. Расписание

Для добавления нового расписания необходимо нажать кнопку «Новый», в открывшемся окне заполнить обязательные поля:

- ввести наименование
- выбрать коннектор к источникам данных: «quartz»
- выбрать ETL-задачу, для которой необходимо настроить расписание

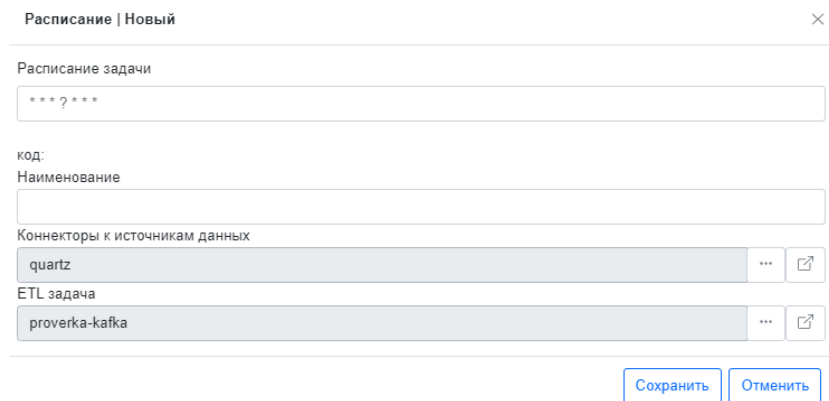


Рисунок 130. Экран настройки расписания

- для установки расписания необходимо нажать левой кнопкой мыши на поле «Расписание задачи», после чего раскроется область для выбора параметров нового расписания, где необходимо будет задать частоту повторов выполнения задачи. Область для выбора параметров расписания представлена в виде нескольких вкладок – «Секунды», «Минуты», «Часы», «Дни», «Месяц», «Год». На каждой из вкладок необходимо задать нужные параметры, по умолчанию на каждой вкладке установлены параметр, означающий, что задача будет выполняться каждую секунду, каждую минуту, каждый день и т.д. В поле «Расписание задачи» формируется Cron-выражение. Cron-выражение представляет собой строку, содержащую 7 символов, которые описывают отдельные значения расписания. Cron-выражение может быть простым, например: «* * * ? * * *», или более сложным, например: «0/5 14,18,3-39,52 * ? JAN,MAR,SEP MON-FRI 2020-2024». Под закодированным Cron-выражением формируется выражение в читаемом текстовом формате, например: «Каждые 30 мин».

Расписание | Новый

Расписание задачи
* 30 * ? * * *

Расшифровка: ["Каждую секунду, в 30 минут"]

Каждую минуту
 Каждую минуту начиная с
 Конкретная минута 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
 16 17 18 19 20 21 22 23 24 25 26 27 28 29
 30 31 32 33 34 35 36 37 38 39 40 41 42 43
 44 45 46 47 48 49 50 51 52 53 54 55 56 57
 58 59
 Каждую минуту между и

код:
 Наименование
 Расписание для загрузки данных в kafka каждые 30 мин
 Коннекторы к источникам данных
 quartz ...
 ETL задача
 proverka-kafka ...

Рисунок 131. Настройка выражения

После заполнения всех параметров необходимо сохранить расписание нажатием на кнопку «Сохранить».

Для просмотра расписания по конкретной задаче, необходимо в «Журнале заданий ETL» выбрать задачу и нажать кнопку «Расписание задач ETL».

Для активирования расписания необходимо выбрать задачу и нажать кнопку «Запланировать через Quartz».

Журнал заданий ETL

Код	Дата создания	Дата обновления	Дата выполнения	Шаблон ETL	Ссылка на задачу ETL	Quartz	Запланирована	E
1	18/08/2022-17:01:01			111			<input type="checkbox"/>	[
2	24/08/2022-13:59:47	2022-08-24 15:15:24.434...		proverka-kafka	dd5eae93-ff0f-4f71-b38...		<input type="checkbox"/>	[
3	24/08/2022-15:14:59			proverka-druid			<input type="checkbox"/>	[

Рисунок 132. Запуск расписания

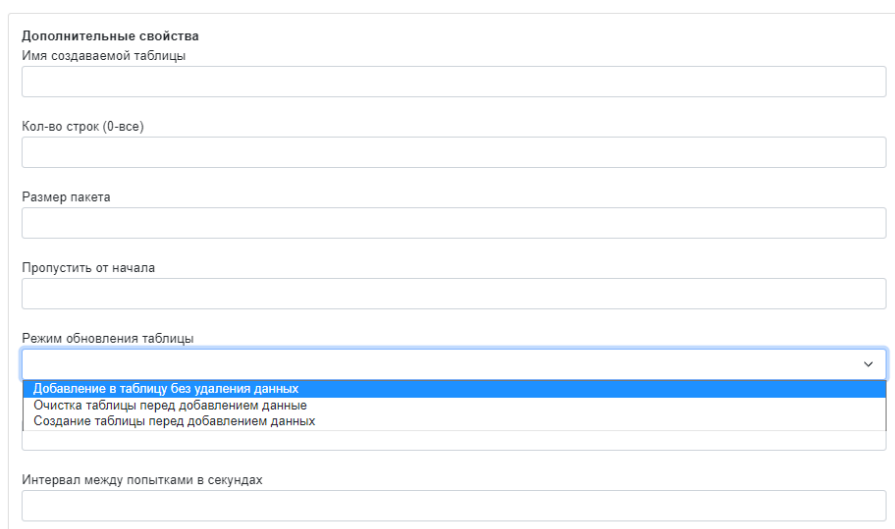
8.4 Перезапись данных, инкрементальное обновление данных

Когда объекты источников данных содержат большое количество данных из постоянно обновляющихся исходных баз данных, перезагрузка набора данных полностью может занять много времени. В этом случае из базы данных

загружаются только новые или измененные записи. Все остальные данные должны быть уже доступны в системе.

Для добавления новых записей используется инкрементальная загрузка данных.

Функционал инкрементальной загрузки данных реализован на этапе создания шаблонов задач ETL. При создании шаблона задачи на извлечение данных в ETL-хранилище в поле «Режим обновления таблицы» необходимо выбрать параметр «Добавление в таблицу без удаления данных».



The image shows a configuration window titled 'Дополнительные свойства' (Additional properties) for creating a table. It contains several input fields: 'Имя создаваемой таблицы' (Table name), 'Кол-во строк (0-все)' (Number of rows), 'Размер пакета' (Batch size), 'Пропустить от начала' (Skip from start), 'Режим обновления таблицы' (Table update mode), and 'Интервал между попытками в секундах' (Interval between attempts in seconds). The 'Режим обновления таблицы' dropdown menu is open, showing three options: 'Добавление в таблицу без удаления данных' (Add to table without deleting data), 'Очистка таблицы перед добавлением данных' (Clean table before adding data), and 'Создание таблицы перед добавлением данных' (Create table before adding data). The first option is highlighted in blue.

Рисунок 133. Режимы обновления таблиц

Если необходимо полностью обновить все данные или изменить набор полей в источнике, то нужно предусматривать сценарии для полной перезагрузки данных.

Механизм полной перезагрузки данных реализован двумя способами:

- Очистка содержимого таблицы и загрузка обновленными данными. Данный способ подразумевает полное стирание содержимого таблицы и загрузку таблицы новыми данными, таблица при этом не удаляется. Для использования данного механизма необходимо в поле «Режим обновления таблицы» шаблона заданий ETL выбрать параметр «Очистка таблицы перед добавлением данных».
- Удаление таблицы и загрузка новых данных. Данный способ подразумевает удаление, а затем создание таблицы и загрузку

таблицы новыми данными. Для использования данного механизма необходимо в поле «Режим обновления таблицы» шаблона заданий ETL выбрать параметр «Создание таблицы перед удалением данных».

8.5 Настройка кросс-фильтрации

Фильтры – это функции системы, которые помогают пользователю глубже вникать в данные, чтобы получить понимание контекста. Значение фильтров в визуальной аналитике заключается в том, что отфильтрованные данные напрямую влияют на визуализацию, облегчая сбор информации. В системе используются такие фильтры как кросс-фильтрация и глобальные фильтры.

Кросс-фильтрация является одним из элементов управления визуализацией, который обеспечивает интерактивное взаимодействие связанных между собой графиков. Механизм кросс-фильтрации заключается в том, что если выбрать элемент одной диаграммы, то по нему отфильтруются данные в других диаграммах. Выбирать можно один или несколько элементов диаграммы, для множественного выбора необходимо удерживать клавишу Ctrl.

Кросс-фильтрация системы позволяет применять фильтры в обе стороны, т.е. позволяет сначала применять фильтр по одной диаграмме, затем по другим диаграммам. С помощью кросс-фильтра можно с легкостью сделать несколько выборов, чтобы набор данных был именно такой, как вам нужно. С помощью точно определенного набора данных можно исследовать те данные, которые представляют особый интерес. При этом система запоминает состояние фильтрации и позволяет производить движение пользователя «вперед – назад» для ускорения получения ответов на вопросы в ходе сценарного анализа.

Для настройки кросс-фильтрации на дашборде следует для начала разместить все необходимые графики на дашборде.

Затем перейти в «Конструктор», зайти в раздел «Настройки визуализации» и выбрать пункт «Дашборды».

Найти нужный дашборд, выбрать его и нажать кнопку «Настроить правила кросс-фильтрации».

Для того, чтобы просмотреть правила кросс-фильтрации, необходимо выбрать нужный дашборд и нажать кнопку «Правила кросс-фильтрации по дашборду».

Пометка	код	Наименование	Наименование прил...
✓	1	ГФ	Кросс-фильтрация
✓	2	КФ	Кросс-фильтрация
✓	3	Дебиторская задолженность	Глобальные фильтры
✓	4	КФО	Глобальные фильтры
✓	5	Помещения	Новое приложение
✓	6	Сроки строительства	Ново приложение 2
✓	7	Сроки	Сроки
✓	8	Объекты строительства	Недвижимость

Рисунок 134. Правила кросс-фильтрации

После этого откроется окно, в котором будут доступны все правила, сформированные для конкретного дашборда для просмотра и корректировки.

Пометка	Правило пользователя	Наименование дашб...	Виджет источник	Поле источник	is UP	Правило	is UP
✓	<input type="checkbox"/>	Объекты строительства	кол-во комнаты	room5_komnaty0	<input checked="" type="checkbox"/>	to right	<input type="checkbox"/>
✓	<input type="checkbox"/>	Объекты строительства	Кол-во объектов группа	room5_obektnedvizhim...	<input checked="" type="checkbox"/>	to right	<input type="checkbox"/>
✓	<input type="checkbox"/>	Объекты строительства	кол-во комнаты	room5_komnaty0	<input checked="" type="checkbox"/>	to right	<input type="checkbox"/>
✓	<input type="checkbox"/>	Объекты строительства	карточка_сумма	price5_summa1	<input checked="" type="checkbox"/>	to right	<input type="checkbox"/>
✓	<input type="checkbox"/>	Объекты строительства	Соотношение объектов	price5_ploshad2	<input checked="" type="checkbox"/>	to right	<input type="checkbox"/>
✓	<input type="checkbox"/>	Объекты строительства	Кол-во объектов группа	staty5_sinomim4	<input type="checkbox"/>	to left	<input checked="" type="checkbox"/>
✓	<input type="checkbox"/>	Объекты строительства	Кол-во объектов группа	room5_obektnedvizhim...	<input checked="" type="checkbox"/>	to right	<input type="checkbox"/>
✓	<input type="checkbox"/>	Объекты строительства	Кол-во объектов группа	room5_grupparplosch...	<input checked="" type="checkbox"/>	to right	<input type="checkbox"/>
✓	<input type="checkbox"/>	Объекты строительства	кол-во комнаты	room5_komnaty0	<input checked="" type="checkbox"/>	to right	<input type="checkbox"/>
✓	<input type="checkbox"/>	Объекты строительства	кол-во комнаты	room5_komnaty0	<input checked="" type="checkbox"/>	to right	<input type="checkbox"/>
✓	<input type="checkbox"/>	Объекты строительства	карточка_сумма	room5_obektnedvizhim...	<input type="checkbox"/>	to left	<input checked="" type="checkbox"/>
✓	<input type="checkbox"/>	Объекты строительства	Кол-во объектов группа	room5_grupparplosch...	<input checked="" type="checkbox"/>	to right	<input type="checkbox"/>
✓	<input type="checkbox"/>	Объекты строительства	кол-во комнаты	room5_komnaty0	<input checked="" type="checkbox"/>	to right	<input type="checkbox"/>
✓	<input type="checkbox"/>	Объекты строительства	Кол-во объектов группа	room5_obektnedvizhim...	<input checked="" type="checkbox"/>	to right	<input type="checkbox"/>
✓	<input type="checkbox"/>	Объекты строительства	Соотношение объектов	room5_obektnedvizhim...	<input checked="" type="checkbox"/>	to right	<input type="checkbox"/>
✓	<input type="checkbox"/>	Объекты строительства	Соотношение объектов	room5_obektnedvizhim...	<input checked="" type="checkbox"/>	to right	<input type="checkbox"/>

Рисунок 135. Правила кросс-фильтрации

Если диаграммы на дашборде построены на разных таблицах, то следует соединить таблицы с помощью виртуальных типов данных. Для начала

необходимо добавить сам виртуальный тип, в разделе «Модели данных» выбрать пункт «Виртуальные типы данных», нажать кнопку «Создать» и в открывшемся окне заполнить наименование виртуального типа.

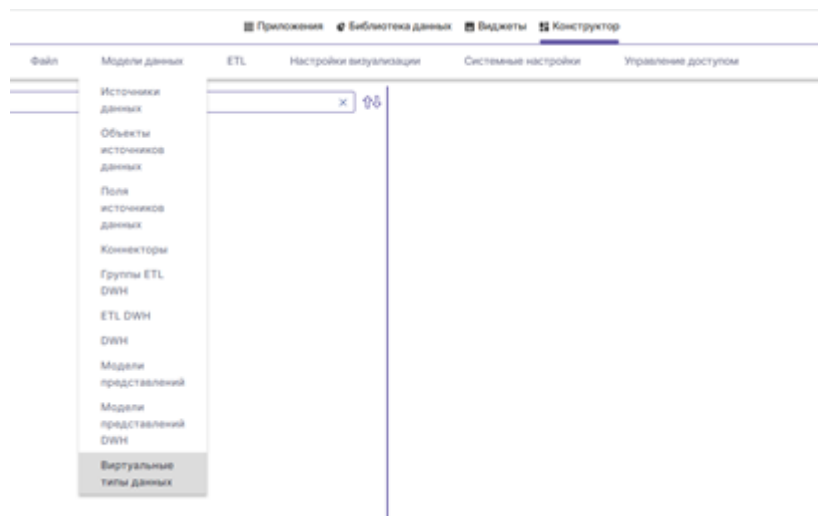


Рисунок 136. Виртуальные типы данных

Далее в разделе «Модели данных» выбрать пункт «Объекты источников данных».

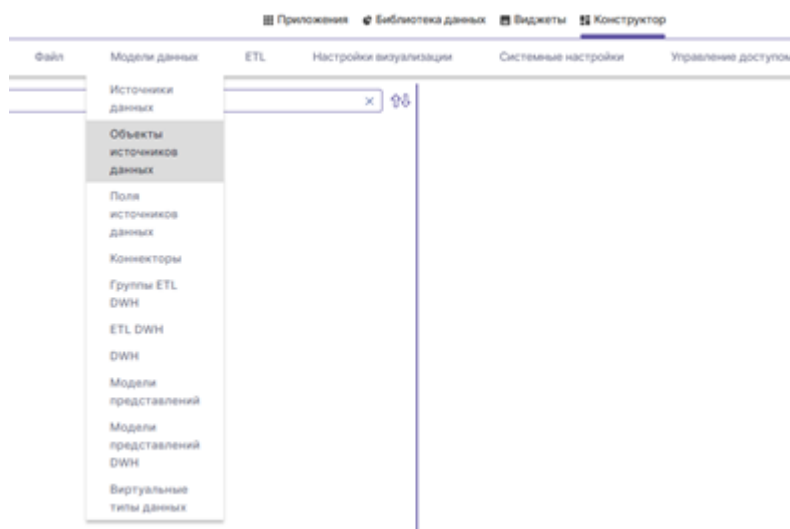


Рисунок 137. Объекты источников данных

Далее поочередно выбрать таблицы, которые необходимо соединить, нажать кнопку «Поля источников данных».

Объекты источников данных

Изменить Пометить на удаление Поля источника данных Просмотр источника данных Экспорт данных

Удаленные Действующие Все Обновить

Пометка	Опубликовано	код	Наименование	Имя представления	Источник данных	Тип таблицы
✓	<input type="checkbox"/>	39	room_objects		Druid	Исходные данные
✓	<input type="checkbox"/>	40	demo_kfo_aggr		Druid	Исходные данные
✓	<input type="checkbox"/>	41	eventsdruid		Druid	Исходные данные
✓	<input type="checkbox"/>	42	demo_dz_aggr		Druid	Исходные данные
✓	<input type="checkbox"/>	43	room_obj_4		Druid	Исходные данные
✓	<input checked="" type="checkbox"/>	44	room_druid2		Druid	Исходные данные
✓	<input type="checkbox"/>	45	v_room		ETL-хранилище (PSQL)	Исходные данные
✓	<input type="checkbox"/>	46	s_price6	s_price6	ETL-хранилище (PSQL)	Исходные данные
✓	<input type="checkbox"/>	47	v_room_druid		Druid	Исходные данные
✓	<input type="checkbox"/>	48	cena	cena	ETL-хранилище (PSQL)	Исходные данные
✓	<input checked="" type="checkbox"/>	49	nedvigimost		ETL-хранилище (PSQL)	Исходные данные
✓	<input type="checkbox"/>	50	room5	room5	ETL-хранилище (PSQL)	Исходные данные
✓	<input type="checkbox"/>	51	srok5	srok5	ETL-хранилище (PSQL)	Исходные данные

Рисунок 138. Поля источников данных (1)

В отрывшемся окне найти и выбрать поле, по которому будут соединяться таблицы.

Назад

Поля источников данных

Изменить Пометить на удаление

Удаленные Действующие Все Обновить

Пометка	код	Наименование	Имя представления	Объект данных	Имя поля	Тип поля
✓	434	obekt0	Объект недвижимости	nedvigimost	obekt0	исходный тип
✓	435	dataregistracii1	Дата регистрации	nedvigimost	dataregistracii1	исходный тип
✓	436	mesyac2	Месяц	nedvigimost	mesyac2	исходный тип
✓	437	kvartal3	Квартал	nedvigimost	kvartal3	исходный тип
✓	438	god4	Год	nedvigimost	god4	исходный тип
✓	439	ploschad5	Площадь	nedvigimost	ploschad5	исходный тип
✓	440	razdel6	Раздел	nedvigimost	razdel6	исходный тип
✓	441	dataoperacii7	Дата операции	nedvigimost	dataoperacii7	исходный тип
✓	442	etapsdelki8	Этап сделки	nedvigimost	etapsdelki8	исходный тип
✓	443	organizaciyazastroichik9	Организации	nedvigimost	organizaciyazastroichik9	исходный тип
✓	444	summadogovora10	Сумма договора	nedvigimost	summadogovora10	исходный тип
✓	445	ploschadprivedennaya11	Площадь приведенная	nedvigimost	ploschadprivedennaya11	исходный тип
✓	446	cenam212	Цена за м2	nedvigimost	cenam212	исходный тип
✓	447	summatogo13	Сумма Итого	nedvigimost	summatogo13	исходный тип
✓	448	tippomesheniya14	Тип помещения	nedvigimost	tippomesheniya14	исходный тип
✓	449	gruppировка15	Группировка	nedvigimost	gruppировка15	исходный тип

Рисунок 139. Поля источников данных (2)

Двойным нажатием левой кнопки мыши зайти в настройки поля источников данных и в поле «Виртуальный тип поля» выбрать нужный виртуальный тип. Виртуальный тип данных для обеих таблиц должен совпадать.

Поля источников данных | Изменить ×

Объект: pedvigimost

код: 448	Наименование: tipromescheniya14		
Имя представления			
Тип помещения			
Тип поля			
... [X] [↗]			
Виртуальный тип поля			
Тип помещения			
... [X] [↗]			
Имя поля			
tipromescheniya14			
Свойства поля			
Тип	<input type="checkbox"/> Должно быть заполнено (NOT NULL)	Длина	Точность
text		2147483647	0
Описание назначения поля			

Сохранить Отменить

Рисунок 140. Виртуальный тип поля

При необходимости можно скорректировать, добавить или удалить правила кросс-фильтрации. Для добавления правила кросс-фильтрации необходимо нажать кнопку «Новый». В открывшемся окне заполнить все обязательные и необходимые для корректной работы поля.

Параметры для правил кросс-фильтрации, если диаграммы построены на одной таблице:

- в поле «Левый элемент витрины данных» необходимо выбрать диаграмму;
- в поле «Левый источник данных» выбрать хранилище, в котором хранится таблица, на основании которой построена диаграмма;
- поле «Поле левой таблицы данных» выбрать таблицу, на основании которой построена диаграмма;
- заполнить «Псевдоним поля слева», если признак «is up» активный;
- параметр «is up» в настройках левого элемента: установить флажок, если в поле «Логическая операция» значение «в обе стороны» или «вправо»;
- заполнить «Наименование»;
- заполнить «Имя представления»;
- заполнить «Описание правила»;
- логическая операция:

- а. если на диаграмме, где производится отбор и фильтрующей диаграмме совпадают поля, то в поле «логическая операция» необходимо выбрать «в обе стороны»;

Правила кросс-фильтрации | Изменить

Настройки левого элемента
Левый элемент витрины данных
test_выполнение

Левый источник данных
ETL хранилище (PostgreSQL)

Левая таблица данных
test_total

Поле левой таблицы данных
vyroplenie13

Виртуальный тип левого поля

Псевдоним поля слева
test_total_vyroplenie13_Avg

is UP

Правило глобального фильтра

Наименование
bidirectional

Имя представления
bidirectional

Описание правила
bidirectional

Дашборд
test1

Логическая операция
В обе стороны

Правило пользователя
 Правило активно
 Правило корректно

Настройки правого элемента
Правый элемент витрины данных
test_организации

Правый источник данных
ETL хранилище (PostgreSQL)

Правая таблица данных
test_total

Поле правой таблицы данных
vyroplenie13

Виртуальный тип правого поля

Псевдоним поля справа
test_total_vyroplenie13_Avg

is UP

Рисунок 141. Правила кросс-фильтрации (1)

- б. если на диаграмме, где производится отбор и фильтрующей диаграмме поля не совпадают, то в поле «логическая операция» необходимо выбрать «вправо» и в настройках левого элемента указать диаграмму, на которой осуществляется отбор, а в настройках правого элемента – фильтрующуюся диаграмму;

Правила кросс-фильтрации | Изменить

Настройки левого элемента
Левый элемент витрины данных
test_выполнение

Левый источник данных
ETL хранилище (PostgreSQL)

Левая таблица данных
test_total

Поле левой таблицы данных
uk5

Виртуальный тип левого поля

Псевдоним поля слева
test_total_uk5

is UP

Правило глобального фильтра

Наименование
to right

Имя представления
to right

Описание правила
to right

Дашборд
test1

Логическая операция
Вправо

Правило пользователя
 Правило активно
 Правило корректно

Настройки правого элемента
Правый элемент витрины данных
test_организации

Правый источник данных
ETL хранилище (PostgreSQL)

Правая таблица данных
test_total

Поле правой таблицы данных
uk5

Виртуальный тип правого поля

Псевдоним поля справа

is UP

Рисунок 142. Правила кросс-фильтрации (2)

- с. если на диаграмме, где производится отбор и фильтрующей диаграмме поля не совпадают, то в поле «логическая операция» необходимо выбрать «влево» и в настройках правого элемента указать диаграмму, на которой осуществляется отбор, а в настройках левого элемента – фильтрующуюся диаграмму;

Правила кросс-фильтрации | Изменить

Рисунок 143. Правила кросс-фильтрации (3)

- установить флажок в «Правила пользователя», если правила настраивались и корректировались вручную, в ином случае при очередном обновлении правила кросс-фильтрации правило созданное вручную не сохранится;
- «Правило активно»: установить флажок, чтобы активировать правило;
- в поле «Правый элемент витрины данных» необходимо выбрать диаграмму;
- в поле «Правый источник данных» выбрать хранилище, в котором хранится таблица, на основании которой построена диаграмма;
- поле «Поле правой таблицы данных» выбрать таблицу, на основании которой построена диаграмма;
- заполнить «Псевдоним поля справа», если признак «is up» активный;
- параметр «is up» в настройках правого элемента: установить флажок, если в поле «Логическая операция» значение «в обе стороны» или «ВЛЕВО».

Параметры для правил кросс-фильтрации, если диаграммы построены на разных таблицах:

- Правила кросс-фильтрации настраиваются по аналогии с правилами, описанными выше (для диаграмм, построенных на одной таблице), только для поля, по которому соединяются таблицы нужно указать виртуальный тип данных.

Рисунок 145. Правила кросс-фильтрации (4)

Использовать функционал кросс-фильтрации можно только в режиме просмотра дашборда.

8.6 Настройка глобальных фильтров (виртуальные типы данных, настройка)

Для удобного и наглядного анализа данных в системе предусмотрена возможность отбора данных на уровне приложений с помощью глобальных фильтров.

Как правило, дашборд дает сводную информацию по ключевым показателям компании и, если её недостаточно, данные можно детализировать и фильтровать. Например, если необходимо отфильтровать все графики дашборда и приложения по периоду, то это можно сделать с помощью глобального фильтра, который создается на основе общего поля для всех источников данных диаграмм. Отбор, установленный в глобальном фильтре, влияет на все дашборды приложения.

Независимо от того построены ли все диаграммы дашборда на информации из одного источника или из разных источников, глобальные фильтры настраиваются одинаково при помощи виртуальных типов данных.

Для настройки глобальных фильтров необходимо зайти в раздел «Настройки визуализации» и выбрать пункт «Глобальные фильтры».

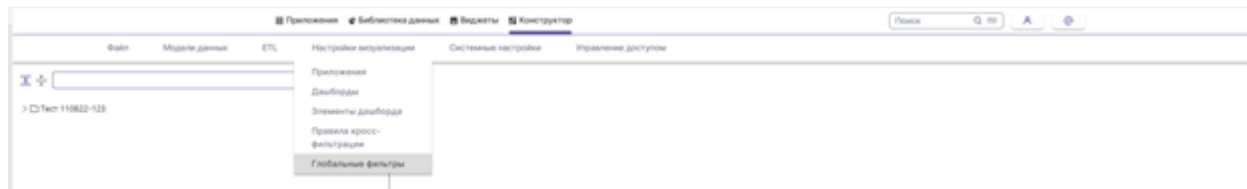


Рисунок 146. Глобальные фильтры (1)

Для добавления нового глобального фильтра нажать кнопку «Новый» и заполнить все обязательные поля, необходимые для корректной работы фильтра:

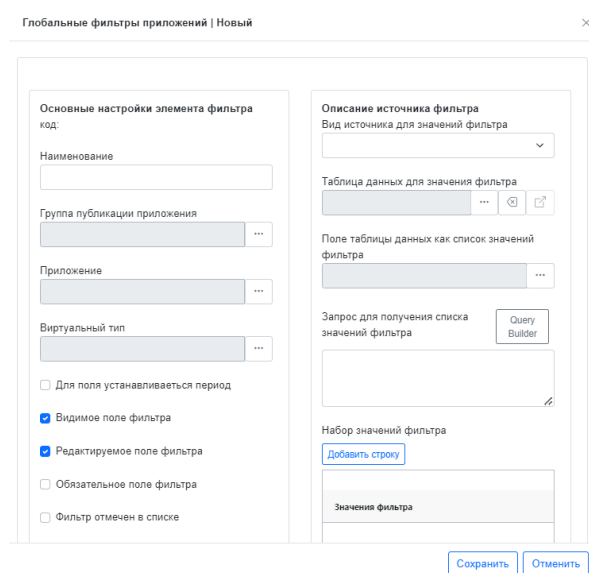


Рисунок 147. Глобальные фильтры (2)

- заполнить наименование;
- в поле «Группа публикации приложения» выбрать предварительно настроенную группу, включающую в себя приложение, для которого настраивается фильтр и пользователей, которые будут работать с приложением;
- в поле «Приложение» выбрать непосредственно приложение, для которого настраивается фильтр;

- поле «Виртуальный тип»: полю, по которому будет устанавливаться отбор, необходимо присвоить виртуальный тип данных. Для начала необходимо добавить сам виртуальный тип, в разделе «Модели данных» выбрать пункт «Виртуальные типы данных», нажать кнопку «Создать» и в открывшемся окне заполнить наименование виртуального типа. Далее в разделе «Модели данных» выбрать пункт «Объекты источников данных», поочередно выбрать таблицы, на основании которых построены таблицы, нажать кнопку «Поля источников данных».

Пометка	Опубликовано	код	Наименование	Имя представления	Источник данных	Тип таблицы
✓	<input checked="" type="checkbox"/>	1	test_dost100		ETL хранилище (Postgre...	Исходные данные
✓	<input type="checkbox"/>	2	TOTAL		ETL хранилище (Postgre...	Исходные данные
✓	<input checked="" type="checkbox"/>	3	test_total		ETL хранилище (Postgre...	Исходные данные
✓	<input checked="" type="checkbox"/>	4	test_bk		ETL хранилище (Postgre...	Исходные данные
✓	<input checked="" type="checkbox"/>	5	test_total2		ETL хранилище (Postgre...	Исходные данные
✓	<input checked="" type="checkbox"/>	6	test_dost100222		ETL хранилище (Postgre...	Исходные данные
✓	<input type="checkbox"/>	7	no_zkgu_total1		ETL хранилище (Postgre...	Исходные данные

Рисунок 148. Объекты источников данных

В отрывшемся окне найти и выбрать поле, по которому будет осуществляться отбор в фильтрах, двойным нажатием левой кнопки мыши зайти в настройки поля источников данных и в поле «Виртуальный тип поля» выбрать нужный виртуальный тип. Виртуальный тип данных для всех таблиц должен совпадать.

Поля источников данных | Изменить ×

Объект: test_total

код: 26 Наименование: period1

Имя представления
Период

Тип поля

Виртуальный тип поля
test3

Имя поля
period1

Свойства поля

Тип Должно быть
заполнено (NOT
NULL) Длина Точность

Описание назначения поля

[Сохранить](#) [Отменить](#)

Рисунок 149. Виртуальный тип поля

- «Видимое поле фильтра»: если не установлен флажок, то фильтр, с предустановленным значением по умолчанию, будет невидимым для пользователя;
- «Редактируемое поле фильтра»: если не установлен флажок, то фильтр, с предустановленным значением по умолчанию, будет недоступным для редактирования;
- «Обязательное поле фильтра»: если установлен флажок, то фильтр обязателен для заполнения;
- «Фильтр отмечен в списке»: обязательный параметр при установленном значении по умолчанию;
- выбрать операцию сравнения;
- выбрать тип ввода поля;
- при необходимости установить значение фильтра по умолчанию;
- «Вид источника для значений фильтра»: выбрать тип «набор значений»;

- «Набор значений фильтра»: добавить значения фильтра, нажать кнопку «Сохранить».

Глобальные фильтры приложений | Изменить

Основные настройки элемента фильтра
код: 3

Наименование
test3

Группа публикации приложения
test3

Приложение
test3

Виртуальный тип
test3

Для поля устанавливается период

Видимое поле фильтра

Редактируемое поле фильтра

Обязательное поле фильтра

Фильтр отмечен в списке

Операция сравнения
=

Тип поля ввода
Список

Значение фильтра по умолчанию
2022.01 - 2022.07

Описание источника фильтра
Вид источника для значений фильтра
Набор значений

Таблица данных для значения фильтра

Поле таблицы данных как список значений фильтра

Запрос для получения списка значений фильтра
Query Builder

Набор значений фильтра
Добавить строку

Значения фильтра	
2022.01	
2022.01 - 2022.02	
2022.01 - 2022.03	

Сохранить Отменить

Рисунок 150. Набор значений фильтра

Далее выбрать сохраненное правило и нажать кнопку «Обновить правила кросс-фильтрации».

Если произведены корректные настройки глобального фильтра, то фильтр появится на второстепенной панели инструментов при открытии приложения (если в фильтре настроены видимые поля).



Рисунок 151. Выбор глобальных фильтров

9. Безопасность (создание пользователей, состав ролей, создание групп, назначение ролей на группу, разграничение доступа к приложениям, разграничение доступа к данным в ETL-хранилище и аналитическом хранилище)

Сочетание назначенных пользователю ролей и групп формирует набор функциональных возможностей по работе с подсистемой, доступных пользователю, и конкретных разделов аналитического хранилища, к которым пользователь имеет доступ, и возможности управлять и настраивать ETL-процесс.

9.1 Добавление пользователя

Добавить в систему нового пользователя можно во вкладке “Конструктор” → “Управление доступом” → “Пользователи”.

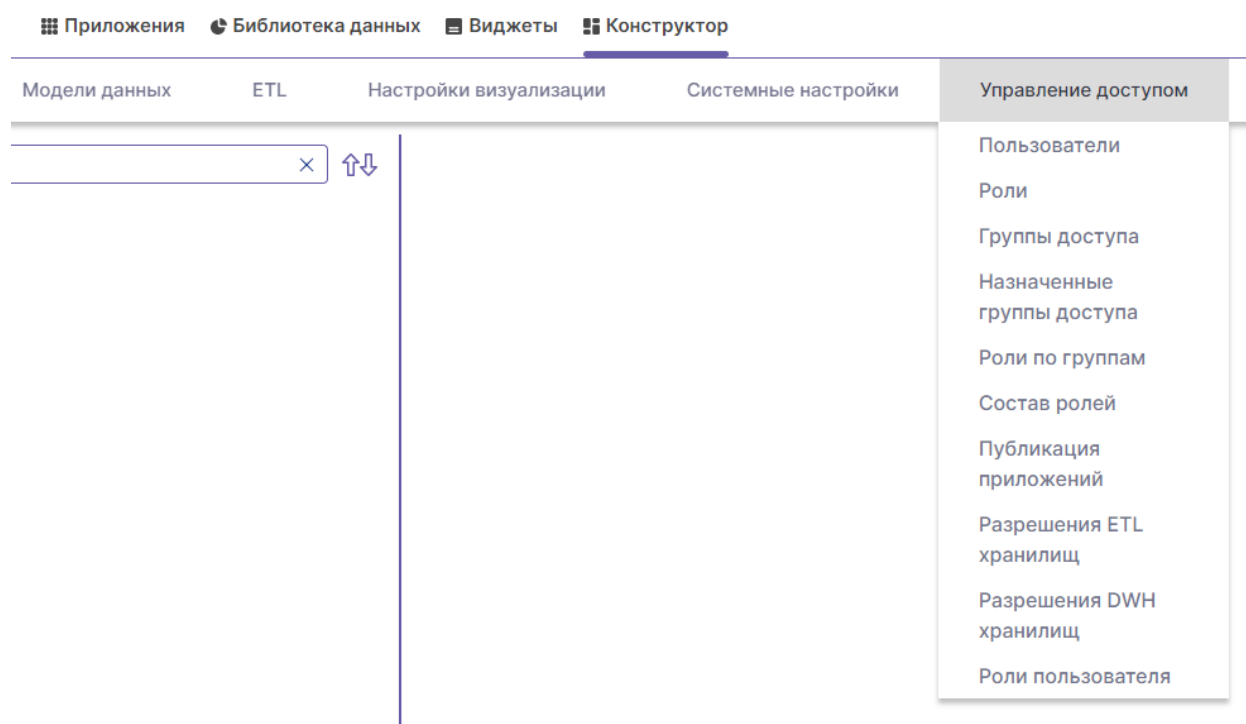


Рисунок 158. Управление доступом

Перейдя во вкладку, “Пользователи” вы увидите окно со списком всех пользователей, добавленных в домен (домен - автономная область, в том числе безопасности, которая характеризуется следующими признаками: уникальными

пользователями, ETL-хранилищами, аналитическим хранилищем, приложениями, дашбордами, данными).

Пользователи

Новый Изменить Пометить на удаление Снять пометку на удаление

Удаленные Действующие Все Обновить

Пометка	код	Наименование	Пользователь заблокирован	Ссылка на Spring Boot Secure
✓	1	kasv	false	d229b5f8-4777-4d17-bc4a-b393c1274ed7
✓	2	ruda	false	a82eb821-f271-4784-a7f8-a20e773f75bd
✓	5	artem_rudakov	false	8e2501ae-6d93-4c55-8217-79591aee4d18

Рисунок 159. Добавление нового пользователя

Для того, чтобы добавить нового пользователя выбираем “Новый”. Перед вами откроется окно следующего вида:

Пользователи | Новый

код:
Secure uuid:
Наименование
Имя
Фамилия
Адрес электронной почты

Пользователь заблокирован

Зарегистрировать пользователя Заблокировать пользователя Сбросить пароль Обновить Сохранить Отменить

Рисунок 160. Заполнение полей для нового пользователя

Где:

- Наименование – то, как пользователь будет отображаться в системе;
- Имя, фамилия – данные пользователя;
- Адрес электронной почты – почтовый ящик, на который будут приходить уведомления системы.

После заполнения указанных данных их необходимо сохранить, кликнув по кнопке “Сохранить” и только потом можно будет зарегистрировать пользователя, кликнув по соответствующей кнопке.

Данные для входа в систему придут на электронную почту.

9.2 Роли

Функциональные возможности работы пользователей с системой определяются такой сущностью, как «Роль». Доступ к информации, содержащейся в системе, и аналитическим приложениям определяются такой сущностью, как «Группы». Каждому пользователю может быть назначена одна или несколько следующих ролей:

1. «Мастер-администратор» — управляет доменами, настройками исходящей почты.
2. «Администратор домена» — назначение остальным пользователям ролей и групп, регистрация ETL хранилища, Аналитического хранилища, ETL-сервисов.
3. «Инженер данных» (Data Developer):
 - настройка процесса извлечения, преобразования данных и их загрузки в аналитическое хранилище, настройка расписания процесса (ETL-процесс);
 - работа с моделями данных в хранилище, создание дополнительных структур данных.
4. «Дизайнер приложений» (Data Designer) — конструирование аналитических приложений, предназначенных для работы остальных пользователей, на основе информации в аналитическом хранилище.
5. «Аналитик» — работа с назначенными по политике безопасности аналитическими приложениями, возможность создания собственных дополнительных дашбордов и графиков на них.
6. «Пользователь» — работа с аналитическими приложениями без возможности создания собственных дополнительных дашбордов и корректировки существующих, только режим просмотра.

9.3 Группы доступа

Доступ к информации, содержащейся в системе, и аналитическим приложениям определяются такой сущностью, как «Группы».

Группы доступа

Новый	Изменить	Пометить на удаление	Снять пометку на удаление	Копировать
Удаленные	Действующие	Все		
Пометка	код	Наименование	Группа - родитель	
✓	1	Администраторы		
✓	2	Инженеры данных		
✓	3	Дизайнеры приложений		
✓	4	Аналитики		
✓	5	Пользователи		

Рисунок 161. Группы доступа

Каждой «Группе» в подсистеме может быть назначены для доступа неограниченное количество аналитических приложений и источников (моделей данных) в аналитическом хранилище.

После создания необходимых групп в них добавляются конкретные пользователи, которые получают доступ к соответствующей информации, содержащейся в подсистеме.

9.4 Назначенные группы доступа

Данный раздел отображает список пользователей и группы, в которые они добавлены.

Назначенные группы доступа

Новый Изменить Пометить на удаление Снять пометку на удаление

Удаленные Действующие Все

Пометка	код	Дата изменения	Группа доступа	Пользователь
✓	1	2022-09-08	Администраторы	kasv
✓	3	2022-09-19	Администраторы	guda

Рисунок 162. Назначение группы доступа (1)

Для того, что изменить группу доступа или добавить в нее нового пользователя, выбираем группу и кликаем по кнопке “Изменить”.

Назначенные группы доступа | Изменить ×

код: 1
Начало действия: 2022-09-08 16:34:02.033230
Пользователи

kasv	...	✕	↗
------	-----	---	---

Группы доступа

Администраторы	...	✕	↗
----------------	-----	---	---

[Сохранить](#) [Отменить](#)

Рисунок 163. Назначение группы доступа (2)

Далее сохраняем внесенные изменения.

9.5 Роли по группам

В данном разделе можно назначить группе конкретную роль.

Роли по группам

Новый Изменить Пометить на удаление

Удаленные Действующие Все

Пометка	код	Группа доступа	Назначенная роль
✓	1	Администраторы	Администратор проекта
✓	2	Аналитики	Аналитик
✓	3	Дизайнеры приложений	Дизайнер приложений
✓	4	Инженеры данных	Инженер данных
✓	5	Пользователи	Пользователь

Рисунок 164. Назначение ролей

Алгоритм изменений аналогичный предыдущему пункту: выбираем группу, кликаем по кнопке “Изменить”. После сохраняем внесенные изменения.

Роли по группам | Изменить ×

код: 1
Группа доступа

Администраторы ... ⊗ ↗

Роль

Администратор проекта ... ⊗ ↗

Сохранить Отменить

Рисунок 165. Сохранение назначенной роли

10. Конструктор нейросетей

10.1 Шаг 1. Создание шаблона ML-задачи

Для того, чтобы перейти к созданию шаблона необходимо выбрать вкладку меню «Предиктивная аналитика» - «Шаблоны ML задач».

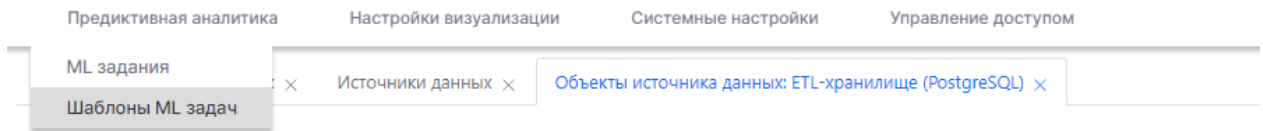


Рисунок 166. Вкладка «Предиктивная аналитика»

При переходе будет отображаться список всех созданных ранее шаблонов.

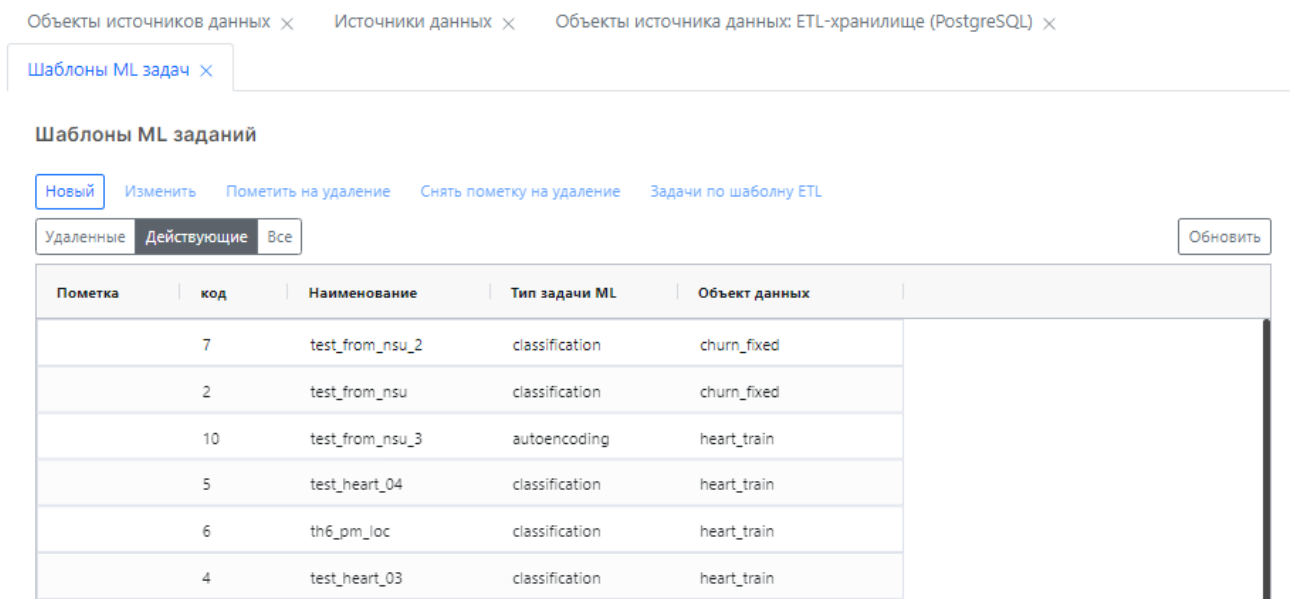


Рисунок 167. Список шаблонов ML задания

Создание шаблона ML-задачи в свою очередь делится на несколько основных этапов:

1. «Источник данных»

В данном разделе необходимо задать наименование шаблона, выбрать ETL-хранилище, которое будет использоваться, источник данных и таблицу. А также нужно указать количество строк (0 - все), размер пакета – 50000, пропустить от начала – 0.

Шаблон ML задачи
код: 17
Наименование
test2111

Свойства шаблона

Источник данных | Таблицы результата | Интерпретация | Преобработка | Тип задачи машинного обучения | Оценка качества модели | Настройка модели | Параметры обучения

ETL хранилище (база размещения результата)
ETL-хранилище

Источник данных
ETL-хранилище (PostgreSQL)

Таблица данных
train_attrition_report01

Кол-во строк (0-все)
0

Размер пакета
50000

Пропустить от начала
0

Сохранить | Отменить

Рисунок 168. Выбор источника данных

2. «Таблицы результатов»

На данной вкладке предзаполнены системные таблицы, которые будут хранить в себе результаты обучения и тестирования нейронной сети, сохраняя ее контрольные точки.

Шаблон ML задачи
код: 17
Наименование
test2111

Свойства шаблона

Источник данных | Таблицы результата | Интерпретация | Преобработка | Тип задачи машинного обучения | Оценка качества модели | Настройка модели | Параметры обучения

Таблица моделей
checkpoint_table

Таблица результатов предобработки
preprocess_table

Таблица результатов обучения
train_table

Сохранить | Отменить

Рисунок 169. Таблица результатов

3. «Интерпретация»

Необходимо из выбранного набора данных определить значимые признаки для обучения. Нажать «Добавить строку» и выбрать столбец таблицы, а также определить каким признаком он является: числовой, категориальный (текстовые столбцы) или целевой признак классификации (предсказываемый признак).

Шаблон ML задачи

код: 17

Наименование
test2111

Свойства шаблона

Источник данных Таблицы результата **Интерпретация** Препроцессинг Тип задачи машинного обучения Оценка качества модели Настройка модели Параметры обучения

Интерпретация

Добавить строку

Имя поля	Числовое значение	Категориальное знач...	Целевой признак кл...
age0	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
attrition1	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
department4	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>

Сохранить Отменить

Рисунок 170. Таблица результатов

4. «Препроцессинг»

Для того, чтобы провести препроцессинг (преобразовать данные для последующего обучения нейронной сети) признаков, которые мы выбрали на предыдущей вкладке, необходимо повторно их добавить (кроме целевого признака классификации). Далее определить каким способом мы будем преобразовывать данные: стандартизация (для числовых значений), повышения шкалы (для текстовых, категориальных признаков).

Шаблон ML задачи

код: 17

Наименование
test2111

Свойства шаблона

Источник данных Таблицы результата Интерпретация **Препроцессинг** Тип задачи машинного обучения Оценка качества модели Настройка модели Параметры обучения

Блоки препроцессинга

Добавить строку

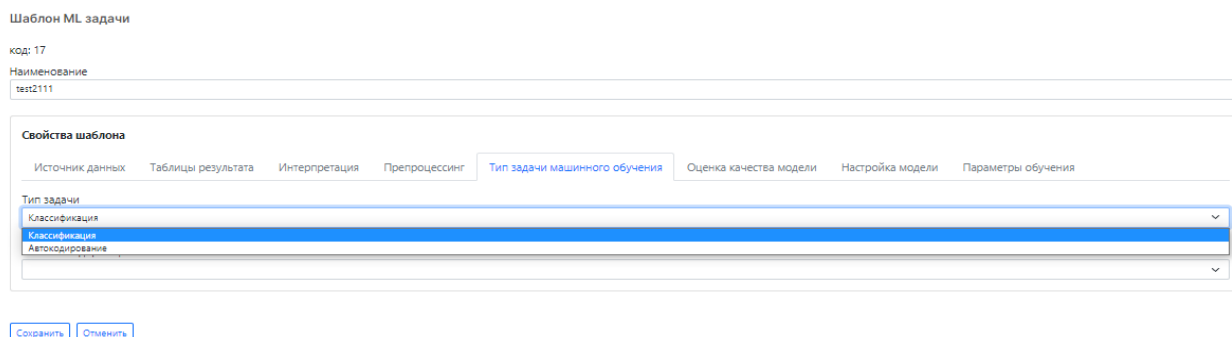
Имя признака	Тип блока
age0	Стандартизация
department4	Повышение шкалы

Сохранить Отменить

Рисунок 171. Препроцессинг

5. «Тип задачи машинного обучения»

На данной вкладке необходимо выбрать тип задачи машинного обучения: классификация, автоэнкодер.



Шаблон ML задачи
код: 17
Наименование
test2111

Свойства шаблона

Источник данных Таблицы результата Интерпретация Преобработка **Тип задачи машинного обучения** Оценка качества модели Настройка модели Параметры обучения

Тип задачи

- Классификация
- Классификация**
- Автокодирование

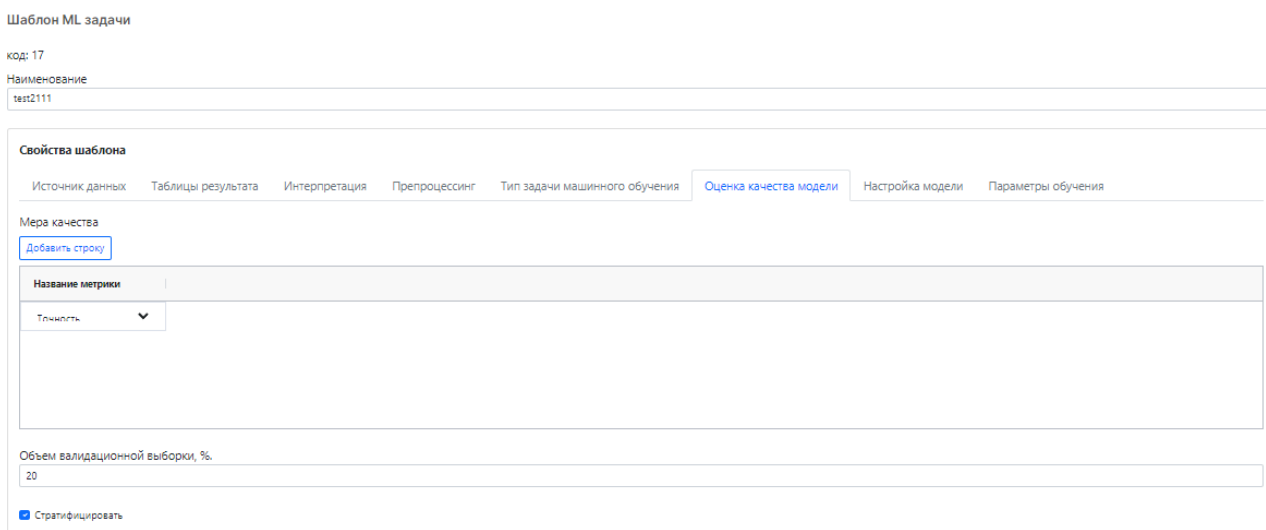
Сохранить Отменить

Рисунок 172. Выбор типа задачи машинного обучения

6. «Оценка качества модели»

Для определения метрики, которая будет отражать качество обучения необходимо «добавить строку» и выбрать метрику «точность».

Далее нужно указать объем валидационной выборки (применимый стандарт 20-25).



Шаблон ML задачи
код: 17
Наименование
test2111

Свойства шаблона

Источник данных Таблицы результата Интерпретация Преобработка Тип задачи машинного обучения **Оценка качества модели** Настройка модели Параметры обучения

Мера качества

Добавить строку

Название метрики
Точность

Объем валидационной выборки, %
20

Стратифицировать

Рисунок 173. Оценка качества модели

7. «Настройка модели»

На данной вкладке выбирается преднастроенная архитектура нейронной сети или «Без шаблона» для самостоятельного построения сети.

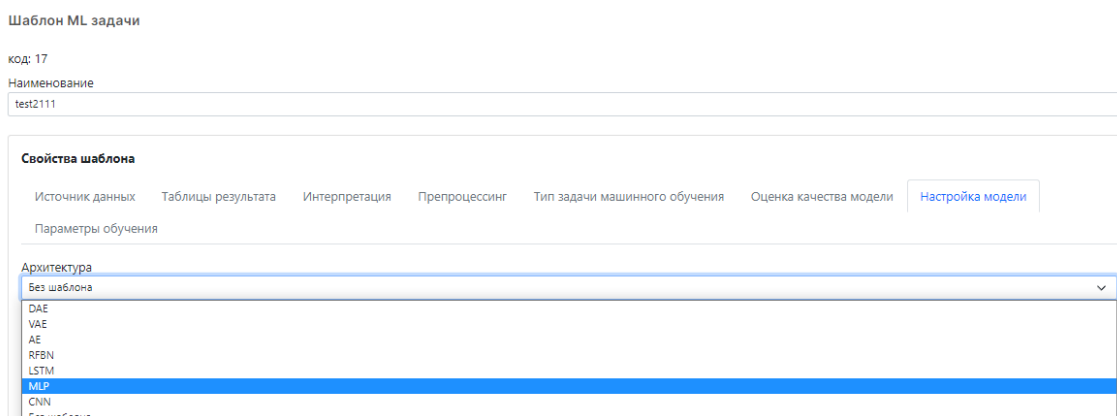


Рисунок 174. Выбор архитектуры нейронной сети

Для выстраивания архитектуры прописываются необходимые слои, устанавливается количество нейронов и функция активации для каждого слоя.

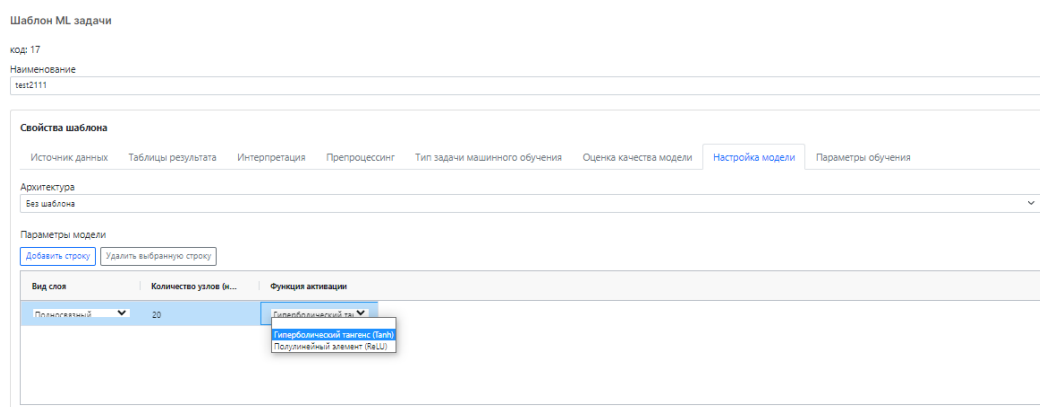


Рисунок 175. Настройка модели

8. «Параметры обучения»

В параметрах обучения указываются следующие пункты:

- Размер порции
- Функция потерь (кросс-энтропия, бинарная кросс-энтропия)
- Метод оптимизации (Adam)
- Шаг обучения (0.01)
- Количество эпох для ранней остановки
- Максимальное количество эпох при обучении

Шаблон ML задачи

код: 17

Наименование
test2111

Свойства шаблона

Источник данных Таблицы результата Интерпретация Препроцессинг Тип задачи машинного обучения Оценка качества модели Настройка модели **Параметры обучения**

Размер порции
1024

Функция потерь
Кросс-энтропия

Метод оптимизации
'Adam' - метод адаптивной оценки момента

Шаг обучения
0.01

Количество эпох для ранней остановки обучения
20

Максимальное количество эпох при обучении
100

Рисунок 176. Параметры обучения

10.2 Шаг 2. Запуск ML-задачи

Следующим шагом нужно создать новую задачу во вкладке «Предиктивная аналитика» - «ML задания».

Задания ML

Информация о задаче

Код: 28

Наименование
test2111_1

Шаблон задачи ML
test2111

Дата создания: 2022-11-21 10:09:41.161316

Дата обновления: 2022-11-21 10:09:41.071511

Дата выполнения: 2022-11-21 07:09:55.284000

Ссылка на задачу ML: e95a3b7a-2db3-4b43-960d-a60e99ede7d2

Quartz:

Запланирована:

Есть ошибки:

Завершена: true

Спецификация задачи

```

Tree +
├── object (6)
│   ├── task (2)
│   ├── input (2)
│   ├── output (1)
│   ├── network (7)
│   ├── preprocess (2)
│   └── validation (2)

```

Содержимое статуса

```

Tree +
├── object (8)
│   ├── stage : finished
│   ├── status : finished
│   ├── is_done : true
│   ├── is_error : false
│   ├── progress : 1
│   ├── completed : 2022-11-21T07:09:55.284Z
│   ├── is_planned : false
│   └── is_running : false

```

Рисунок 177. Создание ML-задания

На рисунке 177 отображается окно создания ML-задания, где нужно задать наименование и выбрать шаблон задачи ML, на основе которого будет выполнено задание.

После сохранения задача попадет в журнал заданий. Для того, чтобы провести задачу, необходимо в верхнем меню нажать кнопку «Запустить», далее «Обновить данные по задаче».

На рисунке 178 отображается полный список созданных заданий ML.

Журнал ML заданий

[Новый](#)
[Изменить](#)
[Пометить на удаление](#)
[Обновить данные по задаче](#)
[Запустить](#)
[Запланировать через Quartz](#)
[Расписание задач ML](#)

[Обновить](#)

К...	Дата создания	Дата обновления	Дата выполнения	Имя шаблона ML	Имя задачи	Ссылка на задачу ML	Quartz	Запланирована	Есть ошибки
29	27.01.2023 10:26:34	27.01.2023 10:26:34		ttt	ttt				
28	21.11.2022 10:09:41	21.11.2022 10:09:41	21.11.2022 07:09:55	test2111	test2111_1	e95a3b7a-2db3-4b43-960d-e80e99ede7c2			
27	21.11.2022 10:07:34	21.11.2022 10:07:34	21.11.2022 07:09:22	test2111	test2111	e519c0bf-0582-4fab-b2a4-757360aace16			
26	18.11.2022 11:12:52	18.11.2022 11:12:52	18.11.2022 08:12:58	test1811	test1811	588006d0-fc77-4335-b7c3-5e5097c7b94e			
25	14.11.2022 13:25:16	14.11.2022 13:25:16		t_141122	t_141122	24308e12-d50e-441c-aacc-5fac00fab2e			
24	07.11.2022 15:02:37	07.11.2022 15:02:37		test071122	test071122	91501df3-8652-4ca9-9a5c-f06cb024540			
23	03.11.2022 14:37:41	03.11.2022 14:37:41	03.11.2022 11:37:45	tt	tt	9cd874e6-9d8e-4290-bd6c-e805e4b12345			
22	03.11.2022 10:43:56	03.11.2022 10:43:56		t1	t1	01f4dccc-af66-4df9-83aa-cb45191446f7			
21	01.11.2022 11:30:04	01.11.2022 11:30:04	01.11.2022 08:30:11	test_01	test_01	0b6a6a29-4eb4-437a-8462-a1496968005d			
20	01.11.2022 10:52:17	01.11.2022 10:52:17		test_from_nsu_3	test_from_nsu_3_AE	dfc434b2-bb8f-4e45-84d3-6e5280e5569e			
19	01.11.2022 10:04:12	01.11.2022 10:04:12	01.11.2022 07:04:16	test_train_report_02	test_train_report_02	4f142440-caf1-4bcb-a1cd-0a039490a5e			
18	01.11.2022 09:58:45	01.11.2022 09:58:45	01.11.2022 08:58:49	test_train_report_01	test_train_report_01	64dcb237-df5c-46bc-a761-ba3f9f5ba7aa			
17	01.11.2022 08:46:21	01.11.2022 08:46:21		test_from_nsu_2	test_from_nsu_2	4f06c8a7-6009-475e-8a5d-ee07d027218			

Рисунок 178. Журнал заданий

10.3 Шаг 3. Мониторинг

После создания и проведения задачи на дашборде отображаются результаты обучения настроенной нейронной сети на выбранном наборе данных.

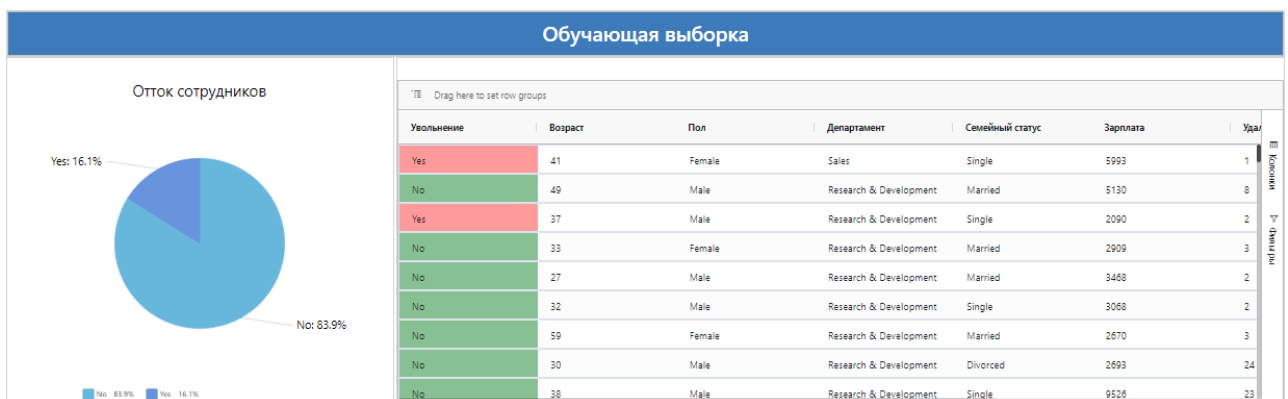


Рисунок 179. Данные выбранной для эксперимента таблицы

На рисунке 179 в левой части экрана расположена круговая диаграмма, на которой отображается отношение уволенных/не уволенных сотрудников. В правой части экрана таблица, которая будет использоваться для выполнения ML-задачи.

Для того, чтобы просмотреть данные по конкретной ML-задаче необходимо выбрать нужный эксперимент в таблице «Обучения модели», после чего на графиках и таблицах будут отображаться данные по выбранному эксперименту.

На рисунке 180 в правой части экрана отображается таблица с результатами тренировки (Точность, функция потерь на валидации, функция потерь на

обучении), а в нижней левой части экрана линейная диаграмма, отображающая значения из таблицы.

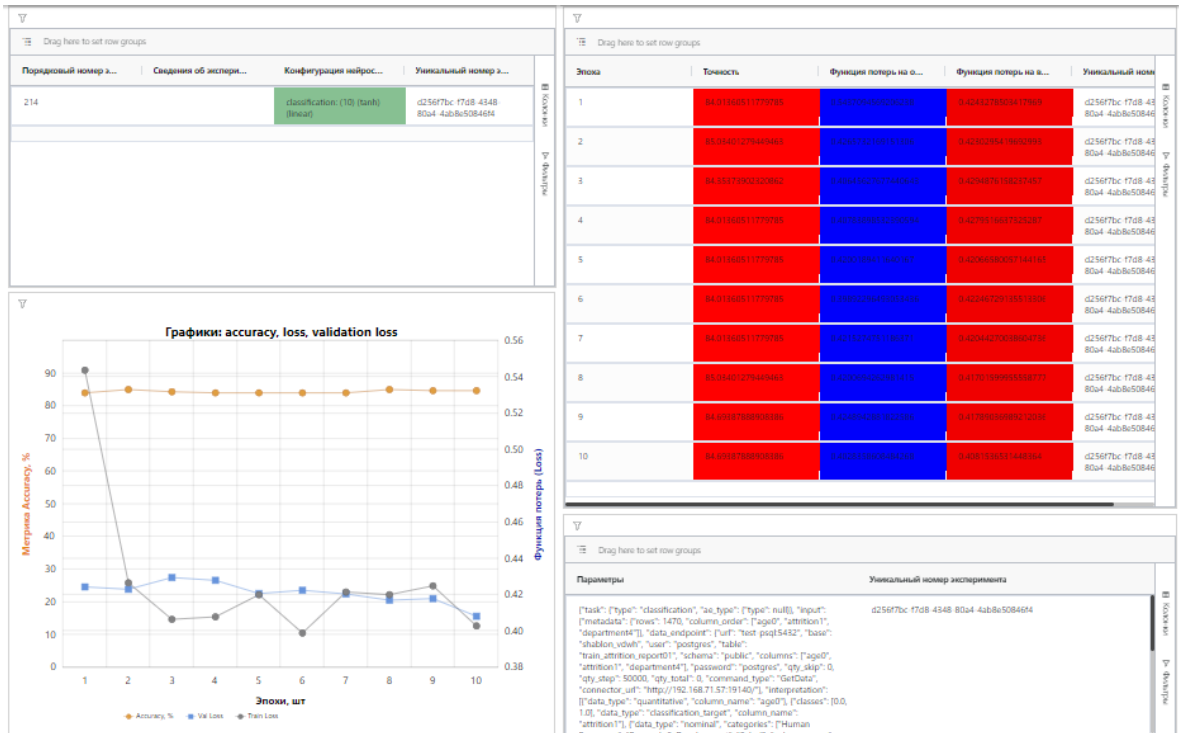


Рисунок 180. Дашборд результатов обучения

11. Заключение

В данном документе были указаны основные понятия, используемые в системе. Был описан порядок входа в систему, а также механизм восстановления пароля.

Подробно описано, как добавлять пользователей и назначать им роли, как создавать приложения, дашборды, графики, в том числе описан механизм редактирования графиков исходя из нужд пользователя. Уделено внимание кросс-фильтрации, глобальным фильтрам, безопасности и конструктору нейросетей.