**InnData**

**Руководство пользователя**

2024

**Содержание**

[Обозначения и сокращения 4](#_Toc140240485)

[1 Общие Сведения 5](#_Toc140240486)

[1.1 Наименование проектируемой системы 5](#_Toc140240487)

[1.2 Назначение системы 5](#_Toc140240488)

[1.3 Цели системы 5](#_Toc140240489)

[2 Введение 6](#_Toc140240490)

[3 Интерфейс Superset 8](#_Toc140240491)

[3.1 Работа с SQL 8](#_Toc140240492)

[3.1.1 Просмотр структуры таблиц и представлений 9](#_Toc140240493)

[3.1.2 Создание запросов 14](#_Toc140240494)

[3.1.3 Сохранение запросов 15](#_Toc140240495)

[3.1.4 Работа с результатами запросов 17](#_Toc140240496)

[3.2 Работа с дашбордами 19](#_Toc140240497)

[3.2.1 События 20](#_Toc140240498)

[3.2.2 Отчет по вендору, журналу и хосту 20](#_Toc140240499)

[3.2.3 Сетевое взаимодействие 21](#_Toc140240500)

[3.2.4 Количество событий по коду события 22](#_Toc140240501)

[3.2.5 Активность учетных записей 23](#_Toc140240502)

[3.2.6 Отчет по учетной записи 24](#_Toc140240503)

[4 Интерфейс Hue 25](#_Toc140240504)

[4.1 Работа с SQL-запросами в Hue 29](#_Toc140240505)

[4.2 Результаты выполнения запроса 35](#_Toc140240506)

[4.3 Панель листинга файлов в HDFS через Hue 39](#_Toc140240507)

[Приложение A (обязательное) Схема данных 44](#_Toc140240508)

# Обозначения и сокращения

|  |  |  |
| --- | --- | --- |
| БД | – | База данных |
| ИБ | – | Информационная безопасность |
| ЛВС | – | Локальная вычислительная сеть |
| СКИ | – | Система классификации инцидентов |
| СУБД | – | Система управления базами данных |
| CSV | – | Comma-Separated Values |
| JSON | – | JavaScript Object Notation |
| MP SIEM | – | Security Information and Event Management |
| SQL | – | Structured Query Language |

# Общие Сведения

## Наименование системы

Полное наименование программного продукта InnData: Платформа InnData.

Условное обозначение: Система.

## Назначение системы

InnData предназначена для:

* получения/сбора, обработки, хранения и анализа данных;
* долгосрочного исторического хранения данных с необходимой глубиной при условии выделения соответствующей емкости жёстких дисков;
* представления хранимых данных в виде дашбордов;
* консолидации в едином хранилище, данных из различных информационных систем;
* мониторинга состояния;
* реализации работы моделей машинного обучения;
* создания моделей машинного обучения;
* хранения артефактов результатов машинного обучения.

## Цели системы

Система предназначена для консолидации в едином пространстве разнородной информации для её хранения, анализа и моделирования при помощи моделей машинного обучения (далее ML-алгоритмы).

# Введение

В рамках работы Системы выполняется:

1. Получение и хранение данных из различных информационных систем, в том числе систем учета и управления инцидентами ИБ – MP SIEM, Microfocus ArcSight, Kaspersky KUMA.
2. Формирование представлений для возможности взаимодействия с хранимыми данными.

Для обеспечения возможности взаимодействия пользователя с хранимыми данными используются следующие компоненты:

* Apache Superset;
* Clickhouse;
* Greenplum;
* Impala
* Shiny;
* Apache Hue.

Apache Superset – программное обеспечение для исследования и визуализации данных, ориентированное на большие данные. Обеспечивает выполнение следующих функций:

* Предоставляет редактор для выполнения запросов к данным, находящимся в хранилище данных;
* Интерфейс для создания собственных страниц представления данных в табличном и графическом виде;
* Возможность анализа и сохранения результатов.

Clickhouse – колоночная аналитическая СУБД, позволяющая выполнять аналитические запросы в режиме реального времени на структурированных больших данных. Обеспечивает выполнение следующих функций:

* Хранение данных;
* Выполнение SQL запросов к данным;
* Создание и поддержание агрегированных выборок данных;
* Использование метаданных и индексов для ускорения выполнения запросов.

Greenplum – СУБД, основанная на популярной СУБД Postgres, классической объектно-реляционной СУБД. Обеспечивает выполнение следующих функций:

* Хранение данных;
* Выполнение SQL запросов к данным;
* Возможность выполнения полного спектра операций с данными – добавление, изменение, удаление.

Impala - это механизм запросов SQL для данных, хранящихся в Hadoop.

Impala реализует:

* Выполнение запросов к данным хранящимся в Hadoop;
* Поддержку различных типов хранилищ данных HDFS, S3, ABFS, Apache HBase and Apache Kudu;
* Поддержку чтения различных форматов данных включая text, LZO, SequenceFile, Avro, RCFile, Parquet и ORC;
* Поддержку ролевой модели доступа к данным.

Shiny – фреймворк для создания интерактивных веб-приложений с графическим интерфейсом пользователя для анализа данных.

Возможности:

* Быстрое проектирование приложений на языке высшего уровня;
* Готовые интерактивные элементы форм;
* Возможность подключения большого количества визуализаций данных.

Apache Hue – многопользовательское интерактивное браузерное программное средство для выполнения запросов к данным, совместной работы над данными. Обеспечивает выполнение следующих функций:

* Возможность выполнения запросов к данным с помощью SQL-запросов;
* Возможность работы с данными в частности копирование, удаление и пр.
* Создание визуальных представлений данных.

# Интерфейс Superset

Для начала работы с Apache Superset необходимо перейти по адресу http://{hostname}:32688. В окне авторизации указать доменную учетную запись (при использовании авторизации через LDAP) или локальную учётную запись (при использовании локальной авторизации).

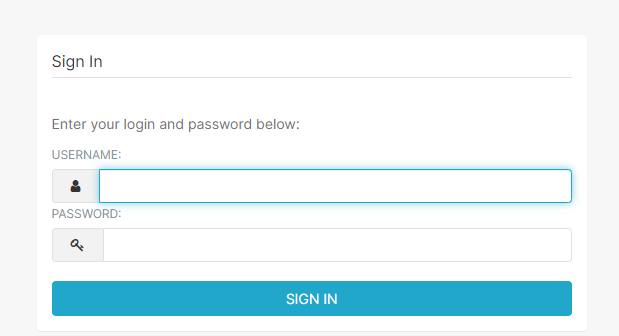


Рисунок ‑ – Окно авторизации

Работа с данными может осуществляться несколькими способами. Первый – это непосредственное выполнение SQL запросов в SQL Lab, второй – анализ чартов (визуализаций), скомпонованных в сборные страницы – дашборды. Чарты в свою очередь создаются на основе подготовленных соединений к таблицам или выборок на основе SQL – датасетах.

## Работа с SQL

Superset предоставляет среду для работы с данными напрямую через SQL-запросы. Для этого существует специальный раздел с отдельными правами доступа – SQL Lab.

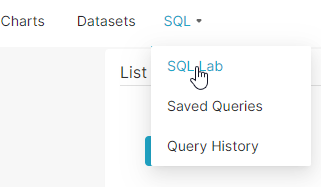


Рисунок ‑ – Раздел SQL Lab в меню навигации

На рисунке выше (Рисунок 3.2) показана компоновка раздела SQL Lab. На рисунке ниже (Рисунок 3.3), в верхней части раздела располагаются табы (под цифрой 1), нажимая на которые можно переключаться между разными запросами. Под каждым табом находится рабочее пространство, состоящее из панели навигации по БД (под цифрой 2), поля ввода запроса (под цифрой 3) и блок выдачи результатов (под цифрой 4). Рабочее пространство отображается только для активного таба. Отображение панели навигации по БД можно выключать (три точки у названия таба – Hide tool bar).

Табы с запросами можно добавлять, переименовывать, дублировать, удалять. Список табов с запросами и список сохраненных запросов у каждого пользователя свой и недоступен другим пользователям.

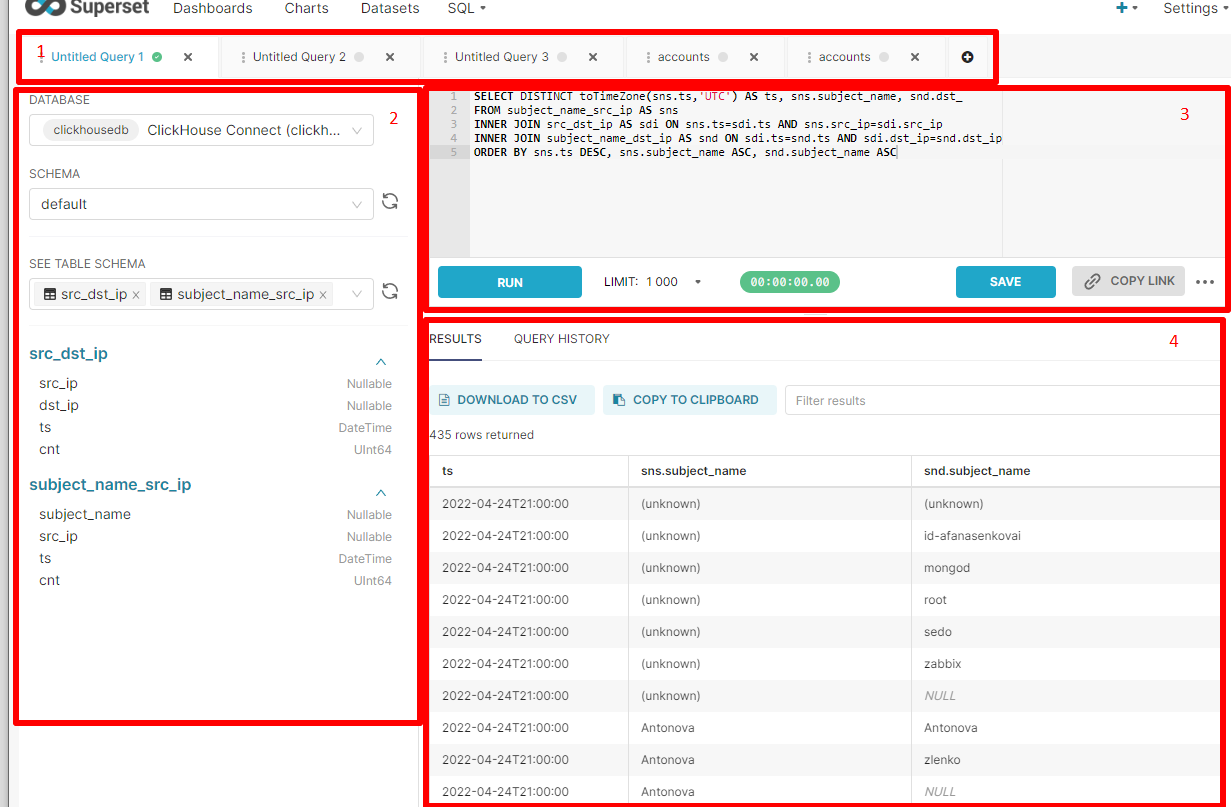


Рисунок ‑ – Компоновка раздела SQL Lab

### Просмотр структуры таблиц и представлений

Для просмотра информации о доступных таблицах нужно выполнить следующие действия.

Войти в SQL lab (Рисунок 3.3), добавить новый таб или использовать существующий (Рисунок 3.4).

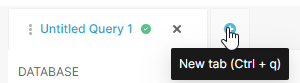


Рисунок ‑ – Добавление нового таба

Далее в выпадающем списке под заголовком «DATABASE» необходимо выбрать соединение с БД, например - «Clickhouse Datasets», в выпадающем списке «SEE TABLE SCHEMA» выбрать одну или несколько таблиц. Пиктограмма около названия таблицы подсказывает ее тип, «таблица» - собственно таблица, «глаз» - представление, view.

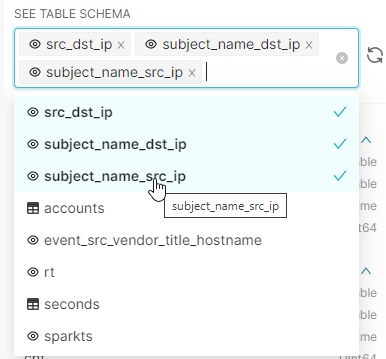


Рисунок ‑ – Выбор таблицы для просмотра

После выбора таблиц ниже появится их перечень со списком колонок для каждой. Для каждой колонки таблицы указано название и тип данных. Эта информация помогает понять схему таблиц и видеть названия их колонок для написания корректного SQL запроса по ним. Также при наведении на название таблицы появляются пиктограммы сортировки списка полей, копирования готового запроса выборки по таблице (при нажатии в буфер clipboard скопируется готовый запрос SELECT для выборки из указанной таблицы с перечислением полного списка полей) и удаления таблицы из списка выбранных.

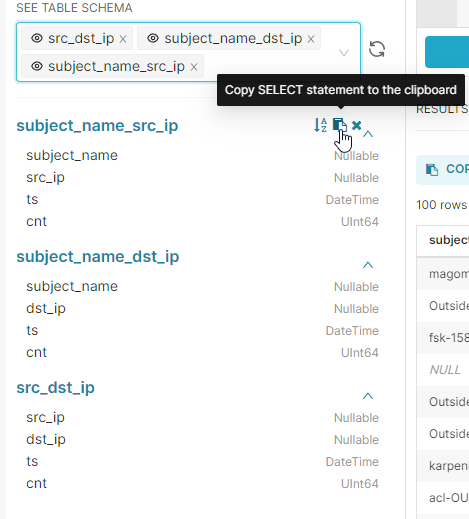


Рисунок ‑ – Схемы таблиц

При выборе таблиц в блоке результатов появятся табы с шаблонным именем «PREVIEW: `YOUR\_TABLE\_NAME`», где YOUR\_TABLE\_NAME это имя каждой из выбранных таблиц. Для каждого таба при его выборе ниже выводится таблица с результатами выборки из указанной таблицы, небольшой (100 строк) набор данных для иллюстрации того, какие данные могут быть выбраны из таблицы рисунок ниже (Рисунок 3.7).

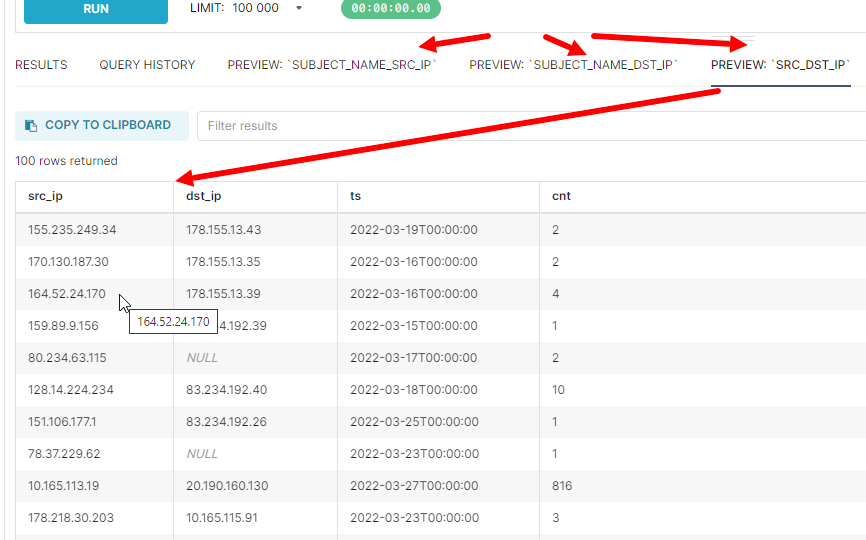


Рисунок ‑ – Предпросмотр случайного набора данных из таблицы

В панели навигации по БД (выпадающий список DATABASE) можно выбрать соединение с БД, далее схему в SCHEMA (список можно обновить, нажав на иконку рядом со списком) и в списке таблиц SEE TABLE SCHEMA (также можно обновить до актуального, если в БД были изменения) выбрать несколько таблиц. По каждой из выбранных таблиц будет показана схема и типы колонок, в блоке результатов показаны семплы данных (до 100 строк) для выбранных таблиц. Эта информация помогает строить SQL запрос непосредственно в поле его ввода. На рисунке выше (Рисунок 3.7) также показана функция копирования готового шаблона SQL запроса SELECT для выбранной таблицы, что позволяет ускорить ввод и уменьшает количество ошибок в наборе полей.

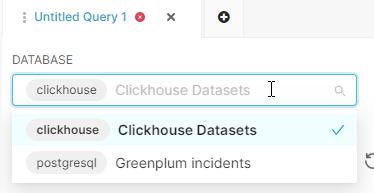


Рисунок ‑ – Выбор соединения с БД

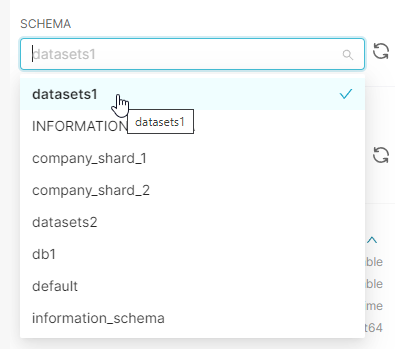


Рисунок ‑ – Выбор схемы БД

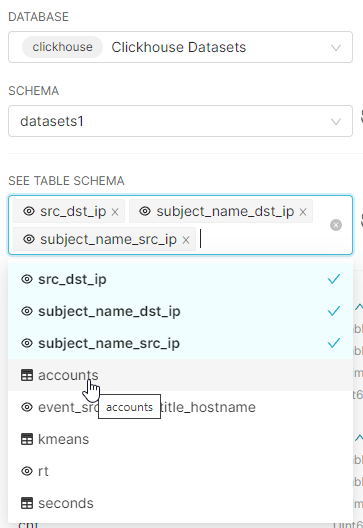


Рисунок ‑ – Выбор таблиц/представлений

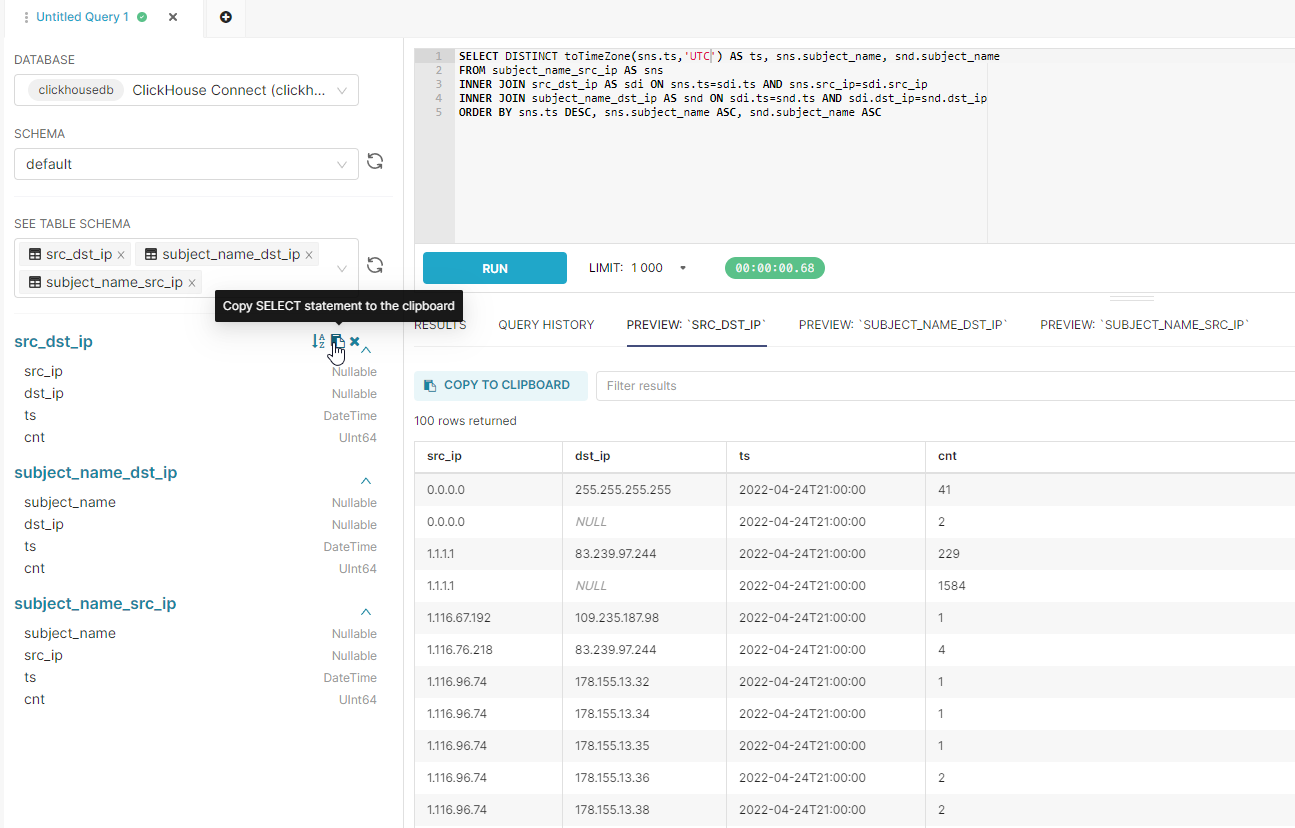


Рисунок ‑ – Просмотр схем и данных таблиц

### Создание запросов

Запросы SQL пишутся в поле ввода запроса и выполняются при нажатии на кнопку «RUN». Во время выполнения запроса кнопка «RUN» меняется на кнопку «STOP», нажатием на нее можно остановить выполнение запроса.

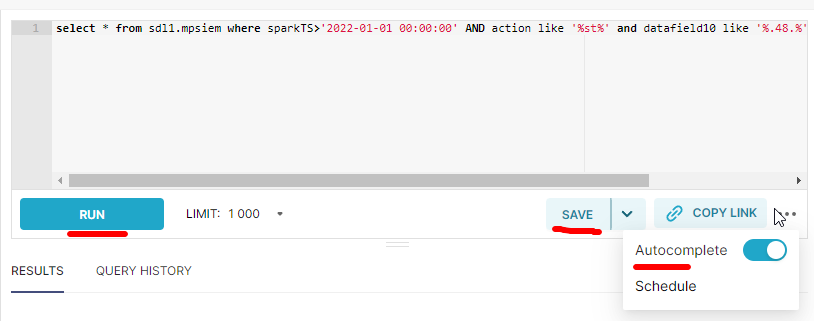


Рисунок ‑ – Поле ввода запроса, кнопки запуска и сохранения

Поддерживается автодополнение текста запросов с возможностью его выключения. В этом режиме при вводе предлагается подсказка для быстрого и безошибочного ввода терминов SQL и названий схем и таблиц. Если выбрать имя таблицы с помощью автодополнения, то она добавляется в список выбранных таблиц в панели навигации по БД, там же покажется ее схема колонок.

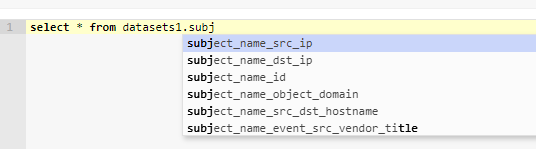


Рисунок ‑ – Автодополнение в запросах

### Сохранение запросов

Запросы можно сохранять для дальнейшего использования. Для этого запрос надо выполнить (можно даже ошибочно), затем нажать кнопку «SAVE» под текстом запроса, в вышедшем окне ввести название и краткое описание и нажать кнопку «SAVE». При этом сохраняется название запроса с кратким описанием, значения в поле выбора соединения с БД, выбранная схема и сам текст запроса. При работе с ранее сохраненным запросом в выпадающем окне вместо кнопки «SAVE» будут кнопки «SAVE AS NEW» – для сохранения отдельно и «UPDATE» для изменения уже сохраненного ранее запроса.

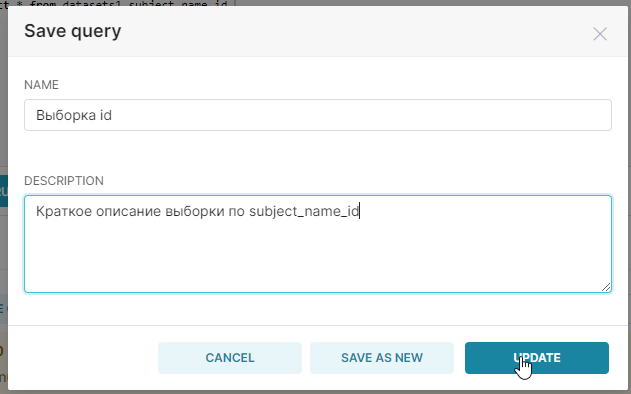


Рисунок ‑ – Сохранение отредактированного ранее сохраненного запроса

Сохраненные запросы можно просмотреть, нажав на ссылку в меню SQL – Saved Queries.

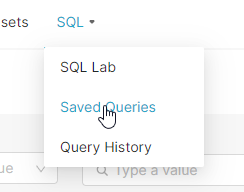


Рисунок ‑ – Ссылка на сохраненные запросы

По этой ссылке можно увидеть список сохраненных запросов, они будут доступны даже если в SQL Lab закрыть таб с этими запросами. Каждый пользователь видит только свой список сохраненных запросов.

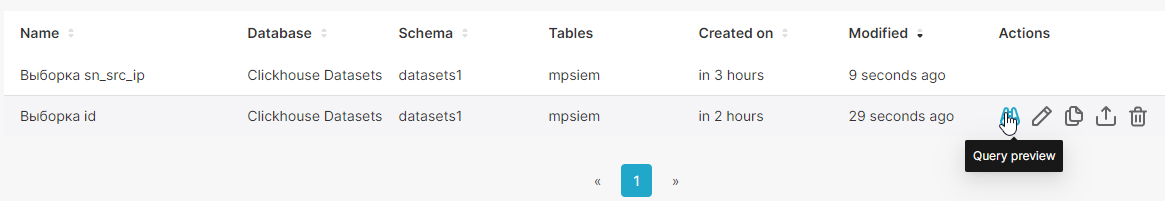


Рисунок ‑ – Список сохраненных запросов

С сохраненными вопросами возможны следующие операции: предпросмотр запроса (query preview), редактирование, копирование ссылки на запрос, экспорт и удаление. При клике на предпросмотр появляется модальное окно с текстом запроса.

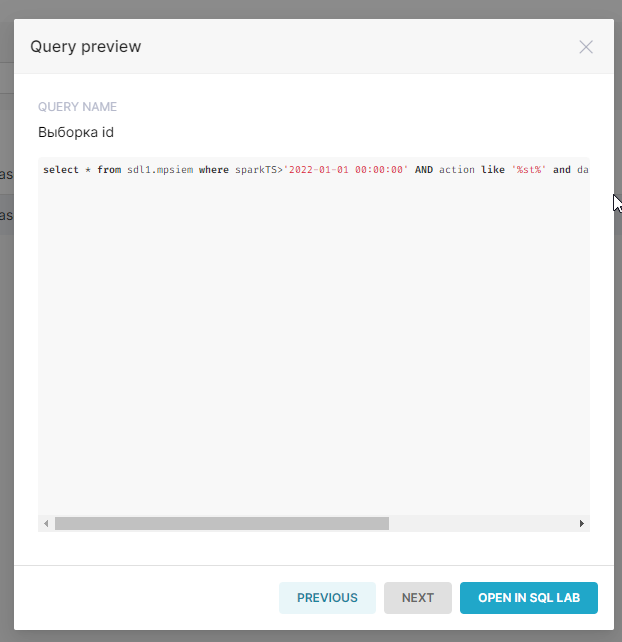


Рисунок ‑ – Предпросмотр запроса

При клике на редактирование запроса в SQL Lab открывается новая вкладка с названием как у сохраненного запроса и с его текстом SQL в поле ввода запроса. Экспорт сохраненного запроса сохраняет его в отдельный архив, который можно перенести на другой сервер при необходимости.

### Работа с результатами запросов

При выполнении запроса в SQL Lab количество строк в результирующей выборке по умолчанию ограничено 1000, это число можно поменять, выбрав в соответствующем выпадающем списке один из вариантов: 10, 100, 1000, 10000, 100000.

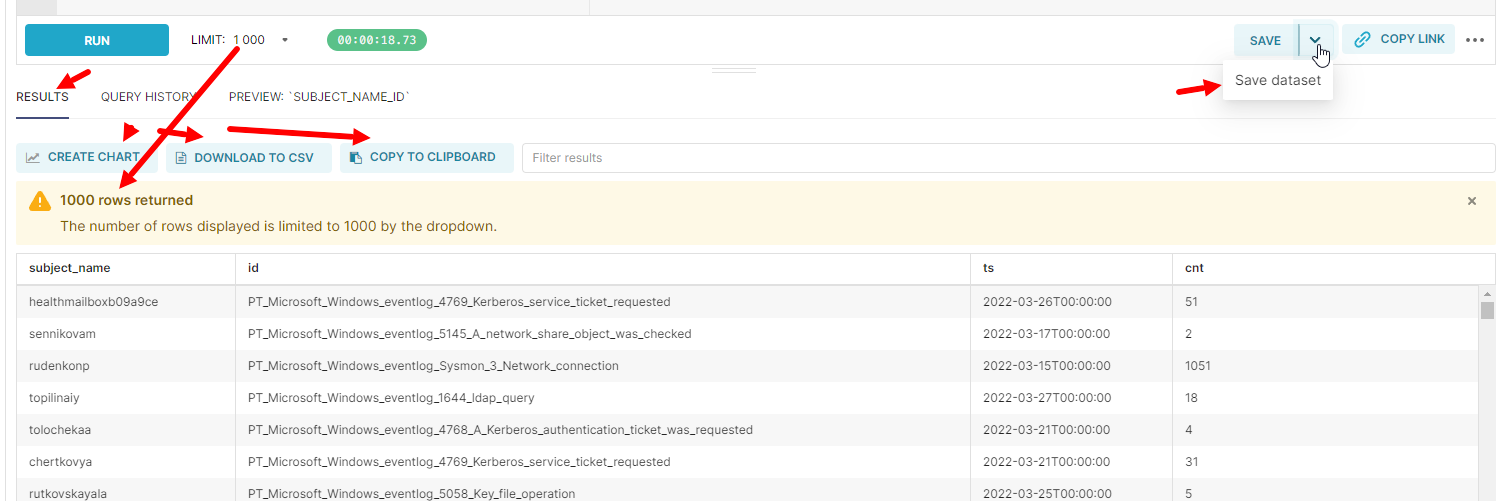


Рисунок ‑ – Работа с результатами запросов

Результаты выводятся в виде таблицы. Результаты можно фильтровать, введя текст в поле ввода «Filter results», при этом в таблице останутся те строки, в которых есть вхождение введенного текста. Для полей с JSON-структурой внутри возможен при клике на поле ее вывод в виде дерева в отдельном окне, как показано на рисунке ниже (Рисунок 3.19).

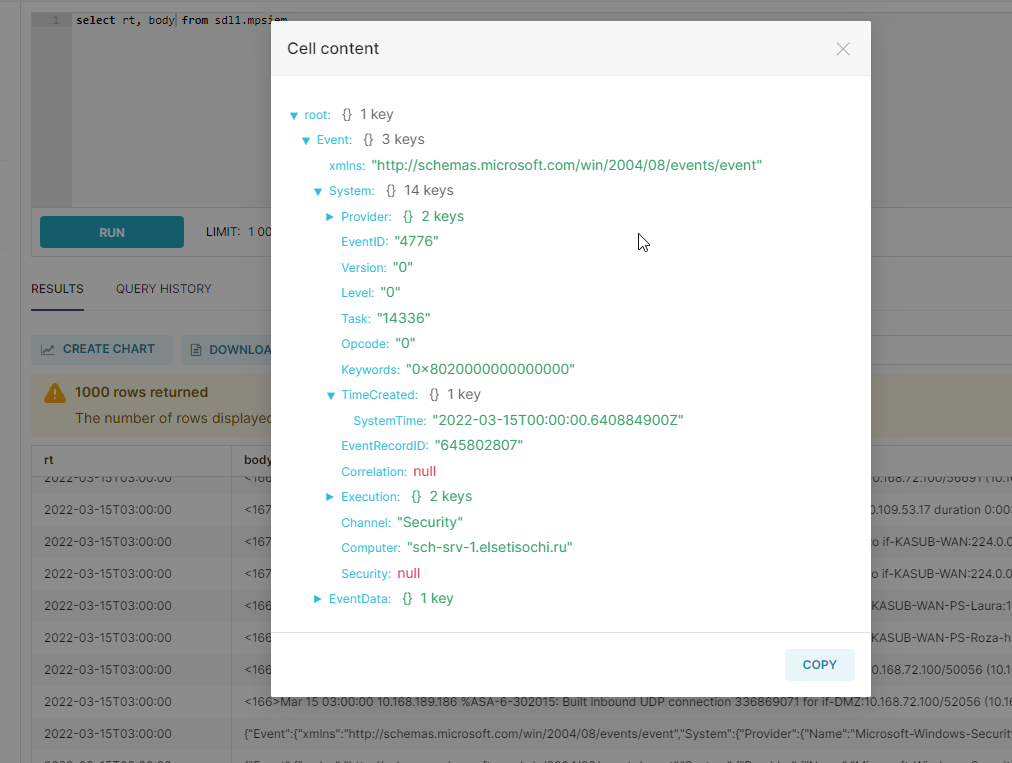


Рисунок ‑ – Просмотр сложной структуры в поле выборки

Результаты можно сохранить в файле формата CSV, нажав на кнопку «DOWNLOAD TO CSV», скопировать в буфер обмена (кнопка «COPY TO CLIPBOARD»).

## Работа с дашбордами

Для работы с дашбордами необходимо перейти в раздел «Dashboards» (Рисунок 3.20).

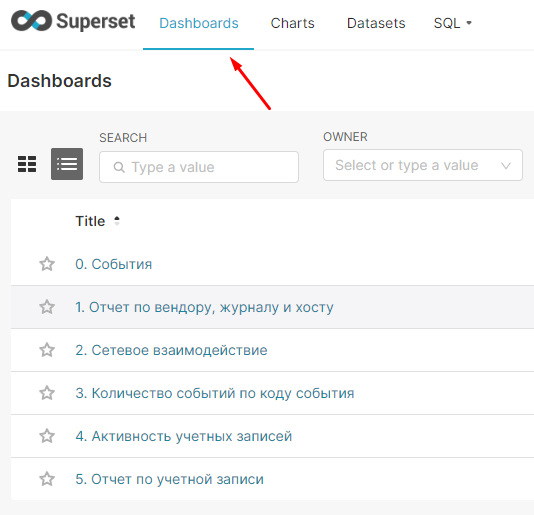


Рисунок ‑ – Раздел Dashboards

В разделе представлены следующие дашборды:

* 0. События;
* 1. Отчет по вендору, журналу и хосту;
* 2. Сетевое взаимодействие;
* 3. Количество событий по коду события;
* 4. Активность учетных записей;
* 5. Отчет по учетной записи.

### События

Системный дашборд, отображает сводную информацию по событиям в системе.

### Отчет по вендору, журналу и хосту

Пользовательский дашборд, отображает сводную информацию по вендору (поле «event\_src\_vendor») и типу журнала (поле «event\_src\_title») для указанного источника (поле «src\_hostname»).

Для просмотра отчета необходимо указать параметры фильтра (Рисунок 3.22) в следующей последовательности:

* Временной период – временной диапазон, в котором происходили события. На рисунке (Рисунок 3.21) указан режим выбора периода «Custom», который позволяет использовать календарь для выбора дат.

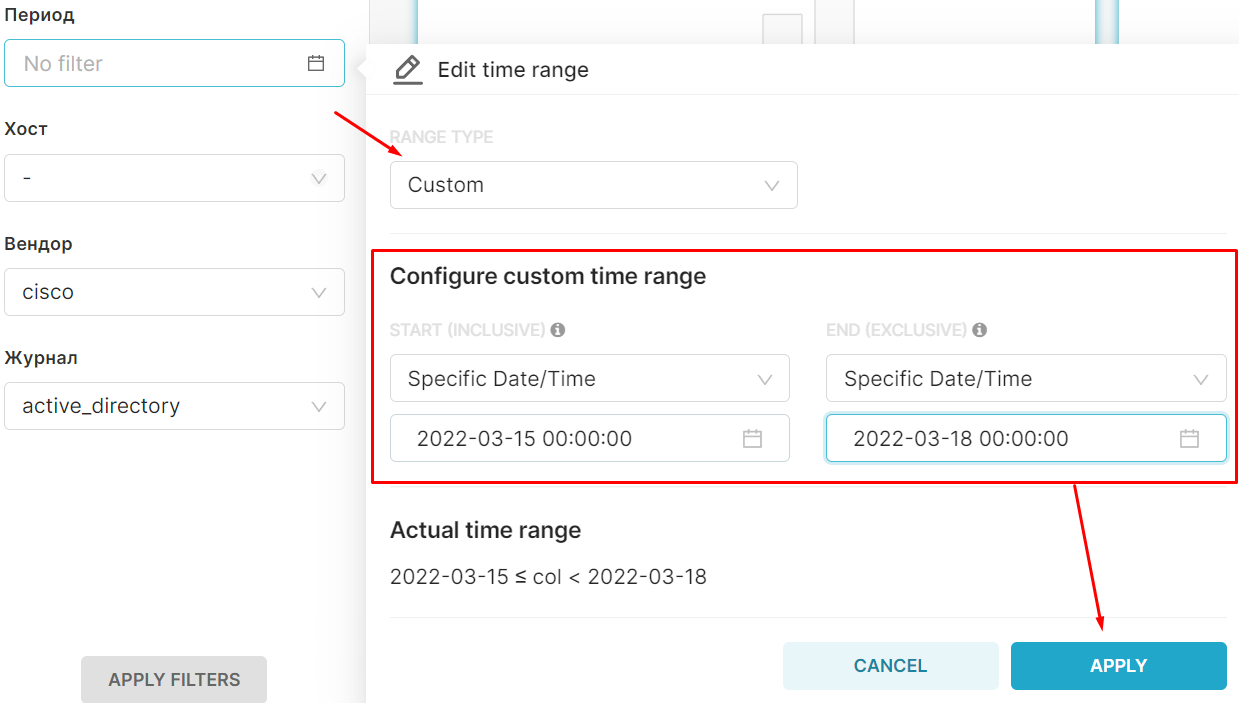


Рисунок ‑ – Выбор даты

* Хост (по умолчанию выбирается 1-й из списка).
* Вендор – в выпадающем списке будут отображены вендоры, по которым имеются события для выбранного хоста (по умолчанию выбирается 1-й из списка).
* Журнал - в выпадающем списке будут отображены типы журналов, по которым имеются события для выбранного хоста и вендора (по умолчанию выбирается 1-й из списка).

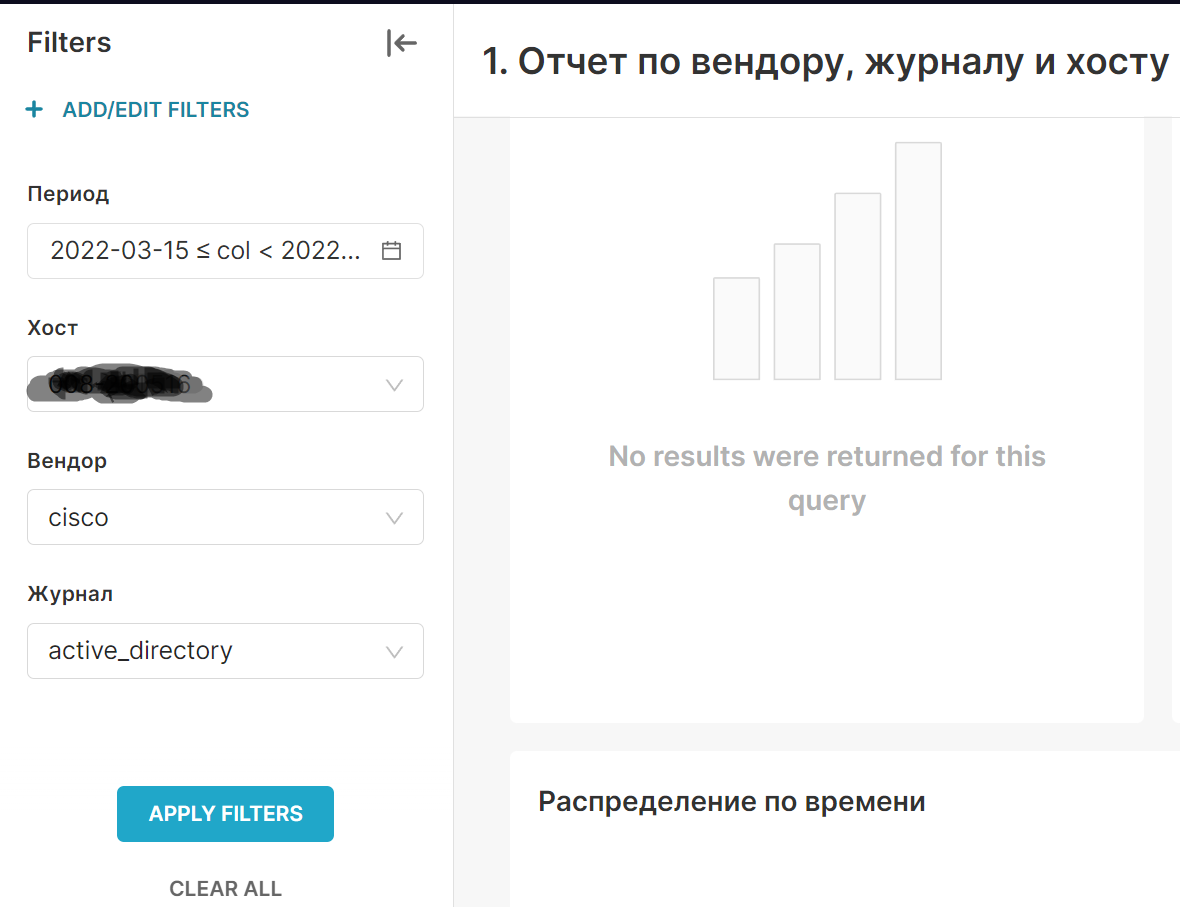


Рисунок ‑ – Отчет по вендору

После выбора всех параметров фильтра нажать кнопку «Apply Filters» для обновления данных в отчете.

### Сетевое взаимодействие

Отображает распределения количества событий по узлам и доменам сети. В дашборде используются поля «src\_hostname», «dst\_hostname», «src\_ip», «dst\_ip», «object\_domain».

Для просмотра отчета необходимо указать параметр фильтра – Период (Рисунок 3.21).

При необходимости можно указать дополнительные фильтры (Рисунок 3.23), в фильтре можно выбрать несколько значений.

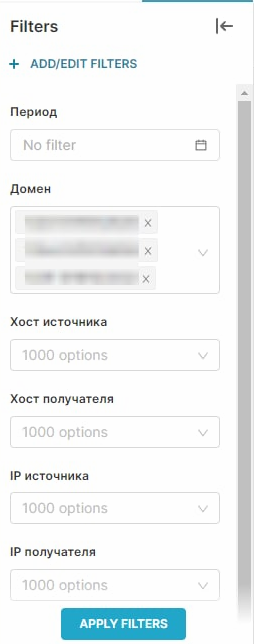


Рисунок ‑ – Фильтры отчета «Сетевое взаимодействие»

### Количество событий по коду события

Отображает распределение событий по коду события безопасности (поле «id»)

Для просмотра отчета необходимо указать параметр фильтра – Период (Рисунок 3.21).

При необходимости можно указать дополнительный фильтр по коду события (Рисунок 3.24), в фильтре можно выбрать несколько значений.

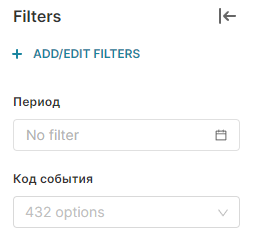


Рисунок ‑ – Фильтры отчета «Количество событий по коду события»

### Активность учетных записей

Отображает сводную информацию по наиболее активным учетным записям (поле «subject\_name») и узлам (поля «src\_hostname», «dst\_hostname»).

Для просмотра отчета необходимо указать параметр фильтра – Период (Рисунок 3.21).

При необходимости можно указать дополнительный фильтр по коду события (Рисунок 3.25), в фильтре можно выбрать несколько значений.

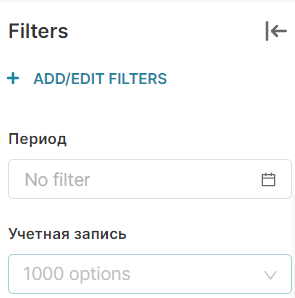


Рисунок ‑ – Фильтры отчета «Активность учетных записей»

### Отчет по учетной записи

Отображает подробную информацию по вендору (поле «event\_src\_vendor»), типу журнала (поле «event\_src\_title»), коду события безопасности (поле «id»), доменам (поле «object\_domain») и узлам (поле «src\_hostname») для указанной учетной записи (поле «subject\_name»).

Для просмотра отчета необходимо указать параметр фильтра – Период (Рисунок 3.21).

При необходимости можно указать дополнительный фильтр по учетной записи (Рисунок 3.25), в фильтре можно выбрать несколько значений. Также можно выбрать единицу времени, по которой группируются события на графике – день, неделя, месяц.

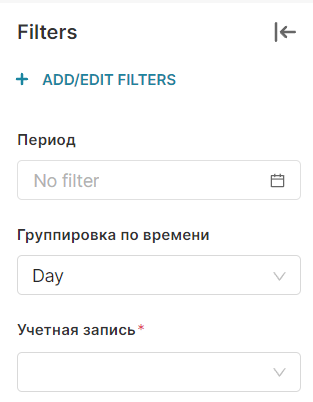


Рисунок ‑ – Фильтры отчета «Отчет по учетной записи»

# Интерфейс Hue

Hue предоставляет среду для работы с данными напрямую через SQL-запросы. Для этого существует специальный Editor в котором пользователь выполняет запросы к данным.

Для начала работы с Hue необходимо перейти по адресу http://<IP>:31890 и выполнить авторизацию:

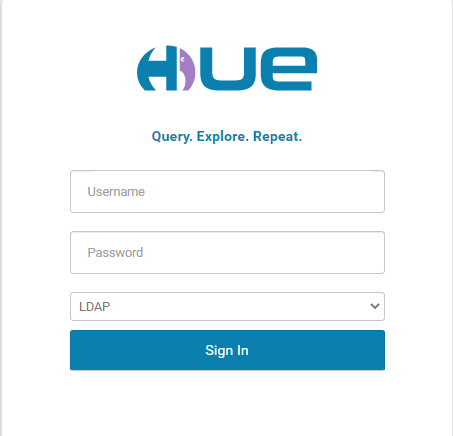
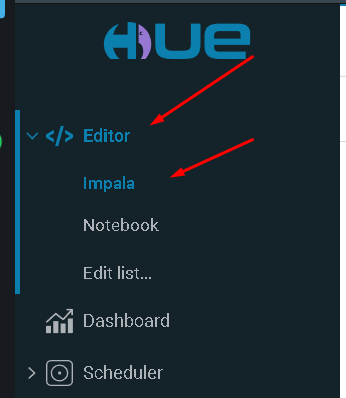


Рисунок ‑ Окно авторизации в Hue

После авторизации в Hue можно работать с редактором запросов, составлять запросы, просматривать и выгружать результаты.

Для написания запросов необходимо открыть «Impala Editor», для этого под логотипом «Hue» в левой части экрана нажать «Editor» -> «Impala»

Рисунок ‑ Editor impala

Работа в редакторах заключается в выполнении запросов к данным. Запросы формируется с помощью языка запросов SQL. Редактор поддерживает интеллектуальное автозаполнение, предупреждения о рисках и рекомендации по устранению неполадок.

Типичный сценарий анализа данных состоит в поиске или импорте данных, их изучении, а затем в построении диаграмм и экспорте результатов.

После выбора редактора Impala отобразится панель с доступными пользователю базами данных и таблицами в них (Рисунок 4‑3) и (Рисунок 4‑4):

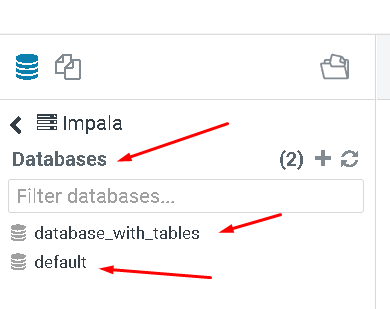


Рисунок ‑ Панель с доступными базами данных

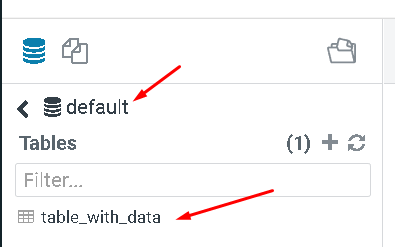


Рисунок ‑ Панель с доступными таблицами в БД default

Схема и набор данных этих таблиц\представления назначается администратором подсистемы.

В правой части экрана находится каталог быстрого просмотра структуры таблицы\представления, к которому выполняется запрос, SQL функций и справочника SQL (**Рисунок 4‑5**)

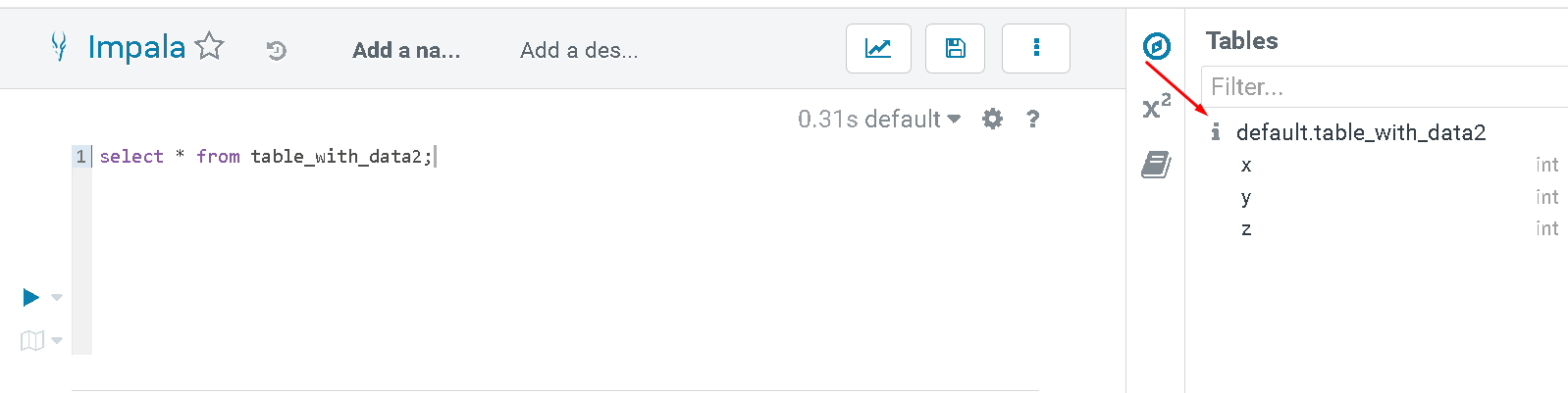


Рисунок ‑ Структура таблицы

Любой запрос в редакторе можно сохранить как документ для дальнейшего использования (**Рисунок 4‑6** и **Рисунок 4‑7**)



Рисунок ‑ Сохранение имени запроса

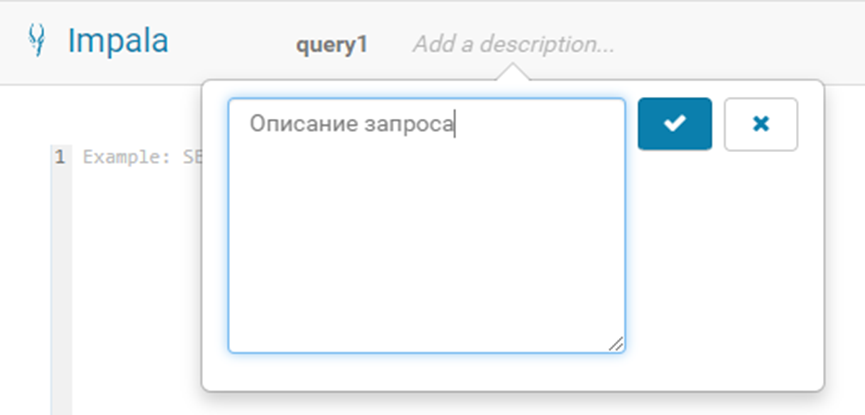


Рисунок ‑ Описание запроса

Для сохранения запроса выбрать соответствующий символ

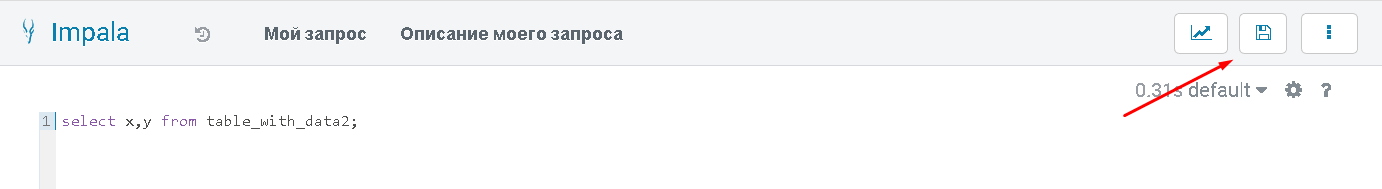


Рисунок 4‑8 Сохранение пользовательского запроса

Список сохраненных запросов отображается во вкладке «Saved Queries»

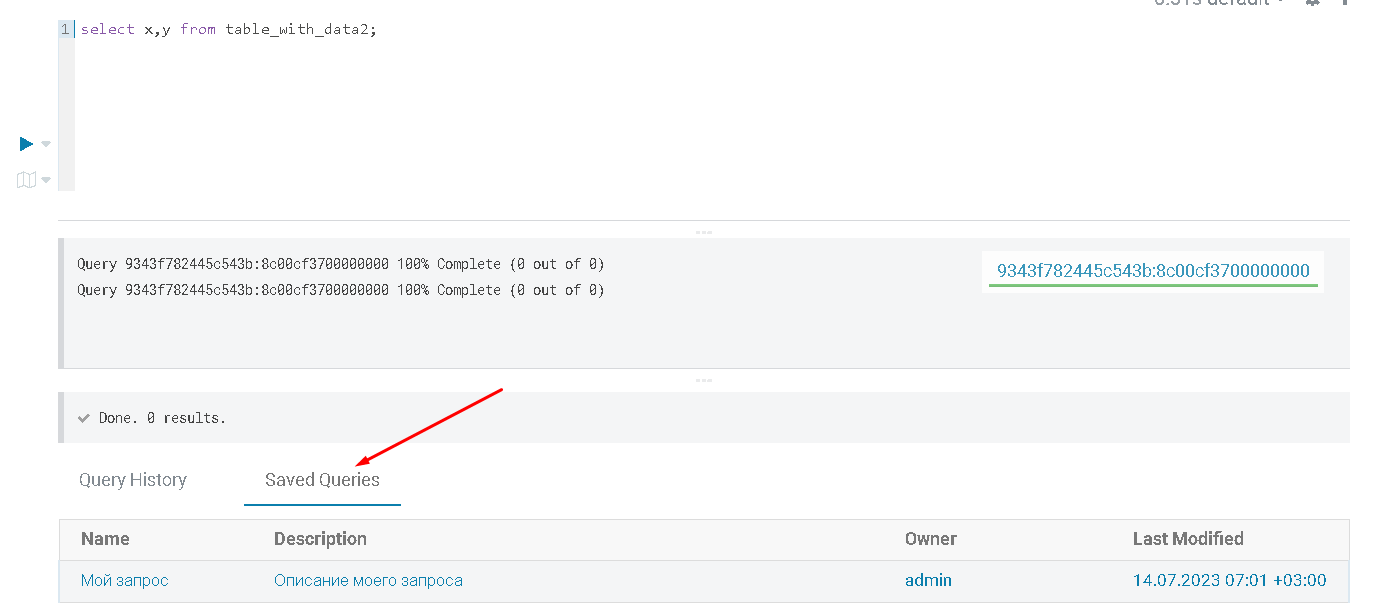


Рисунок 4‑9 Список сохраненных запросов

## Работа с SQL-запросами в Hue

Для выполнения запроса необходимо:

Ввести текст запроса в окно редактора запроса. Для выполнения части запроса необходимо выделить исполняемую часть запроса.

Нажать на кнопку запуска запроса «» или сочетание клавиш клавиатуры «Ctrl+Enter». Во время выполнения запроса индикатор выглядит следующим образом «» до тех пор, пока запрос не будет выполнен полностью.

Для быстрого просмотра значений отдельных столбцов запроса необходимо в окне редактора запроса выбрать столбец и нажать на правую кнопку мыши. Откроется окно значений столбца. Выведенные данные можно отфильтровать, отсортировать, выбрать максимальное и минимальное значения (Рисунок 4‑10)

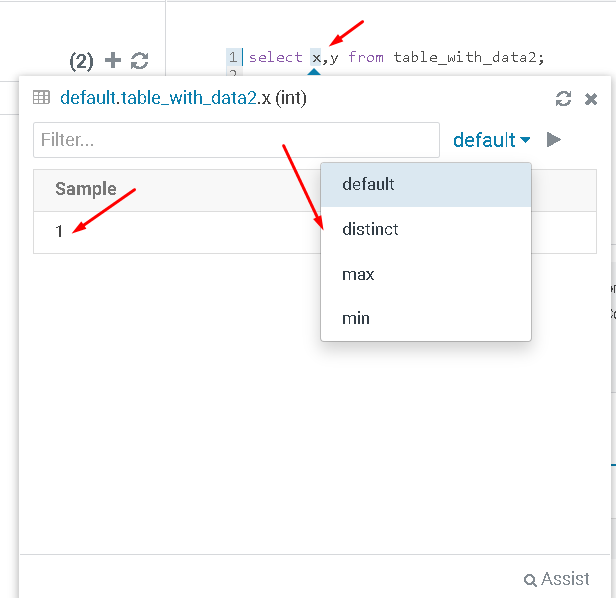


Рисунок ‑ Окно просмотра результатов и варианты выборки

В Hue доступна функция параметризации запросов. Для этого необходимо в запросе указать следующую конструкцию: WHERE Название\_столбца= "${Название\_столбца}" (**Рисунок 4‑11**)

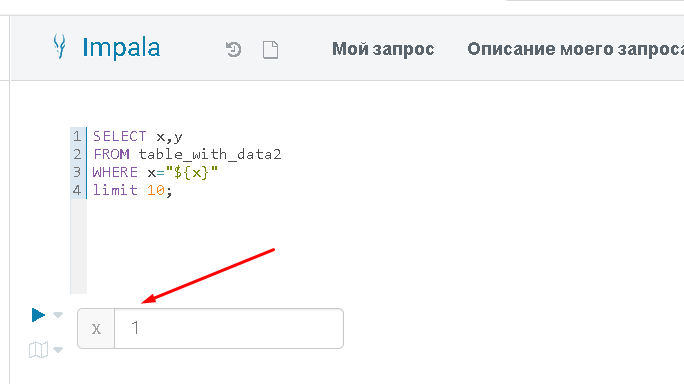


Рисунок 4‑11 Пример запроса

Под редактором запроса располагается лог выполнения запроса. Если запрос выполнен без ошибок, то лог окрашен в серый цвет (**Рисунок 4‑12**)

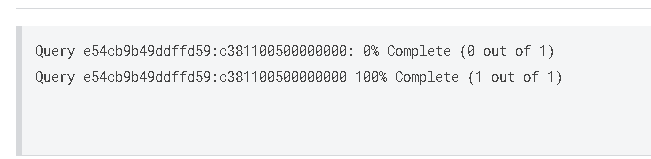


Рисунок 4‑12 Состояние выполнения запроса

В зависимости от запроса в строке состояния может отображаться разный статус исполнения. При выполнении запроса для отображения данных, без вычислений и условий на ограничение строк результата - будет отображаться статус сканирования файлов, для отображения 100 результатов. В этом случае будет запись как на рисунке (**Рисунок 4‑13**)

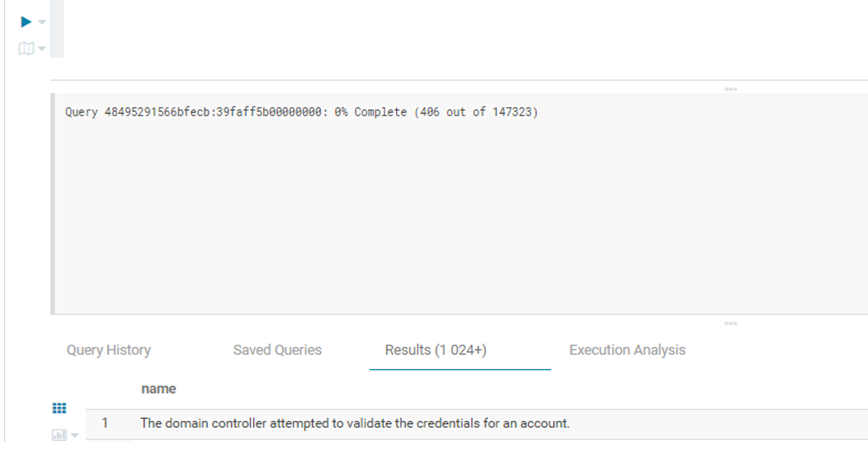


Рисунок 4‑13 Статус выполнения запроса

В этом случае запрос считается выполненным, но не закрытым, так как возвращено только 100 строк результата. При прокрутке окна с результатом до 101 строки будет выполнено дополнительное сканирование файлов для получения еще 100 строк. Также при выгрузке результата подобного запроса в файл будет выполнено дополнительное сканирование для получения 100 тыс. строк. – это значение является максимальным для выгрузки.

Пример выполнения запроса с ограничением (**Рисунок 4‑14**)

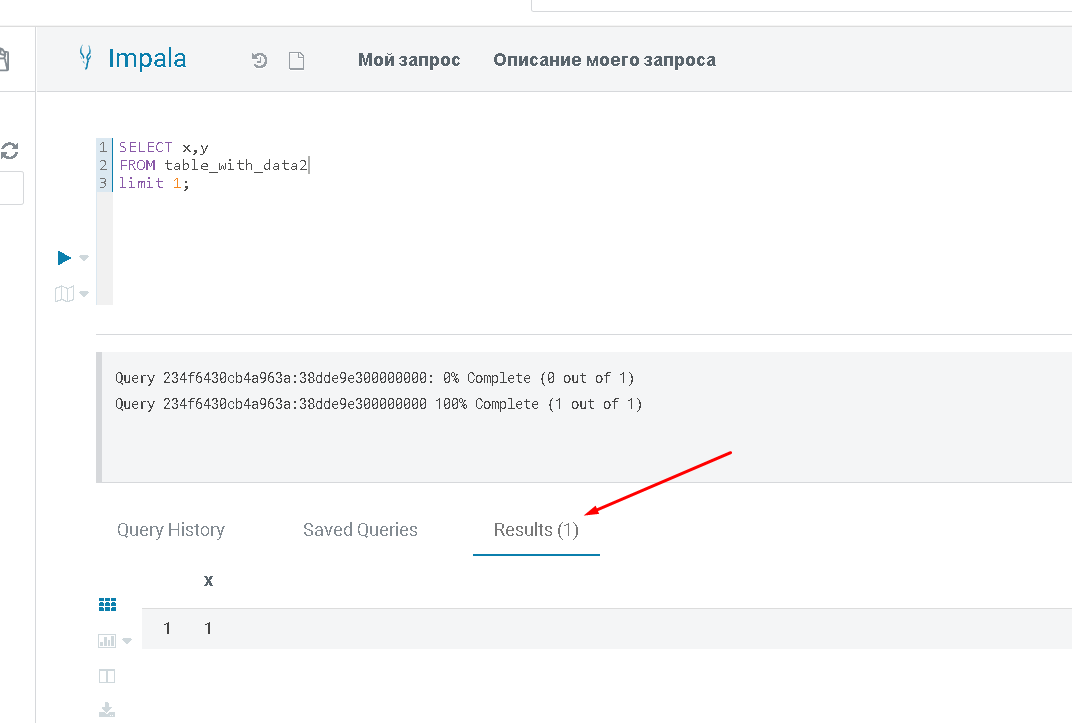


Рисунок 4‑14 Выполнение запроса с ограничением

В этом случае запрос выполнен на 100%, так как получено запрашиваемое количество строк. Для получения этого результата был просканирован только 1 файл, поэтому состояние сканирования отображается как 0%.

Здесь же отображаются ошибки в случае некорректного выполнения запроса (Рисунок 4‑15)

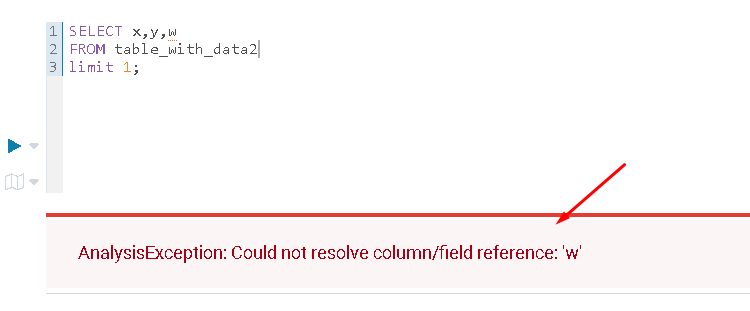


Рисунок 4‑15 Отображение ошибок при выполнении запроса

История всех выполненных запросов хранится в разделе «Query History». Для повторения выполнения запроса необходимо щелкнуть по запросу левой кнопкой мыши (**Рисунок 4‑16**)



Рисунок 4‑16 История выполнения запросов пользователя

После выполнение запросов его результаты можно просматривать еще некоторое время, без выполнения повторного запроса и времени ожидания на его исполнение. Запросить повторный результат можно только для тех запросов, у которых рядом с временем выполнения стоит значок галки.

## Результаты выполнения запроса

Результаты выполнения запроса приведены на вкладке «Результаты» («Results»), расположенной под окном лога хода выполнения запроса. Данный режим можно развернуть на всю страницу («Expand results»). Результаты запроса можно просмотреть в различном виде:

1. Таблица («Grid») «». Здесь данные можно отсортировать, кликнув по заголовку столбца сортировки (Рисунок 4‑17)

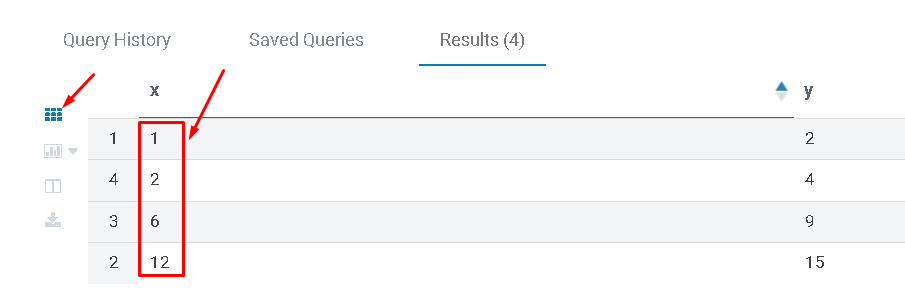


Рисунок ‑ Результаты запроса

Для просмотра информации о строке, кликните на значок расширения слева от необходимой строки, откроется окно просмотра информации о строке (**Рисунок 4‑18** и **Рисунок 4‑19**)

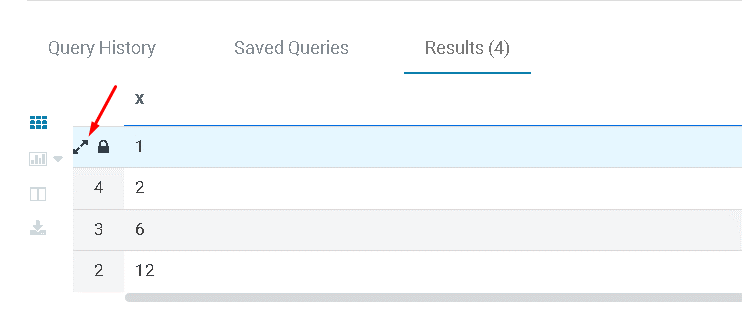


Рисунок 4‑18 Описание строки

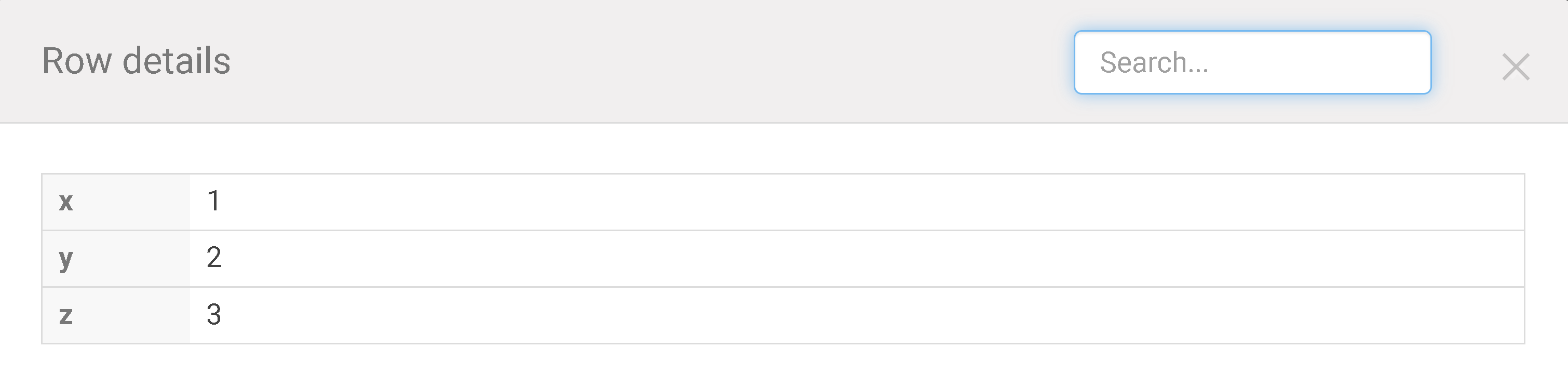


Рисунок 4‑19 Строка результата

Для блокировки выбранной строки воспользуйтесь кнопкой блокировки строки («Lock this row»), расположенной в начале каждой строки. При этом заблокированная строка переместится вверх таблицы и будет подкрашена в желтый цвет (**Рисунок 4‑20**)

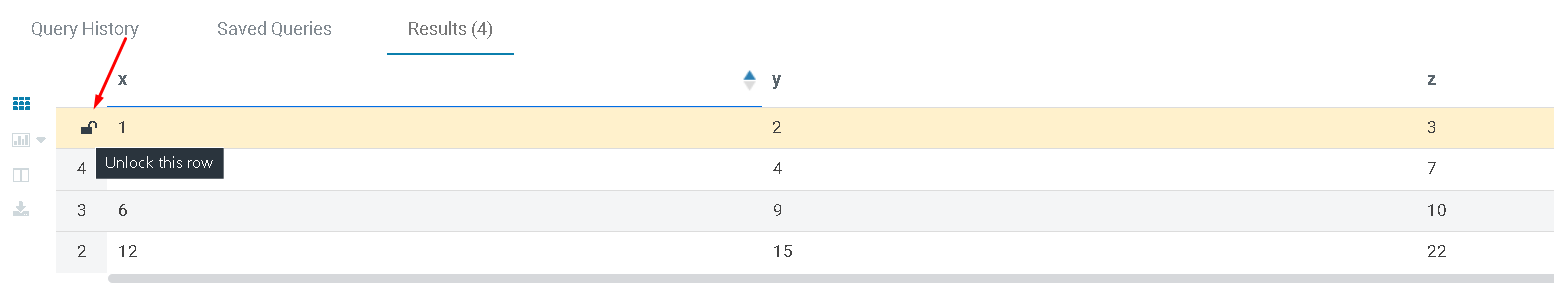


Рисунок 4‑20 Закрепление строк

Для разблокировки строки воспользуйтесь обратной кнопкой разблокировки строки («Unlock this row»). При этом строка вернется на свое место в первоначальном результате.

1. График («Bars») «» - здесь можно просмотреть результаты запроса в виде графика («Line» или «Bars»), выбрать отображаемые колонки, настроить группировку и сортировку данных (Рисунок 4‑21)

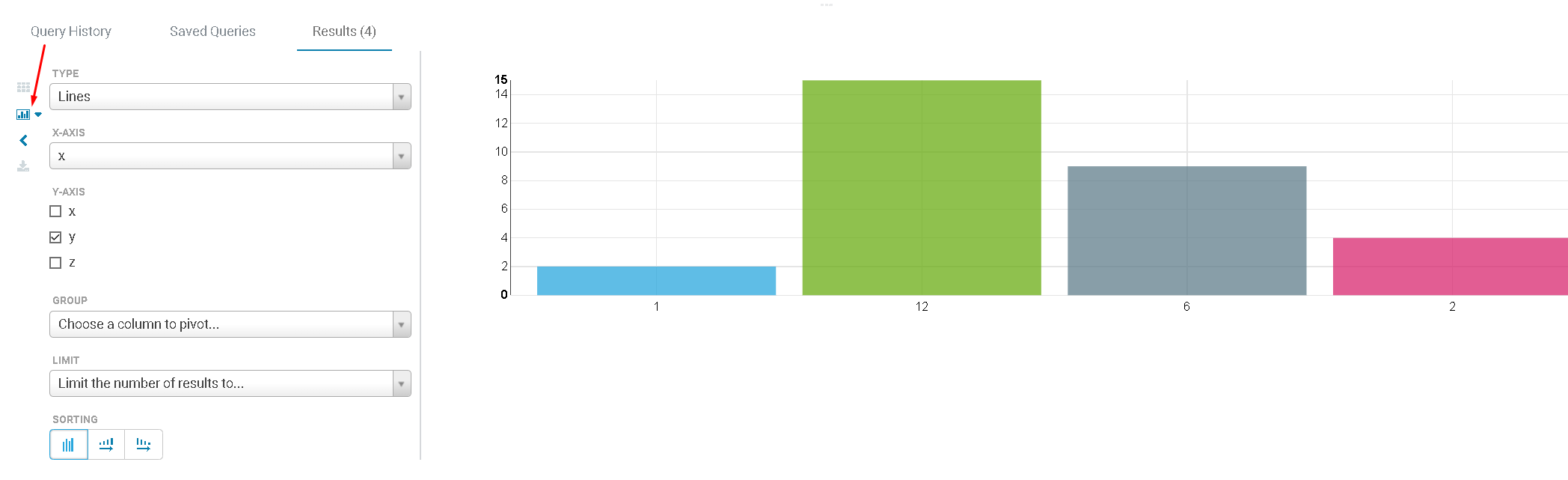


Рисунок 4‑21 Результаты запроса в виде графика

1. Поля («Columns») – здесь можно выбрать необходимые поля для просмотра результата выполнения запроса (**Рисунок 4‑22**)

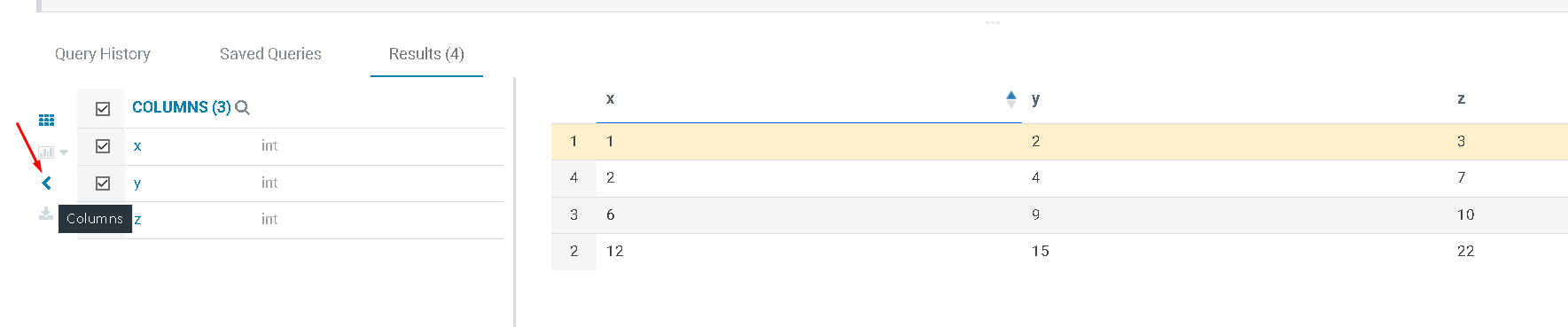


Рисунок 4‑22 Выбор отображаемых полей

1. Экспортировать результаты выборки («Export results») «» в файлы различных форматов (**Рисунок 4‑23**)

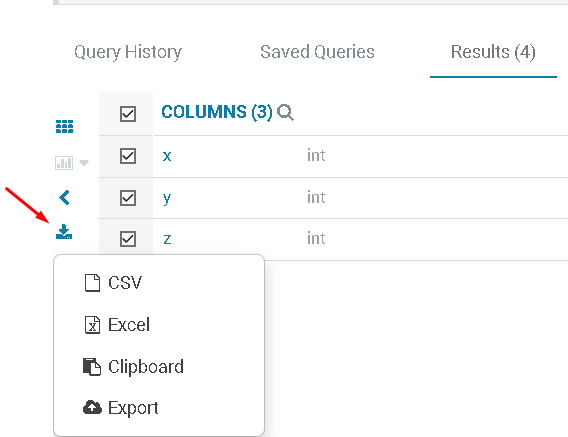


Рисунок 4‑23 Окно выгрузки результатов запроса

Выполните любое из следующих действий, чтобы загрузить или сохранить результаты запроса:

1. Нажмите «Загрузить как CSV», чтобы загрузить результаты в файл значений, разделенных запятыми, который можно использовать в других приложениях.
2. Нажмите «Загрузить как XLS», чтобы загрузить результаты в файл листа Microsoft Office Excel.
3. Нажмите «Сохранить», чтобы сохранить результаты в виде таблицы или файла HDFS.
4. Чтобы сохранить результаты в новой таблице, выберите «В новой таблице», введите имя таблицы и нажмите «Сохранить».
5. Чтобы сохранить результаты в файле HDFS, выберите «В каталоге HDFS», введите путь и нажмите «Сохранить». Затем вы можете скачать файл с помощью браузера файлов.

## Панель листинга файлов в HDFS через Hue

На боковой панели слева, под логотипом Hue, есть раздел «Files». При выборе этого раздела пользователь может просматривать директории и файлы в HDFS (если у пользователя есть соответствующие права). (Рисунок 4‑24)

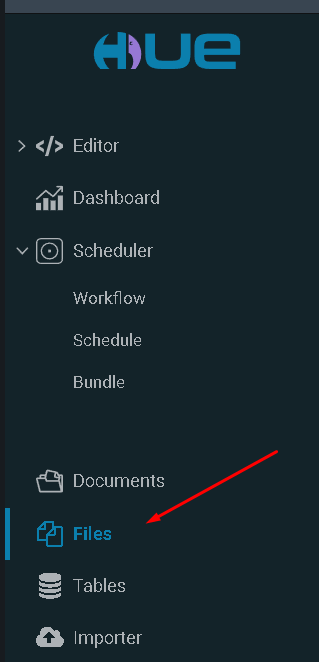


Рисунок ‑ Раздел Files

В разделе «Files» отображены директории, существующие в HDFS (**Рисунок 4‑25**)

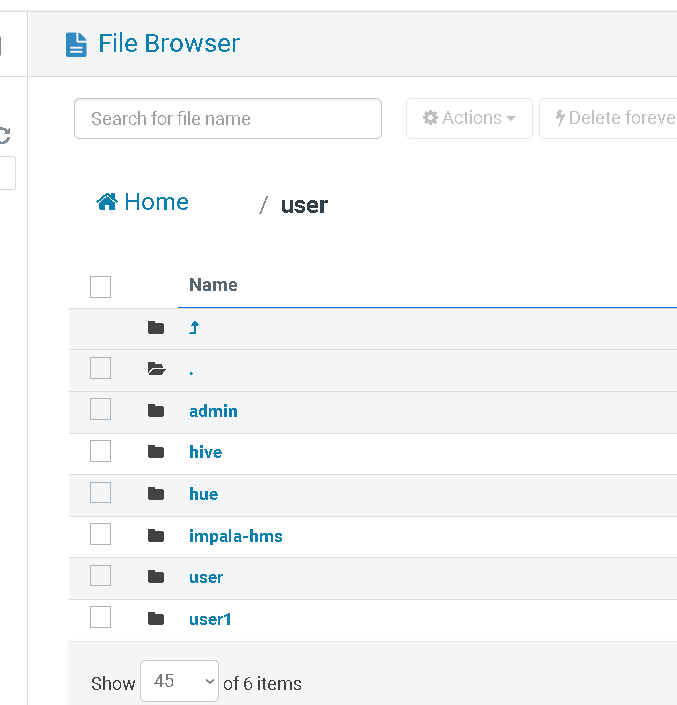


Рисунок 4‑25Листинг директорий в HDFS

Файлы в директориях можно просматривать и редактировать, при нажатии на соответствующий файл (если файл доступен для редактирования и отображения, например, текстовый формат файла) (**Рисунок 4‑26**)

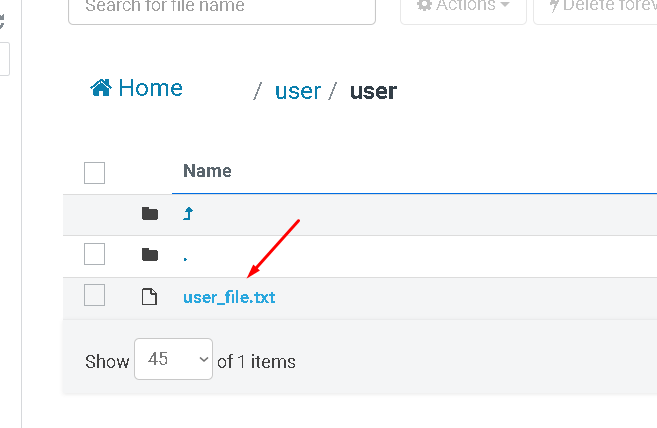


Рисунок 4‑26 Просмотр файлов в HDFS

Нажав на файл - отобразится его содержание (**Рисунок 4‑27**)

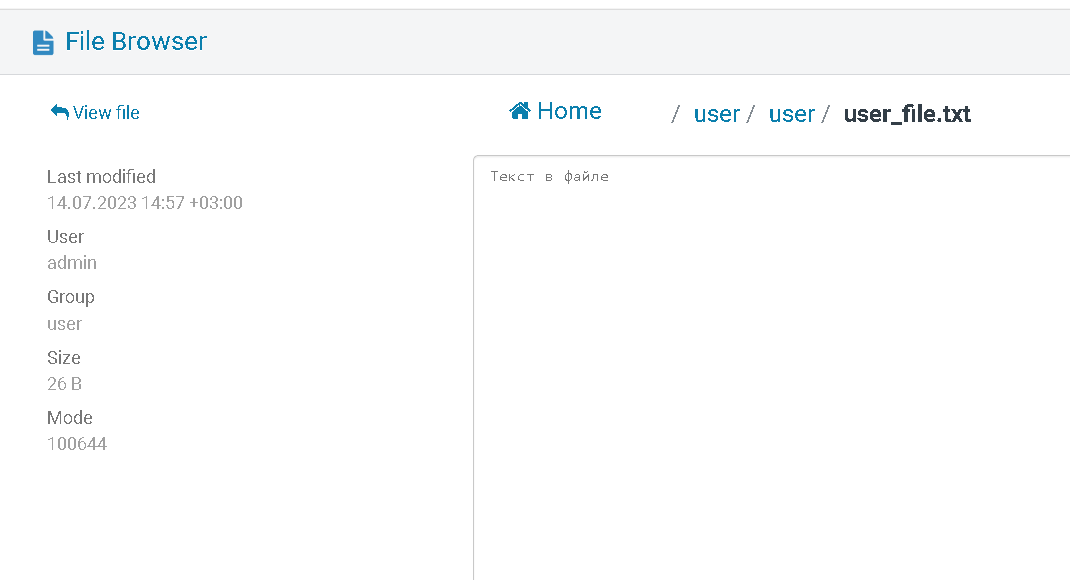


Рисунок 4‑27 Просмотра содержания файла

Существует возможность выполнять действия с файлами, предоставленные на рисунке (**Рисунок 4‑28**)

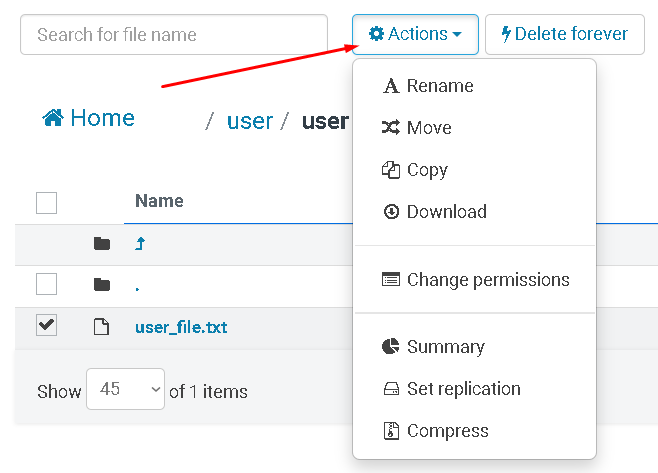


Рисунок 4‑28 Действия с файлами

Существует возможность выполнять загрузку файлов в HDFS, создавать новые файлы и директории (**Рисунок 4‑29**)

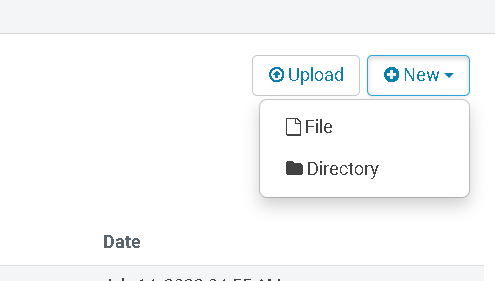


Рисунок 4‑29 Примеры действий с файлами

# Приложение A (обязательное) Схема данных MP SIEM

Схема полей таблиц MP SIEM представлен в таблице ниже (Таблица А.1).

Таблица А.1 – Схема полей таблиц MP SIEM

| Имя поля | Тип поля |
| --- | --- |
| 1 | 2 |
| \_checkpoint | long |
| action | string |
| agent\_id | string |
| aggregation\_name | string |
| asset\_ids | string |
| assets | string |
| assigned\_dst\_host | string |
| assigned\_dst\_ip | string |
| assigned\_dst\_port | long |
| assigned\_src\_host | string |
| assigned\_src\_ip | string |
| assigned\_src\_port | long |
| attacking\_assets | string |
| body | string |
| category\_generic | string |
| category\_high | string |
| category\_low | string |
| chain\_id | string |
| corrections | string |
| correlation\_name | string |
| correlation\_type | string |
| count | long |
| count\_bytes | long |
| count\_bytes\_in | long |
| count\_bytes\_out | long |
| count\_packets | long |
| count\_packets\_in | long |
| count\_packets\_out | long |
| count\_subevents | long |
| datafield1 | string |
| datafield10 | string |
| datafield2 | string |
| datafield3 | string |
| datafield4 | string |
| datafield5 | string |
| datafield6 | string |
| datafield7 | string |
| datafield8 | string |
| datafield9 | string |
| detect | string |
| direction | string |
| dst\_asset | string |
| dst\_fqdn | string |
| dst\_geo\_asn | long |
| dst\_geo\_city | string |
| dst\_geo\_country | string |
| dst\_geo\_org | string |
| dst\_host | string |
| dst\_hostname | string |
| dst\_ip | string |
| dst\_mac | string |
| dst\_port | long |
| duration | long |
| event\_src\_asset | string |
| event\_src\_category | string |
| event\_src\_fqdn | string |
| event\_src\_host | string |
| event\_src\_hostname | string |
| event\_src\_id | string |
| event\_src\_ip | string |
| event\_src\_subsys | string |
| event\_src\_title | string |
| event\_src\_vendor | string |
| event\_type | string |
| external\_link | string |
| generator | string |
| generator\_type | string |
| generator\_version | string |
| historical | boolean |
| id | string |
| importance | string |
| incident\_aggregation\_closed\_behavior | string |
| incident\_aggregation\_key | string |
| incident\_aggregation\_time\_window | string |
| incident\_aggregation\_timeout | long |
| incident\_assigned\_to\_user\_id | string |
| incident\_attacking\_addresses | string |
| incident\_attacking\_assets | string |
| incident\_category | string |
| incident\_description | string |
| incident\_name | string |
| incident\_related\_addresses | string |
| incident\_related\_assets | string |
| incident\_severity | string |
| incident\_severity\_behavior | string |
| incorrect\_time | boolean |
| input\_id | string |
| interface | string |
| job\_id | string |
| logon\_auth\_method | string |
| logon\_service | string |
| logon\_type | long |
| mime | string |
| msgid | string |
| nas\_fqdn | string |
| nas\_ip | string |
| normalized | boolean |
| object | string |
| object\_account\_contact | string |
| object\_account\_dn | string |
| object\_account\_domain | string |
| object\_account\_fullname | string |
| object\_account\_group | string |
| object\_account\_id | string |
| object\_account\_name | string |
| object\_account\_privileges | string |
| object\_account\_session\_id | string |
| object\_domain | string |
| object\_fullpath | string |
| object\_group | string |
| object\_hash | string |
| object\_id | string |
| object\_name | string |
| object\_path | string |
| object\_process\_cmdline | string |
| object\_process\_cwd | string |
| object\_process\_fullpath | string |
| object\_process\_guid | string |
| object\_process\_hash | string |
| object\_process\_id | string |
| object\_process\_meta | string |
| object\_process\_name | string |
| object\_process\_original\_name | string |
| object\_process\_parent\_cmdline | string |
| object\_process\_parent\_fullpath | string |
| object\_process\_parent\_guid | string |
| object\_process\_parent\_hash | string |
| object\_process\_parent\_id | string |
| object\_process\_parent\_name | string |
| object\_process\_parent\_path | string |
| object\_process\_path | string |
| object\_process\_version | string |
| object\_property | string |
| object\_query | string |
| object\_state | string |
| object\_type | string |
| object\_value | string |
| object\_vendor | string |
| object\_version | string |
| origin\_app\_id | string |
| original\_time | timestamp |
| primary\_siem\_app\_id | string |
| protocol | string |
| protocol\_layer7 | string |
| reason | string |
| recv\_asset | string |
| recv\_host | string |
| recv\_ipv4 | string |
| recv\_ipv6 | string |
| recv\_time | timestamp |
| remote | boolean |
| scope\_id | string |
| siem\_id | string |
| site\_id | string |
| src\_asset | string |
| src\_fqdn | string |
| src\_geo\_asn | long |
| src\_geo\_city | string |
| src\_geo\_country | string |
| src\_geo\_org | string |
| src\_host | string |
| src\_hostname | string |
| src\_ip | string |
| src\_mac | string |
| src\_port | long |
| start\_time | timestamp |
| status | string |
| subevents | string |
| subevents\_time | string |
| subject | string |
| subject\_account\_contact | string |
| subject\_account\_dn | string |
| subject\_account\_domain | string |
| subject\_account\_fullname | string |
| subject\_account\_group | string |
| subject\_account\_id | string |
| subject\_account\_name | string |
| subject\_account\_privileges | string |
| subject\_account\_session\_id | string |
| subject\_domain | string |
| subject\_group | string |
| subject\_id | string |
| subject\_name | string |
| subject\_privileges | string |
| subject\_process\_cmdline | string |
| subject\_process\_cwd | string |
| subject\_process\_fullpath | string |
| subject\_process\_guid | string |
| subject\_process\_hash | string |
| subject\_process\_id | string |
| subject\_process\_meta | string |
| subject\_process\_name | string |
| subject\_process\_original\_name | string |
| subject\_process\_parent\_cmdline | string |
| subject\_process\_parent\_fullpath | string |
| subject\_process\_parent\_guid | string |
| subject\_process\_parent\_hash | string |
| subject\_process\_parent\_id | string |
| subject\_process\_parent\_name | string |
| subject\_process\_parent\_path | string |
| subject\_process\_path | string |
| subject\_process\_version | string |
| subject\_state | string |
| subject\_type | string |
| subject\_version | string |
| tag | string |
| task\_id | string |
| taxonomy\_version | string |
| tcp\_flag | string |
| tenant\_id | string |
| time | timestamp |
| type | string |
| uuid | string |
| IndexName | string |
| sparkTS | timestamp |
| sparkpart | long |
| rt | timestamp |

Описание добавленных полей с префиксом Spark представлен в таблице ниже (Таблица А.2).

Таблица А.2 – Описание добавленных полей с префиксом Spark

| Наименование поля | Тип | Описание |
| --- | --- | --- |
| SparkPart | Long | Используется для партиционирования паркет файлов. Формируется из поля rt. Вычисляет начало часа для rt и записывает в поле количество секунд до начала часа |
| sparkTs | Timestamp | Время начала парсинга строки |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Таблица регистрации изменений** | | | | | | | | |
| Изм. | Номера листов (страниц) | | | | Всего  листов  (страниц)  в док. | Номер док. | Подп. | Дата |
| изме-ненных | заменен-ных | новых | аннулиро-ванных |
|  |  |  |  |  |  |  |  |  |