

**ПЛАТФОРМА ДАННЫХ СЕЛЕНА**  
**РУКОВОДСТВО ПОЛЬЗОВАТЕЛЯ**  
**РФ.DIS.00011-01 41**

## АННОТАЦИЯ

Документ РФ.DIS.00011-01 41 «Руководство пользователя» подготовлен на основе рекомендаций государственного стандарта ГОСТ Р 59795—2021 Информационные технологии. Комплекс стандартов на автоматизированные системы. Автоматизированные системы. Требования к содержанию документов.

Электронная версия документа хранится в составе пакета программной и эксплуатационной документации на изделие РФ.DIS.00011 «Платформа данных СЕЛЕНА» (далее по тексту Платформа данных СЕЛЕНА).

Ознакомление с документом «Руководство пользователя» персонала подразделения, принимающего участие в работе автоматизированной системы, производится под роспись с внесением соответствующей записи в журнал первичного инструктажа.

## СОДЕРЖАНИЕ

Аннотация .....	2
Обозначения и сокращения .....	6
Предисловие .....	7
1 Обзор Платформы Данных .....	8
1.1 Обзор продукта .....	8
1.2 Архитектура и компоненты репликации данных .....	10
2 Модуль управления кластером Cluster Manager .....	14
2.2 Вход в систему .....	14
2.3 Установка и обновление лицензии .....	14
2.4 Смена языка .....	14
2.5 Смена пароля учётной записи .....	14
2.6 Добавление виртуальных машин.....	14
2.7 Добавление фронтенд (мастер) узлов .....	15
2.8 Добавление бэкенд (вычислительных) узлов .....	15
2.9 Добавление узлов хранилища .....	16
2.10 Добавление узлов каталога .....	16
2.11 Удаление (декомиссия) узлов .....	16
2.12 Конфигурирование параметров фронтенд-нод .....	17
2.13 Конфигурирование параметров кластера .....	17
2.14 Создание пользователей .....	17
2.15 Изменение пользователя .....	18
2.16 Удаление пользователя .....	18
2.17 Создание роли .....	18
2.18 Изменение роли — добавление привилегий .....	18
2.19 Изменение роли — удаление привилегий.....	19
2.20 Удаление роли .....	19
2.21 Конфигурация LDAP .....	19
2.22 Мониторинг .....	19
3 Управление загрузкой данных, расписанием и задачами .....	21
3.1 Вход в систему .....	21
3.2 Просмотр общей статистики по выполняемым процессам.....	21
3.3 Создание подключений .....	22
3.4 Создание проектов.....	22

3.5	Создание и настройка процессов в проектах .....	22
3.6	Запуск и настройка расписания в процессах .....	23
3.7	Мониторинг и отладка выполняемых процессов .....	24
3.8	Мониторинг состояния сервиса .....	25
3.9	Создание и изменение пользователей .....	25
3.10	Создание и изменение групп пользователей.....	25
3.11	Настройка оповещений.....	26
4	Работа с модулем среды разработки .....	27
4.1	Описание интерфейса .....	27
4.2	Вход в систему .....	27
4.3	Смена языка интерфейса.....	28
4.4	Выход из системы .....	28
4.5	Интерфейс работы с запросами .....	28
4.6	Выполнение Ad-hoc запросов .....	29
4.7	Выполнение Ad-hoc запросов с помощью AI .....	29
4.8	Выполнение поиска.....	29
4.9	Просмотр состава таблиц .....	29
4.10	Просмотр и поиск сниппетов .....	30
4.11	Создание сниппетов .....	30
4.12	Изменение сниппетов .....	31
4.13	Удаление сниппетов .....	31
4.14	Просмотр блокнотов .....	31
4.15	Создание блокнота .....	32
4.16	Создание блокнота с помощью AI .....	32
4.17	Типы блоков в блокноте .....	32
4.18	Управление блоками .....	34
4.19	Добавление переменной в блокноте .....	34
4.20	Дополнительные возможности .....	34
4.21	Работа с ИИ-агентом .....	35
4.22	Интерфейс работы с ИИ-агентом .....	35
4.23	Выполнение запроса через ИИ-агента.....	35
4.24	Начало нового чата.....	36
4.25	Работа с историей общения.....	36
5	Подключение к системе из внешних инструментов .....	37

5.1	Подключение к платформе данных из Apache Superset .....	37
5.2	Подключение к платформе данных из Tableau Desktop .....	38
5.3	Подключение к платформе данных из DataGrip .....	38
5.4	Подключение к платформе данных из DBeaver.....	39
5.5	Подключение к платформе данных из Jupyter.....	40

## ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ

В документе РФ.DIS.00011.0141 «Руководство пользователя» используются следующие обозначения и сокращения, имеющие соответствующие значения:

ОБОЗНАЧЕНИЕ ИЛИ СОКРАЩЕНИЕ	ЗНАЧЕНИЕ
ВМ	Виртуальная машина
ГОСТ	Государственный стандарт
РФ	Российская Федерация
СУБД	Система управления базами данных
ПКМ	Правая кнопка мыши
УЗ	Учетная запись
AI	Искусственный интеллект
BI	Business intelligence. Система сбора, обработки и анализа данных
CPU	Центральный процессор
DDL	Data Definition Language (DDL) (язык описания данных)
DS	Data Science
ETL	Extract, Transform, Load
HDFS	Hadoop Distributed File System
JDBC	Java DataBase Connectivity
ML	Machine learning
ODBC	Open Database Connectivity
OLAP	Online analytical processing
MPP	Massively parallel processing
SIMD	Single instruction, multiple data
S3	Simple Storage Service
SQL	Structured Query Language — «язык структурированных запросов»
URI	Uniform Resource Identifier
URL	Uniform Resource Locator

## Предисловие

Данное руководство пользователя описывает программный комплекс «Платформа хранения и обработки данных Селена» (в дальнейшем Селена), включая его архитектуру, варианты использования, принципы развертывания, источники и целевые объекты, а также основные концепции.

В руководстве также описывается, как использовать интерфейс панели управления для настройки, запуска, мониторинга и администрирования.

## **1 ОБЗОР ПЛАТФОРМЫ ДАННЫХ**

### **1.1 Обзор продукта**

Селена – сверхбыстрая платформа хранения и обработки данных нового поколения с массивно-параллельной обработкой (MPP), разработанная для упрощения и ускорения доступа к данным, быстрой аналитики в реальном времени. MPP и векторизованный механизм выполнения вычислений позволяют пользователям выбирать между различными схемами для разработки многомерных аналитических отчетов.

Платформа Селена предназначена для применения для следующих случаев:

- многомерная аналитика OLAP;
- построение отчетности любого уровня сложности;
- аналитика данных в реальном времени;
- high-concurrency аналитика;
- унифицированная аналитика.

Селена поддерживает самые разнообразные функции, обеспечивая надежную и быструю работу вашего корпоративного хранилища данных, что указано в списке ниже.

- а) MPP-фреймворк: платформа Селена использует фреймворк массивно-параллельной обработки. Один запрос разделяется на несколько физических вычислительных блоков, которые могут выполняться параллельно на нескольких машинах. Каждая машина имеет выделенные ресурсы CPU и памяти.
- б) Оптимизатор: он находит наиболее оптимальный план на основе собранной статистики ваших данных. Это ключ к лучшей в своем классе производительности запросов, особенно для мультитабличных запросов.
- в) Полностью векторизованный механизм выполнения: благодаря колоночному механизму хранения и полностью векторизованным операторам платформа Селена в полной мере использует современные многоядерные процессоры и инструкции SIMD для повышения производительности.
- г) Гибридное строково-столбцовое хранилище: оно обеспечивает более 10 000 запросов в секунду уже на 16-ядерных экземплярах вычислительных серверов



(нод кластера) за счет оптимизированных точечных запросов и первичного ключевого индексируемого ускорения в гибридном строково-столбцовом хранилище.

- д) Кэш данных: встроенная в платформу Селена структура кэширования на основе памяти и дискового пространства специально разработана для минимизации накладных расходов ввода-вывода при извлечении данных из внешнего хранилища для ускорения выполнения запросов.
- е) Аналитика в реальном времени: от потоковой передачи до сбора данных с богатым набором коннекторов возможно загружать данные в платформу Селена в реальном времени для получения самых свежих инсайтов.
- ж) Таблица первичного ключа обеспечивает непревзойденную производительность запросов с обновлениями, вставками и удалениями в реальном времени. Индекс первичного ключа позволяет эффективно разрешать изменения данных во время приема данных, оптимизируя производительность чтения, поддерживая актуальность данных на уровне менее десяти секунд от времени изменения данных.
- з) Синхронное материализованное представление может постепенно обновляться при приеме данных и выполнять переписывание запросов при их исполнении.
- и) Общая архитектура данных - Селена разделяет слои хранения и вычисления посредством сохранения данных в удаленном объектном хранилище, таком как S3 или HDFS.
- к) Единый каталог метаданных: с помощью одной команды при использовании встроенного каталога метаданных платформа Селена позволяет вам легко подключаться и напрямую запрашивать самые свежие данные, хранящиеся во всех озерах и других источниках данных.
- л) Асинхронное материализованное представление разработано для ускорения медленных запросов без каких-либо дополнительных внешних инструментов обработки. Платформа Селена обеспечивает возможность перезаписи запросов и позволяет создавать асинхронное материализованное представление в любое время без необходимости вручную изменять SQL-запрос.

- м) Унификация доступа к данным: платформа Селена поддерживает синтаксис ANSI SQL, протокол MySQL и обеспечивает поддержку диалекта Trino/Presto. Платформа совместима с широким спектром клиентского ПО и инструментов BI и аналитики.

## 1.2 Архитектура и компоненты репликации данных

Платформа Данных Селена включает следующие компоненты:

- а) Мастер сервер
- б) Вычислительный сервер
- в) Панель управления кластером (Cluster Manager)
- г) Модуль загрузки данных
- д) Метакаталог данных
- е) Мониторинг
- ж) Модуль среды разработки
- з) Модуль хранения данных

На рис. 1 показана архитектура Платформы данных Селена.

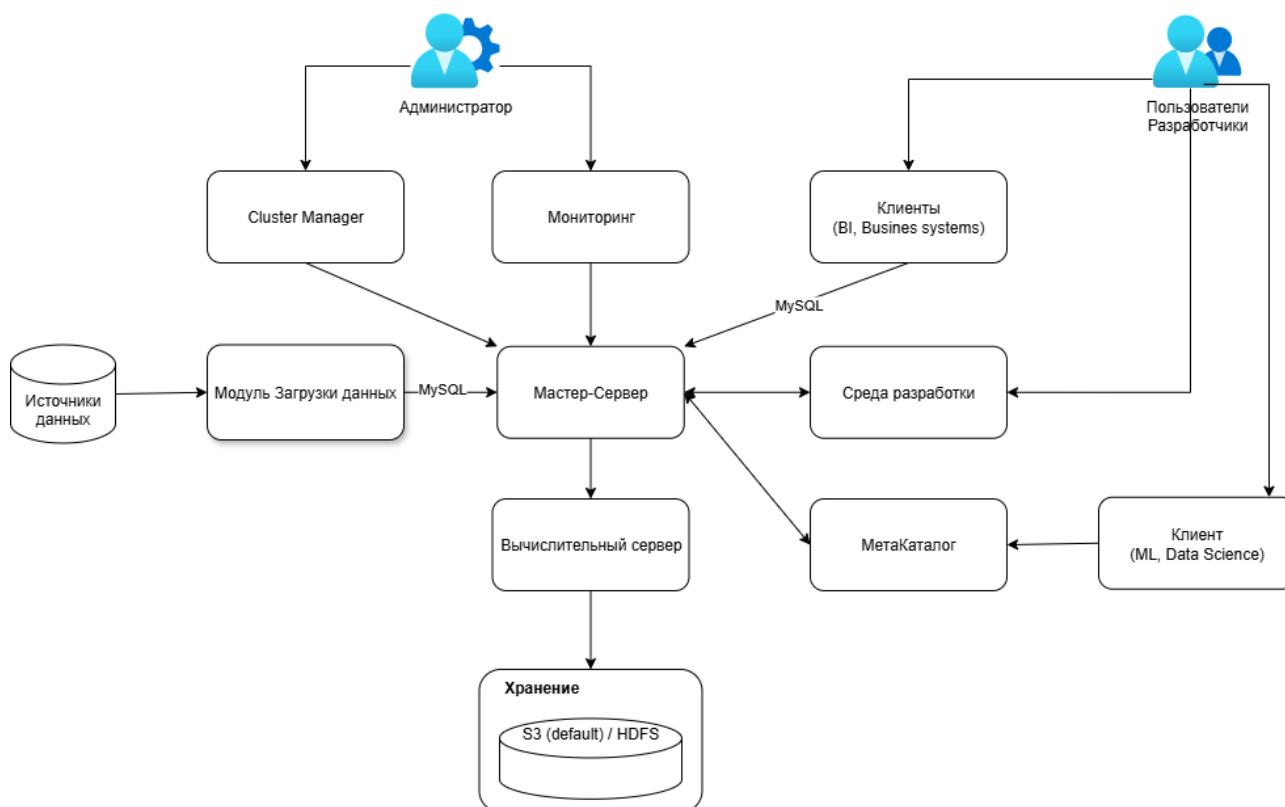


Рисунок 1 - Архитектура и компоненты Платформы Данных Селена

## Мастер сервер

Мастер сервер отвечают за управление метаданными, управление клиентскими соединениями, планирование запросов и диспетчеризацию запросов. Каждый Мастер сервер хранит и поддерживает полную копию метаданных в своей памяти, что гарантирует неизбирательное обслуживание среди Мастер серверов. Мастер сервер может выполнять одну из ролей - лидер, последователь и наблюдатель.

Лидер считывает и записывает метаданные. Лидер обновляет метаданные, а затем синхронизирует изменения метаданных с последователями и наблюдателями. Записи данных считаются успешными только после того, как изменения метаданных синхронизированы с более чем половиной последователей.

Последователи синхронизируют и воспроизводят журналы с лидера для обновления метаданных.

## Вычислительный сервер

Основной задачей Вычислительных серверов является выполнение SQL запросов. Мастер сервер разбирают каждый SQL-запрос на логический план выполнения в соответствии с семантикой запроса, а затем преобразуют логический план в физические планы выполнения, которые могут быть выполнены на Вычислительных серверах. Вычислительные сервера, хранящие данные назначения, выполняют запрос. Это устраняет необходимость в передаче и копировании данных, достигая высокой производительности запроса.

## Панель управления кластером (Cluster Manager)

Cluster Manager - основной инструмент управления кластером, обеспечивает администраторам все необходимые функции по созданию и настройке кластера. Обеспечивает функции добавления и удаления нод, назначение типов и развертывание приложений на новых нодах. Помимо этого, выполняет функции безопасности, обеспечивая контроль доступа как к самим интерфейсам, используемым в платформе компонентов, так и к данным, расположенным в хранилище.

### Модуль загрузки данных

Модуль загрузки данных обеспечивает бесперебойную загрузку данных из различных систем источников данных в целевое хранилище. Компонент обеспечивает возможность извлекать данные из самых разнообразных систем включая самые популярные реляционные и No-SQL базы данных, шины данных и т.д. Кроме задач по загрузке данных, компонент реализует функции визуального формирования потоков данных, и настройки расписания запусков, а также выполнять функции отладки и мониторинга выполнения процессов работы с данными.

### Метакаталог данных

Метакаталог данных содержит информацию о метаданных, хранящихся в объектном хранилище, и обеспечивает быстрое формирование запросов на данные в объектном хранилище. Каталог метаданных в архитектуре платформы Селена обеспечивает единое представление о хранящихся в объектном хранилище данных.

### Мониторинг

Мониторинг в системе обеспечивается выделенным модулем, который собирает всю необходимую информацию о работе кластера, состоянии основных компонентов и вычислительных мощностях. Пользователям предоставляется целый ряд пред настроенных мониторов отображая информацию от загрузки ЦПУ на нодах мастер серверов, заканчивая производительностью и количеством выполняемых запросов на вычислительных узлах кластера.

### Модуль среды разработки

Модуль среды разработки позволяет различным бизнес-пользователям быстро и просто получить данные. Для разработчиков и аналитиков обеспечивается возможность создания ad-hoc запросов и формирование блокнотов. А для бизнес-пользователей для работы с данным компонентом нет необходимости уметь писать sql-запросы. Все обращения в систему формируются используя семантику, далее они автоматически преобразуются в SQL код, который в свою очередь и выполняется, возвращая ожидаемый результат пользователю.

### Модуль хранения данных

В качестве компонента, отвечающего за непосредственное хранение данных, могут выступать два решения для хранения: объектное хранилище S3 и HDFS. Платформа хранения и обработки данных Селена поставляется вместе с S3 на базе Minio, но при этом может быть использовано любое клиентское S3 Compatible решение, а также HDFS для реализации хранения данных. Платформа обеспечивает возможность записывать и хранить данные в различных форматах, в зависимости от использования внешнего метакаталлога данные будут храниться в открытом формате, либо в нативном.

## **2 МОДУЛЬ УПРАВЛЕНИЯ КЛАСТЕРОМ CLUSTER MANAGER**

### **2.2 Вход в систему**

- а) Откройте указанный URL.
- б) Нажмите Login.
- в) Введите логин и пароль.
- г) Нажмите Sign In.

### **2.3 Установка и обновление лицензии**

- а) Войдите в Cluster Manager.
- б) В навигационной панели нажмите кнопку Лицензия.
- в) Нажмите. Обновить лицензию.
- г) В открывшемся окне выберите лицензионный файл.
- д) Нажмите. Открыть

### **2.4 Смена языка**

- а) Войдите в Cluster Manager.
- б) В навигационной панели нажмите Язык / Language.
- в) Выберите Русский или English.

### **2.5 Смена пароля учётной записи**


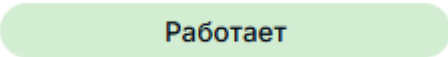
- а) Войдите в Cluster Manager.
- б) В навигационной панели нажмите Профиль.
- в) В открывшемся окне укажите:
  - Текущий пароль;
  - Новый пароль;
  - Повторите новый пароль.
- г) Нажмите Сохранить.

### **2.6 Добавление виртуальных машин**

- а) Войдите в Cluster Manager.
- б) В навигационной панели нажмите Виртуальные машины.
- в) Нажмите Новая виртуальная машина.

- г) Укажите:
- Название ВМ;
  - Публичный IP;
  - Внутренний IP;
  - Имя пользователя.
- д) Выберите тип аутентификации:
- Пароль — укажите пароль пользователя.
  - SSH-ключ — нажмите. Выберите файл, укажите файл с ключом, нажмите Открыть.
- е) Для проверки корректности подключения нажмите. Проверить доступ по SSH.
- ж) Нажмите Сохранить.

## 2.7 Добавление фронтенд (мастер) узлов

- а) Войдите в Cluster Manager.
- б) Нажмите Управление кластером.
- в) В разделе Фронтенды нажмите кнопку  .
- г) В выпадающем списке Роль укажите:
- Master — главный узел кластера (только один);
  - Follower — вспомогательный узел, участвует в выборе нового Master;
  - Observer — вспомогательный узел, не участвует в выборе нового Master.
- д) Выберите виртуальную машину для установки.
- е) Нажмите Сохранить.
- ж) Для отслеживания установки нажмите значок файла.
- з) Дождитесь, пока статус  будет установлен для Master.

## 2.8 Добавление бэкенд (вычислительных) узлов

- а) Войдите в Cluster Manager.
- б) Нажмите Управление кластером.



- в) В разделе Бэкенды нажмите значок .
- г) Выберите виртуальную машину для установки.
- д) Нажмите Сохранить.
- е) Для отслеживания установки нажмите значок файла.

## 2.9 Добавление узлов хранилища

- а) Войдите в Cluster Manager.
- б) Нажмите Управление кластером.



- в) В разделе Хранилище нажмите значок .
- г) Выберите виртуальную машину для установки.
- д) Нажмите Сохранить.
- е) Для отслеживания установки нажмите значок файла.

## 2.10 Добавление узлов каталога

- а) Войдите в Cluster Manager.
- б) Нажмите Управление кластером.





- в) В разделе Каталоги нажмите значок .
- г) Выберите виртуальную машину для установки.
- д) Нажмите Сохранить.
- е) Для отслеживания установки нажмите значок файла.

## 2.11 Удаление (декомиссия) узлов


- а) Войдите в Cluster Manager.
- б) Нажмите Управление кластером.
- в) В строке узла нажмите кнопку с изображением корзины.
- г) Подтвердите удаление.
- д) Дождитесь завершения процесса.



## 2.12 Конфигурирование параметров фронтенд-нод

- а) Войдите в Cluster Manager.
- б) Нажмите Управление кластером.
- в) В строке Фронтенд нажмите кнопку .
- г) Выберите параметр из списка (поиск доступен через поле ввода + Enter).
- д) Нажмите кнопку .
- е) Введите новое значение.
- ж) Нажмите Сохранить.

## 2.13 Конфигурирование параметров кластера

- а) Войдите в Cluster Manager.
- б) Нажмите Конфигурации.
- в) Выберите параметр из списка (поиск доступен через поле ввода + Enter).
- г) Нажмите кнопку .
- д) Введите новое значение.
- е) Нажмите Сохранить.

## 2.14 Создание пользователей

- ж) Войдите в Cluster Manager.
- з) В навигационном меню нажмите Пользователи.
- и) Нажмите Создание и удаление пользователей.
- к) Выберите User Management.
- л) Нажмите Add User.
- м) Укажите:
  - Логин;
  - Отображаемое имя.
- н) При необходимости активируйте флаг Blocked.
- о) При необходимости активируйте флаг User Management.
- п) Нажмите Save and Exit.

## 2.15 Изменение пользователя

- а) Войдите в Cluster Manager.
- б) В навигационном меню нажмите Пользователи.
- в) Нажмите Создание и удаление пользователей.
- г) Выберите User Management.
- д) В строке пользователя нажмите карандаш.
- е) Измените данные.
- ж) Нажмите Save and Exit.


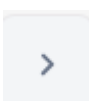

## 2.16 Удаление пользователя

- з) Войдите в Cluster Manager.
- и) В навигационном меню нажмите Пользователи.
- к) Нажмите Создание и удаление пользователей.
- л) Выберите User Management.
- м) В строке пользователя нажмите корзину.
- н) Подтвердите удаление.




## 2.17 Создание роли

- о) Войдите в Cluster Manager.
- п) Нажмите Роли.
- р) Нажмите Новая роль.
- с) Задайте имя.
- т) Нажмите Сохранить.


## 2.18 Изменение роли — добавление привилегий

- а) Войдите в Cluster Manager.
- б) Нажмите Роли.
- в) В строке роли нажмите  → Изменить.
- г) Выберите объекты (каталоги, хранилища данных, представления, системные).
- д) Выберите привилегию.
- е) Нажмите  или  для добавления.

## 2.19 Изменение роли — удаление привилегий

- а) Войдите в Cluster Manager.
- б) Нажмите Роли.
- в) В строке роли нажмите  → Изменить.
- г) Выберите объект.
- д) Выберите привилегию.
- е) Нажмите  или  для удаления.

## 2.20 Удаление роли

- а) Войдите в Cluster Manager.
- б) Нажмите Роли.
- в) В строке роли нажмите  → Удалить.
- г) Подтвердите удаление.

## 2.21 Конфигурация LDAP

- а) Войдите в Cluster Manager.
- б) Нажмите Пользователи.
- в) Нажмите LDAP.
- г) Укажите:
  - Хост LDAP;
  - Порт LDAP;
  - Base DN;
  - Admin Bind DN;
  - Пароль;
  - Фильтр поиска пользователей.
- д) Нажмите Сохранить.

## 2.22 Мониторинг

Для контроля за состоянием кластера:

- а) Войдите в Cluster Manager.

- б) В навигационном меню нажмите "Мониторинг".
- в) В открывшемся окне появится Dashboard.
- г) Метрики сгруппированы по типам:
  - Overview — общее состояние кластера, ключевые показатели производительности и доступности.
  - Cluster Overview — сводная информация по узлам кластера, их статусам и нагрузке.
  - Query Statistic — статистика выполнения запросов, включая время выполнения, количество запросов и ошибки.
  - Resource Group — использование ресурсов по группам, распределение нагрузки между ними.
  - Jobs — состояние фоновых и плановых заданий, время их выполнения и статусы.
  - Transaction — показатели работы транзакций: количество, длительность, ошибки.
  - FE JVM — метрики работы JVM фронтенд-узлов (память, сборка мусора, потоки).
  - BE — общие показатели работы backend-узлов (нагрузка CPU, состояние служб).
  - BE Memory — использование оперативной памяти backend-узлами.
  - BE Tasks — статистика выполнения задач на backend-узлах.
  - BE Load — показатели загрузки backend-узлов, распределение вычислительных ресурсов.

Раскрывая каждую группу, можно просмотреть соответствующие графики и показатели.

### 3 УПРАВЛЕНИЕ ЗАГРУЗКОЙ ДАННЫХ, РАСПИСАНИЕМ И ЗАДАЧАМИ

#### 3.1 Вход в систему

- а) Открыть страницу системы оркестрации и управления потоками данных
- б) Ввести логин и пароль
- в) Нажать «Войти»

#### 3.2 Просмотр общей статистики по выполняемым процессам

- а) Авторизоваться в системе
- б) В левой части экрана нажать «Дашборд»
- в) В появившемся окне будет отображена информация по выполняемым в системе процессам и задачам.
- г) Для того, чтобы отфильтровать процессы и задачи по временному диапазону необходимо задать соответствующие даты начала и конца в соответствующем полях над выводимой информацией
- д) Для того, чтобы отфильтровать процессы с не интересующим вас состоянием необходимо в центральной части нажать по показателю статуса, который следует убрать или добавить

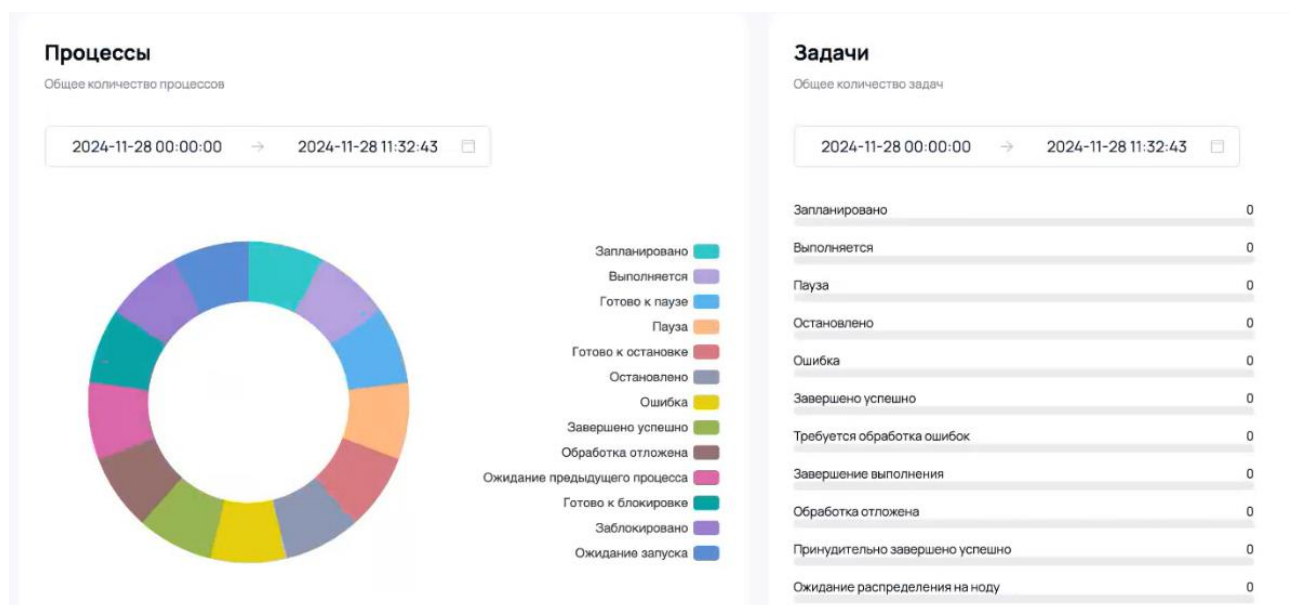


Рисунок 2

### 3.3 Создание подключений

- е) Авторизоваться в системе
- ж) В левой части экрана нажать «Подключения»
- з) В появившемся окне нажать кнопку «новое подключение»
- и) Выбрать из списка подключений необходимую СУБД или источник данных
- к) Указать требуемые настройки в зависимости от системы
- л) Нажать «тест соединения» для проверки корректности введенной информации
- м) Нажмите «сохранить»
- н) Если все указано корректно, то схема данных взятая из добавленного источника автоматически отобразиться в настройках при создании процессов

### 3.4 Создание проектов



- а) Авторизоваться в системе
- б) В левой части экрана нажать «Проекты»
- в) В появившемся окне нажать кнопку «новый проект»
- г) Указать название, владельца и описание, нажать «подтвердить»
- д) Открыть проект нажав по нему
- е) В открывшемся окне, на вкладке Информация собрана общая информация о работе данного проекта, включая все запускаемые процессы и задачи в нем. Настройка фильтрация по времени или состоянию задач осуществляется согласно разделу 3.2 пунктам 4 и 5.

### 3.5 Создание и настройка процессов в проектах

- а) Авторизоваться в системе
- б) В левой части экрана нажать «Проекты»
- в) Открыть необходимый проект нажав по нему
- г) Нажать вкладку «процессы»
- д) Нажать «новый процесс»
- е) В открывшемся окне используя функции drag and drop. Перетащите из левой части экрана задачу, на область для проектирования.
- ж) При помещении задачи на область для проектирования автоматически откроется всплывающее окно с настройками. В нем необходимо задать:

- Название задачи
  - Выполняемые код, запрос или действие в зависимости от выбранной задачи.
  - При необходимости развернуть раздел «дополнительные настройки» и указать необходимые дополнительные параметры
- з) Для соединения задач в единый поток, нажмите ЛКМ на «кружке» одной из них и перетащите появившуюся стрелку к другой задаче
- и) Для редактирования ранее добавленной задачи, нажмите на задачу правой кнопкой мыши, в контекстном меню нажмите «редактировать», измените настройки, нажмите «сохранить».
- к) Для копирования ранее добавленной задачи, нажмите на задачу правой кнопкой мыши, в контекстном меню нажмите «копировать», дубликат задачи автоматически создастся рядом с копируемой задачей.
- л) Для удаления ранее добавленной задачи, нажмите на задачу правой кнопкой мыши, в контекстном меню нажмите «удалить»
- м) По окончании создания потока действий нажмите «сохранить» в правом нижнем углу области проектирования.

### 3.6 Запуск и настройка расписания в процессах

- а) Авторизоваться в системе
- б) В левой части экрана нажать «Проекты»
- в) Открыть необходимый проект нажав по нему
- г) Нажать вкладку «процессы»
- д) Для единоразового запуска процесса, напротив требуемого процесса нажать кнопку редактирования , в контекстном меню нажать «запустить»
- е) Для настройки запуска задачи по расписанию, напротив требуемого процесса нажать кнопку редактирования , в контекстном меню нажать «перевести в онлайн», подтвердить свой выбор нажав «подтвердить».
- ж) В появившемся окне задать настройки выполнения по расписанию:
- Время начала и окончания
  - расписание выполнения

- действие при ошибке
- стратегии уведомления
- приоритет процесса
- группу выполняемых серверов
- группу пользователей

з) Нажать «Подтвердить»

### 3.7 Мониторинг и отладка выполняемых процессов



- а) Авторизоваться в системе
- б) В левой части экрана нажать «Проекты»
- в) Открыть необходимый проект нажав по нему
- г) Нажать вкладку «инстансы процессов»
- д) В открывшемся окне будет отображена таблица со всеми запусками процессов в данном проекте.
- е) По каждому запуску фиксируется следующая информация:
  - Название процесса
  - Тип запуска
  - Статус
  - Расписание
  - Время начала и окончания
  - Продолжительность
  - Повторы
- ж) При нажатии на интересующий нас процесс откроется область проектирования. Нажав на требуемую задачу, откроется контекстное меню, в котором можно выбрать одно из следующих действий для отладки процесса
  - Редактирование
  - Копирование
  - Удаление
  - Просмотр лога
  - Очистка кэша
  - Запуск
  - Запуск в обратном порядке
  - Запуск в прямом порядке



### 3.8 Мониторинг состояния сервиса


- а) Авторизоваться в системе
- б) В левой части экрана нажать «мониторинг»
- в) В открывшемся окне нажать на название Ноды сервиса
- г) В открывшемся окне будет указана сводная информация о работе сервера и создаваемой нагрузке

### 3.9 Создание и изменение пользователей

- а) Авторизоваться в системе
- б) В левой части экрана раскрыть «управление» нажать «Безопасность»
- в) В открывшемся окне в вкладке пользователи нажать на кнопку «создать пользователя»
- г) В открывшейся оснастке задать:
  - Имя
  - Пароль
  - Группа
  - Email
  - Телефон
  - Статус
- д) Нажать «подтвердить»
- е) Для изменения напротив требуемой УЗ нажать кнопку конфигурирования  , в контекстном меню выбрать «редактирование», после внесения изменений нажать подтвердить.
- ж) Для удаления требуемой УЗ нажать кнопку конфигурирования  , в контекстном меню выбрать «удалить», подтвердить свой выбор.

### 3.10 Создание и изменение групп пользователей

- а) Авторизоваться в системе
- б) В левой части экрана раскрыть «управление» нажать «Безопасность»
- в) В открывшемся окне нажать вкладку группы пользователей

- г) Нажать на кнопку «создать группу»
- д) В появившейся оснастке задать имя и описание, нажать подтвердить
- е) Напротив созданной группы нажать кнопку конфигурирования , в контекстном меню выбрать «редактирование», указать пользователей, входящих в группу, нажать «подтвердить»

### 3.11 Настройка оповещений

- а) Авторизоваться в системе
- б) В левой части экрана раскрыть «управление» нажать «Оповещение»
- в) В открывшемся окне нажать «создать оповещение»
- г) В открывшейся оснастке задать:
  - название оповещение
  - Глобальное оповещение
  - Тип предупреждения
  - Используемый плагин
- д) Нажать сохранить

## **4 РАБОТА С МОДУЛЕМ СРЕДЫ РАЗРАБОТКИ**

### **4.1 Описание интерфейса**

Интерфейс среды разработки реализован в виде веб-приложения и содержит следующие элементы:

- **Левая панель инструментов:**
  - Консоль — выполнение Ad-hoc запросов.
  - Поиск — форма поиска по различным объектам.
  - Расписание — список блокнотов, запланированных на запуск по расписанию.
  - ИИ агент — окно работы с искусственным интеллектом.
  - Язык — выбор языка интерфейса.
  - Настройки — конфигурация параметров и выход из системы.
- **Правая панель инструментов:**
  - Блокноты — создание и управление блокнотами.
  - Таблицы — просмотр объектов, доступных в хранилище данных.
  - Снимпеты — создание и сохранение пользовательских фрагментов кода.
  - Вып. запросы — история выполненных запросов и их результаты.
- **Центральная область:**
  - Область редактора — ввод и редактирование запросов.
  - Область результатов — отображение результатов выполнения запросов в табличной форме.

### **4.2 Вход в систему**

Процесс аутентификации пользователя в веб-интерфейсе подсистемы:

- а) Откройте указанный URL-адрес в браузере.
- б) Нажмите кнопку Login.
- в) Введите логин и пароль.
- г) Нажмите кнопку Sign In.

#### 4.3 Смена языка интерфейса

Изменение языка отображения элементов интерфейса:

- а) Войдите в подсистему среды разработки.
- б) В левой панели инструментов нажмите кнопку Язык.
- в) Выберите Русский или English.


#### 4.4 Выход из системы

Завершение сеанса работы пользователя:

- а) Войдите в подсистему среды разработки.
- б) В левой панели инструментов нажмите кнопку Настройки.
- в) Нажмите кнопку Выход.

#### 4.5 Интерфейс работы с запросами

Интерфейс состоит из следующих элементов:

- Область редактирования — ввод SQL-запросов или текстовых инструкций для ИИ.
- Область результатов — отображение данных, полученных в результате выполнения запроса.
- Дополнительные действия  :
  - 1. Форматировать — переформатирование запроса в удобный для чтения вид.
  - 2. Скрыть запрос — удаление текста запроса из области редактирования.
  - 3. Выполнение — запуск выполнения запроса.

В области результатов доступны:

- 1. Скрыть колонки — фильтрация отображения колонок результирующей таблицы.
- 2. Full Screen — разворачивание результирующей таблицы на весь экран.
- 3. Просмотр результатов нескольких запросов — переключение между результатами через выпадающее меню кнопки *Запрос 1 из N*.

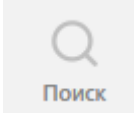
#### 4.6 Выполнение Ad-hoc запросов

- а) Войдите в подсистему среды разработки.
- б) В левой панели инструментов нажмите кнопку Консоль.
- в) В панели редактора введите необходимый SQL-запрос в поле запроса.
- г) Нажмите кнопку Выполнить.
- д) В области результатов отобразится результирующая таблица.

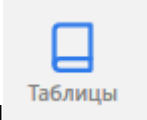
#### 4.7 Выполнение Ad-hoc запросов с помощью AI

- а) Войдите в подсистему среды разработки.
- б) В левой панели инструментов нажмите кнопку Консоль.
- в) В панели редактора введите текстовый запрос в поле работы с AI.
- г) Нажмите Enter.
- д) Подтвердите правильность созданного запроса, нажав Разрешить.
- е) Нажмите Выполнить.
- ж) В области результатов отобразится результирующая таблица.

#### 4.8 Выполнение поиска

- а) Войдите в подсистему среды разработки.
- б) В левой панели инструментов нажмите кнопку Поиск .
- в) В открывшемся окне выберите тип объекта: запрос, блокнот, таблицы, список.
- г) Введите интересующую информацию в поле поиска.
- д) При необходимости используйте фильтры в правой части окна для уточнения поиска.

#### 4.9 Просмотр состава таблиц

- а) Войдите в подсистему среды разработки.
- б) В правой панели инструментов нажмите кнопку Таблицы .
- в) В открывшейся оснастке разверните каталог.
- г) Разверните необходимую базу данных.

д) Нажмите на нужную таблицу.

е) Нажмите. Посмотреть таблицу





ж) В открывшейся оснастке переключайтесь по вкладкам Обзор, Колонки, Примеры строк, Исходный запрос, Предупреждения для просмотра информации.

з) Для работы с таблицей доступны действия:

- Добавить в избранное — нажмите кнопку Добавить в избранное




- Добавить теги — нажмите 

- Добавить себя как владельца — нажмите  напротив поля Владелец.

- Редактировать описание — нажмите. Редактировать



, введите текст и нажмите Сохранить.

- Добавить в список — нажмите. Добавить в список  и выберите нужный.

#### 4.10 Просмотр и поиск сниппетов

а) Войдите в подсистему среды разработки.

б) В правой панели инструментов нажмите кнопку Сниппеты



в) В открывшейся оснастке выберите:

- Публичный — доступен всем пользователям.
- Приватный — доступен только вам.

г) Для поиска введите имя сниппета в поле поиска.

#### 4.11 Создание сниппетов

а) Войдите в подсистему среды разработки.

- б) В правой панели инструментов нажмите кнопку Снимпеты.

- в) В открывшейся оснастке нажмите кнопку .

- г) Введите название, описание, запрос, выберите область видимости (Приватный или Публичный).

- д) Нажмите Сохранить.

#### 4.12 Изменение снимпетов

- а) Войдите в подсистему среды разработки.
- б) В правой панели инструментов нажмите кнопку Снимпеты.
- в) В открывшейся оснастке выберите снимпет.
- г) Внесите изменения в запрос или описание.
- д) Нажмите Обновить.

#### 4.13 Удаление снимпетов

- а) Войдите в подсистему среды разработки.
- б) В правой панели инструментов нажмите кнопку Снимпеты.
- в) В открывшейся оснастке выберите снимпет.
- г) Нажмите Удалить.

#### 4.14 Просмотр блокнотов


- а) Войдите в подсистему среды разработки.

- б) В правой панели инструментов нажмите кнопку Блокноты .

- в) В открывшейся оснастке разверните группированные списки:
- Недавние — последние открытые блокноты.
  - Избранное — блокноты, добавленные в избранное.
  - Списки — пользовательские или общие списки блокнотов.
  - Мои блокноты — блокноты, созданные пользователем.
- г) Выберите необходимый блокнот.


#### 4.15 Создание блокнота

- а) Войдите в подсистему среды разработки.
- б) В правой панели инструментов нажмите кнопку Блокноты.

- в) В открывшейся оснастке нажмите кнопку  .
- г) Выберите команду Создать блокнот.
- д) Введите название блокнота.

#### 4.16 Создание блокнота с помощью AI

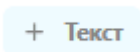
- а) Войдите в подсистему среды разработки.
- б) В правой панели инструментов нажмите кнопку Блокноты.

- в) В открывшейся оснастке нажмите кнопку  .
- г) Выберите команду Создать блокнот с помощью AI.
- д) Выберите каталог.
- е) Выберите базу данных.
- ж) Введите текстовый запрос.
- з) Нажмите Создать.

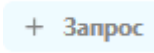
#### 4.17 Типы блоков в блокноте

В блокнотах поддерживается использование трёх типов блоков:

##### Текстовый блок

- а) Нажмите кнопку Текст  .
- б) Введите текстовое описание.

##### Блок SQL-запроса

- а) Нажмите кнопку Запрос  .
- б) Введите название запроса.



в) В блок редактирования введите SQL-запрос.

Использование AI при работе с запросами:

г) В поле работы с AI введите текстовый запрос.

д) Нажмите Enter.

е) Проверьте корректность сгенерированного SQL-запроса.

ж) Нажмите Разрешить.

з) При необходимости измените текстовый запрос и повторно подтвердите.

Выполнение запроса:

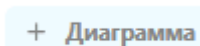


и) Нажмите кнопку Выполнить

Дополнительные действия через меню запроса:

- Форматировать запрос — привести к читаемому виду.
- Объяснить запрос — получить описание запроса с помощью AI.
- Скрыть запрос — оставить только результаты.
- Отобразить шаблон — применить шаблон.

Блок диаграммы



а) Нажмите кнопку Диаграмма

б) Если диаграмма создаётся сразу после запроса, он будет выбран источником данных по умолчанию.

в) Для изменения источника, нажмите Настройки диаграммы

Настройка диаграммы

→ вкладка Данные.



г) Выберите:

- Текущий запрос в блокноте.
- Запрос в блокноте (с указанием номера).
- Из выполненного запроса (с указанием названия и номера).

д) Настройте отображение во вкладках Диаграмма и Визуализация.

е) Нажмите Сохранить.








#### 4.18 Управление блоками

- а) Удалить блок — нажмите кнопку удалить .
- б) Комментарий — нажмите кнопку Комментарий , введите текст и нажмите Comment.

#### 4.19 Добавление переменной в блокноте

- а) Нажмите + Новая переменная.
- б) Укажите имя, тип и значение.
- в) Сохраните изменения.

#### 4.20 Дополнительные возможности

- а)  Запустить все — выполнение всех блоков в блокноте.
- б) Добавить в избранное — значок звёздочки.
- в) Доступ — выбрать Приватный или Публичный режим, при приватном можно назначить других пользователей.
- г) Меню дополнительных действий ):
  - Клонировать  Клонировать — копия блокнота.
  - Экспортировать в PDF  Экспортировать в PDF — сохранить в файл PDF.
  - Списки  Списки — просмотр списка, в котором используется блокнот.
  - Шаблон  Шаблон — сохранить или применить шаблон.
  - Планировщик  Планировщик — задать расписание выполнения.

- Удалить  Удалить — удалить блокнот.

#### 4.21 Работа с ИИ-агентом

ИИ-агент предоставляет возможность интерактивного взаимодействия с моделью LLM через интеграцию по протоколу MCP. Подсистема позволяет формировать запросы на естественном языке, получать автоматически сгенерированные SQL-запросы, выполнять их и просматривать результаты.

#### 4.22 Интерфейс работы с ИИ-агентом

Экран работы с ИИ-агентом состоит из следующих элементов:

- Область ввода запроса — текстовое поле для ввода вопроса или задачи.
- Кнопка "Выполнить" — отправка запроса ИИ-агенту.
- Левая область — чат с ИИ-агентом (история запросов и ответов).
- Правая область — результаты выполнения запросов (таблицы, графики, текстовые описания).
- Кнопка "История" — просмотр предыдущих диалогов с ИИ-агентом.
- Кнопка "Новый чат" — очистка текущей истории и начало нового диалога.

#### 4.23 Выполнение запроса через ИИ-агента

- а) Войдите в подсистему среды разработки.

- б) В левой панели инструментов нажмите кнопку ИИ-агент




- в) В приветственном окне в области ввода запроса введите вопрос или задачу на естественном языке.
- г) Нажмите кнопку Выполнить.
- д) Дождитесь появления ответа в левой области чата.
- е) Если ИИ-агент сгенерировал SQL-запрос:
- Проверьте корректность предложенного запроса.
  - При необходимости внесите изменения.


– Подтвердите выполнение.

ж) Результаты выполнения запроса будут отображены в правой области экрана.

#### 4.24 Начало нового чата

- а) В интерфейсе ИИ-агента нажмите кнопку Новый чат .
- б) Текущая история общения будет очищена.
- в) Введите новый запрос в область ввода и нажмите Выполнить.

#### 4.25 Работа с историей общения

- а) В интерфейсе ИИ-агента нажмите кнопку История .
- б) В открывшемся списке выберите интересующий диалог.
- в) Просмотрите предыдущие запросы и ответы.

## 5 ПОДКЛЮЧЕНИЕ К СИСТЕМЕ ИЗ ВНЕШНИХ ИНСТРУМЕНТОВ

### 5.1 Подключение к платформе данных из Apache Superset

- а) на сервере Apache Superset установите Python client StarRocks, выполнив:
- ```
pip install starrocks
```
- б) В Apache Superset перейдите в раздел подключения БД в качестве источника данных
- в) В разделе поддерживаемых Баз Данных (supported databases) из выпадающего списка выберите StarRocks

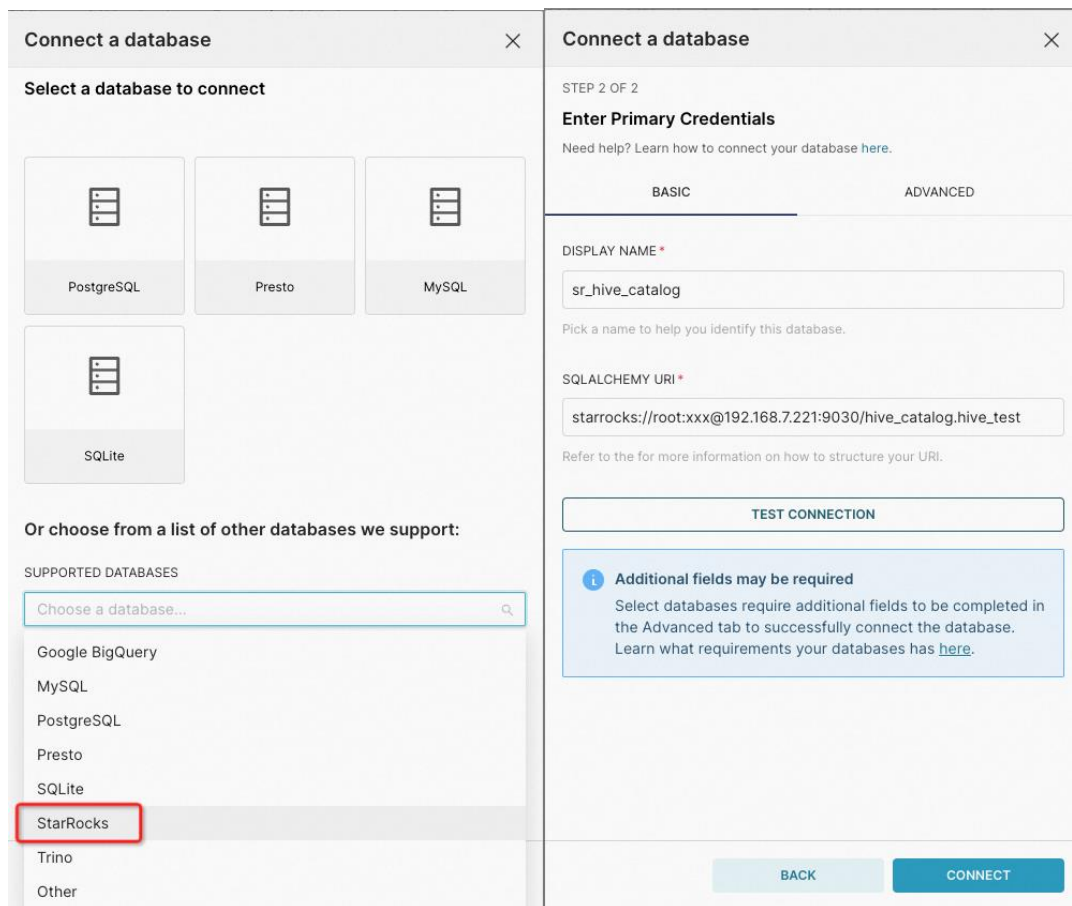


Рисунок 3

- г) На Этапе настройки подключения укажите:
- Отображаемое имя в поле Display Name
  - Путь подключения в поле SQLAlchemy URI:

*starrocks://<User>:<Password>@<Host>:<Port>/<Catalog>.<Database>*

- User – имя пользователя технической УЗ
- Password – пароль от УЗ
- Host – hostname или ip адрес мастер сервера
- Port – порт мастер сервера
- Catalog – имя целевого каталога, поддерживаются внутренние и внешние каталоги
- Database – имя целевой База Данных, поддерживаются внутренние и внешние базы данных

## 5.2 Подключение к платформе данных из Tableau Desktop

Tableau Desktop поддерживает запросы и визуализацию как внутренних, так и внешних данных в Селене.

- а) Создайте базу данных в Tableau
- б) Выберите Other Databases (JDBC) в качестве источника данных.
- в) Для Dialect выберите MySQL.
- г) Для URL введите URL в формате MySQL URI, как показано ниже:

*`jdbc:mysql://<Host>:<Port>/<Catalog>.<Databases>`*

Параметры в URL описаны следующим образом:

- Host: hostname или IP-адрес хоста мастер сервера вашего кластера.
  - Port: порт хоста мастер сервера, например, 9030.
  - Catalog: целевой каталог в вашем кластере. Поддерживаются как внутренние, так и внешние каталоги.
  - Database: целевая база данных в вашем кластере. Поддерживаются как внутренние, так и внешние базы данных.
- д) Настройте имя пользователя и пароль.

## 5.3 Подключение к платформе данных из DataGrip

DataGrip поддерживает запросы как внутренних, так и внешних данных.

Создайте источник данных в DataGrip. Обратите внимание, что в качестве источника данных необходимо выбрать MySQL.

Параметры, которые вам необходимо настроить, описаны ниже:

- Хост: hostname или IP-адрес хоста мастер сервера вашего кластера.

- Порт: порт хоста мастер сервера, например, 9030.
- Аутентификация: метод аутентификации, который вы хотите использовать. Выберите Имя пользователя и пароль.
- Пользователь: имя пользователя, которое используется для входа в ваш кластер, например, admin.
- Пароль: пароль, который используется для входа в ваш кластер.
- База данных: источник данных, к которому вы хотите получить доступ в вашем кластере. Значение этого параметра имеет формат `<catalog_name>.<database_name>`.
- catalog\_name: имя целевого каталога в вашем кластере. Поддерживаются как внутренние, так и внешние каталоги.
- database\_name: имя целевой базы данных в вашем кластере. Поддерживаются как внутренние, так и внешние базы данных.

#### 5.4 Подключение к платформе данных из DBeaver

DBeaver — это клиентское программное обеспечение SQL и инструмент администрирования баз данных.

Чтобы подключиться к базе данных, выполните следующие действия:

- а) Запустите DBeaver.
- б) Щелкните значок плюса (+) в верхнем левом углу окна DBeaver или выберите База данных > Новое подключение к базе данных в строке меню, чтобы получить доступ к помощнику.
- в) Выберите драйвер MySQL.
- г) На этапе выбора базы данных вам будет представлен список доступных драйверов. Нажмите Analytical на левой панели, чтобы быстро найти драйвер MySQL. Затем дважды щелкните значок MySQL.
- д) Настройте подключение к базе данных.
- е) На этапе «Параметры подключения» перейдите на вкладку «Главное» и настройте следующие основные параметры подключения:
  - Хост сервера: hostname или IP-адрес хоста мастер сервера вашего кластера.
  - Порт: порт хоста мастер сервера, например, 9030.

- База данных: целевая база данных в вашем кластере. Поддерживаются как внутренние, так и внешние базы данных
  - Имя пользователя: имя пользователя, которое используется для входа в ваш кластер, например, admin.
  - Пароль: пароль, который используется для входа в ваш кластер.
- ж) Вы также можете просматривать и редактировать свойства драйвера MySQL на вкладке «Свойства драйвера», если это необходимо. Чтобы изменить определенное свойство, щелкните строку в столбце «Значение» для этого свойства.
- з) Протестируйте подключение к базе данных.
- и) Нажмите «Проверить подключение», чтобы проверить точность параметров подключения. Появится диалоговое окно с информацией о драйвере MySQL. Нажмите «ОК» в диалоговом окне, чтобы подтвердить информацию. После успешной настройки параметров подключения нажмите «Готово», чтобы завершить процесс.
- к) После установки подключения вы можете просмотреть его в дереве подключений к базе данных слева, и DBeaver сможет эффективно подключиться к базе данных.

## 5.5 Подключение к платформе данных из Jupyter

Вы можете использовать JupySQL поверх Jupyter для выполнения запросов к Селена.

После загрузки данных в кластер вы можете запрашивать и визуализировать их с помощью построения графиков SQL.

Перед началом работы вам необходимо установить локально следующее программное обеспечение:

- JupySQL: *pip install jupysql*
- Jupyterlab: *pip install jupyterlab*
- SKlearn Evaluation: *pip install sklearn-evaluation*
- Python
- pymysql: *pip install pymysql*



После выполнения вышеуказанных требований вы можете открыть Jupyter lab, просто вызвав `jupyterlab` — это откроет интерфейс блокнота. Если Jupyter lab уже запущен в блокноте, вы можете просто запустить ячейку ниже, чтобы получить зависимости.

```
%pip install --quiet jupysql sklearn-evaluation pymysql
```

Для использования обновленных пакетов может потребоваться перезапуск ядра.

```
import pandas as pd  
from sklearn_evaluation import plot  
# Import JupySQL Jupyter extension to create SQL cells.  
%load_ext sql  
%config SqlMagic.autocommit=False
```

Вам нужно будет настроить строку подключения в соответствии с типом экземпляра, к которому вы пытаетесь подключиться (url, пользователь и пароль). В примере ниже используется локальный экземпляр.

#### Подключение к Селена через JupySQL

В этом примере используется экземпляр Docker, который отражает данные в строке подключения.

Пользователь `root` используется для подключения к локальному экземпляру Селена, создание базы данных и проверки того, что данные действительно могут быть прочитаны и записаны в таблицу.

```
%sql mysql+pymysql://root:@localhost:9030
```

Создать JupySQL БД:

```
%sql CREATE DATABASE jupysql;  
%sql USE jupysql;
```

Создать таблицу:

```
%%sql
```

```
CREATE TABLE tbl(c1 int, c2 int) distributed by hash(c1) properties
("replication_num" = "1");
INSERT INTO tbl VALUES (1, 10), (2, 20), (3, 30);
SELECT * FROM tbl;
```

#### Сохранение и загрузка запросов

Теперь, после создания базы данных, вы можете записать в нее некоторые образцы данных и запросить их.

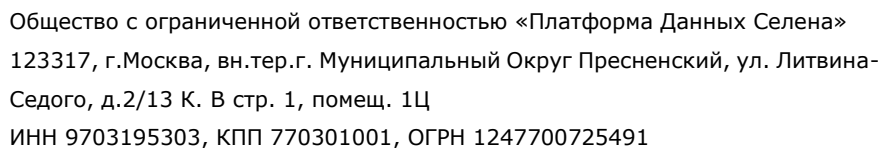
JupySQL позволяет разбивать запросы на несколько ячеек, упрощая процесс создания больших запросов.

Вы можете писать сложные запросы, сохранять их и выполнять при необходимости, аналогично CTE в SQL.

```
%%sql --save initialize-table --no-execute
```

```
CREATE TABLE tbl(c1 int, c2 int) distributed by hash(c1) properties ("replication_num" =
"1");
INSERT INTO tbl VALUES (1, 1), (2, 2), (3, 3);
SELECT * FROM tbl;
```

Обратить внимание, что используется `--with;`, это извлечет ранее сохраненные запросы и добавит их (используя CTE). Затем мы сохраняем запрос в `track_fav`.

43