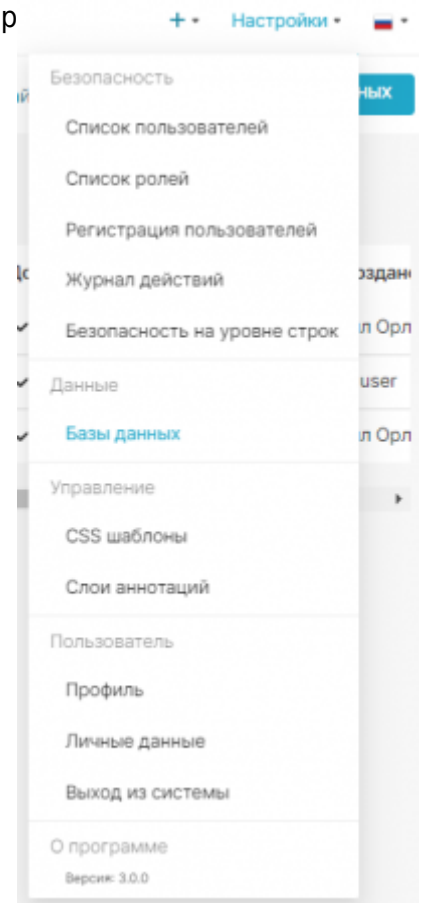
Содержание

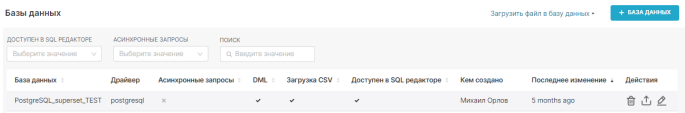
Базы данных (Databases) 2
Подключение базы данных 3
Шаг 1. Создание новой базы данных. 3
Шаг 2. Выбор базы данных. 3
Шаг 3. Настройка базы данных. 4
Для баз данных Postgres и SQLite нужно ввести следующие параметры: 4
Для других поддерживаемых баз данных введите следующие параметры: 4
Шаг 4. Продвинутая настройка 5
Шаг 5. После всех настроек нажмите кнопку Завершить. 6
Импорт из файлов 6
Импорт данных из CSV в базу данных 6
Импорт данных из Excel в базу данных 8
Импорт данных столбчатого формата 10

Базы данных (Databases)

Базы данных – упорядоченный набор структурированной информации или данных, которые обычно хранятся в электронном виде. База данных обычно управляется системой управления базами данных (СУБД). Раздел настройки подключения к разным базам данных находится **Настройки → Данные → Базы данных**.



Интерфейс раздела **Базы данных** выглядит следующим образом:



Справа расположены кнопки для добавления базы данных и таблицы из файлов (CSV, Excel, столбчатый формат).
Ниже - окно **поиска** и окна для **фильтрации** загруженных баз данных:

- возможно ли использовать базу данных в SQL редакторе;
- возможно ли использовать асинхронный режим работы.

Для загруженных баз данных отображаются следующие атрибуты:

- **База данных** – имя базы данных (как вы ее назвали при

добавлении);
- **Драйвер** – используемая система управления базами данных;
- **Асинхронные запросы (AQE, Asynchronous Query Execution)** – поддерживается ли асинхронный режим работы;
- **DML (Data Manipulation Language)** – поддерживается ли язык манипулирования данными (вставка, изменение, удаление);
- **Загрузить CSV** – поддерживается ли загрузка CSV-файлов;
- **Доступен в SQL редакторе** – разрешено ли использовать в SQL редакторе;
- **Кем создано** – кем создана база данных;
- **Последнее изменение** – когда было сделано последнее изменение;
- **Действия** – доступные действия, которые можно произвести с базой данных:
    - Удалить;
    - Экспортировать;
    - Редактировать.

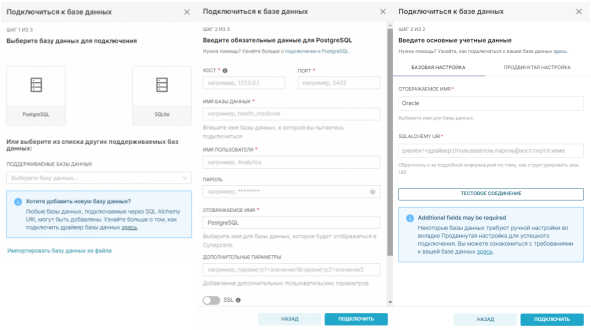
Подключение базы данных

Шаг 1. Создание новой базы данных.

В разделе **Базы данных** нажимаем справа кнопку **+ БАЗА ДАННЫХ**.

Шаг 2. Выбор базы данных.

- Выбираем в диалоговом окне PostgreSQL или SQLite.
- Для подключения другой базы данных воспользуйтесь выпадающим списком ниже «Поддерживаемые базы данных».
- Или импортируйте свою базу данных из файла.



Шаг 3. Настройка базы данных.

Для баз данных Postgres и SQLite нужно ввести следующие параметры:

- **Хост** – IP-адрес или доменное имя;
- **Порт** – порт хоста (5432, по умолчанию для PostgreSQL);
- **Имя базы данных** – имя базы данных;
- **Имя пользователя** – пользователь СУБД, у которого выданы права на подключения к базе данных;
- **Пароль** – пароль пользователя;
- **Отображаемое имя** – имя подключения, отображаемое в списке подключения;
- **Дополнительные параметры** – добавление дополнительных пользовательских параметров;
- **SSL** – включает ssl-шифрование подключения (если поддерживается);
- **SSH Tunnel** – включает SSH-туннели.

Для других поддерживаемых баз данных введите следующие параметры:

- **Отображаемое имя** – имя подключения, отображаемое в списке подключения;
- **SQLAlchemy URI** в виде диалект+драйвер:

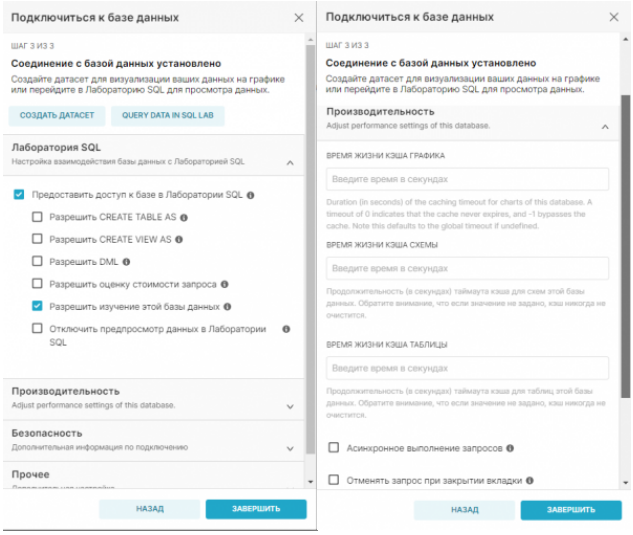
  ```
  //пользователь:пароль@хост:порт/схема
  ```

  , где

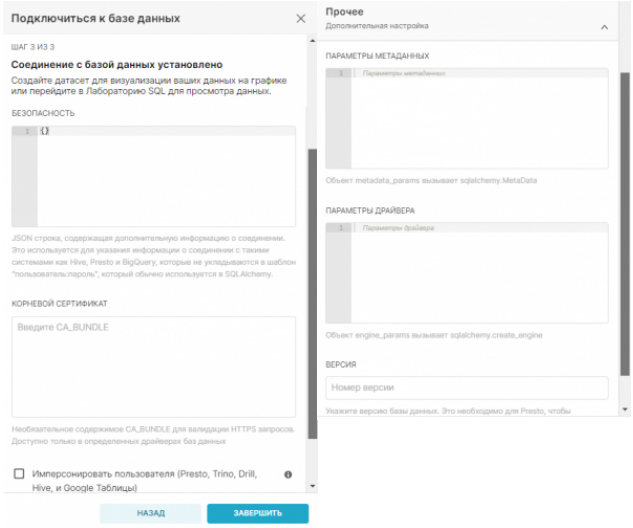
  - **Диалект+Драйвер** – например, Oracle;
  - **Пользователь** – пользователь СУБД, у которого выданы права на подключения к базе данных;
  - **Пароль** – пароль пользователя;
  - **Хост** – IP-адрес или доменное имя;
  - **Порт** – порт хоста (1521, по умолчанию для Oracle);
  - **Схема** – используемая схема базы данных.

Нажмите кнопку **Тестовое соединение**. При успешном соединении, нажмите внизу кнопку **Подключить**.

Шаг 4. Продвинутая настройка



- **Лаборатория SQL** – настройка взаимодействия базы данных с Лабораторией SQL:
  - **Предоставить доступ к базе в Лаборатории SQL** – разрешить запросы к этой базе дынных;
  - **Разрешить CREATE TABLE AS** – разрешить создавать таблицы на основе запросов;
  - **Разрешить CREATE VIEW AS** – разрешить создавать представления на основе запросов;
  - **Разрешить DML** – разрешить команды UPDATE, DELETE, CREATE и пр. над базой данных;
  - **Разрешить оценку стоимости запроса** – показывать кнопку подсчета стоимости запроса перед его выполнением (для Bigquery, Presto и Postgres);
  - **Разрешить изучение этой базы данных** – пользователям разрешено смотреть ответ на запрос к этой базе в Лаборатории SQL;
  - **Отключить предпросмотр данных в Лаборатории SQL** – отключить предпросмотр данных при извлечении метаданных таблицы в Лаборатории SQL (полезно для избежания проблем с производительностью браузера при использовании баз данных с очень широкими таблицами);



- **Производительность** – параметры производительности для базы данных:
  - **Время жизни кэша графика** – длительность (сек.) таймаута кэша для графиков, использующих эту базу данных;
  - **Время жизни кэша схемы** – длительность (сек.) таймаута кэша для схем, использующих эту базу данных;
  - **Время жизни кэша таблицы** – длительность (сек.) таймаута кэша для таблиц, использующих эту базу данных;

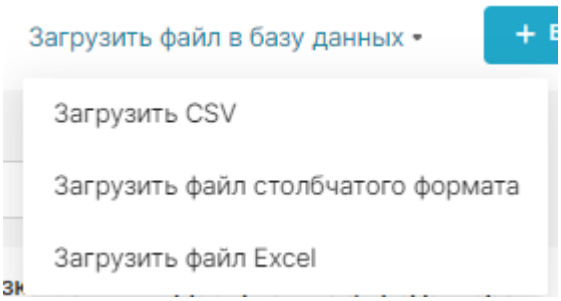
    - **Асинхронное выполнение запросов** – работа с базой данных в асинхронном режиме означает, что запросы выполняются на удаленных серверах, а не на веб-сервере Superset (подразумевается, что у вас есть установка с Celery);
    - **Отменять запрос при закрытии вкладки** – завершать выполнение запросов после закрытия браузерной вкладки или после того, как пользователь переключится на другую вкладку (доступно для Presto, Hive, MySQL, Postgres, Snowflake);
- **Безопасность** – дополнительная информация по подключению:
    - **Безопасность** – JSON-строка, содержащая дополнительную информацию о соединении;
    - **Корневой сертификат**;
    - **Имперсонировать пользователя**;
    - **Разрешить загрузку файлов в базу данных**;
- **Прочее**:
    - **Параметры метаданных**;
    - **Параметры драйвера**;
    - **Версия**.

Шаг 5. После всех настроек нажмите кнопку Завершить.

Импорт из файлов

Импорт данных из CSV в базу данных

В разделе **Базы данных**, нажимаем справа на выпадающий список **Загрузить файл в базу данных** (рядом с кнопкой **+ БАЗА ДАННЫХ**). Из списка выбираем **Загрузить CSV**.



Далее необходимо заполнить следующее:

- **Загрузка CSV** - прикрепите сюда CSV-файл, который будет загружен в базу данных;

- **Имя таблицы** - имя таблицы, которая будет сформирована из данных csv;
- **База данных** – база данных, в которую будет добавляться таблица;
- **Схема** – схема, в которую будет добавлена таблица (если это поддерживается базой данных);
- **Разделитель** – разделитель, используемый в CSV-файле;
- **File Settings**:
    - **Если таблица уже существует** – что должно произойти, если таблица уже существует: Ошибка (Fail) – ничего не делать, Заменить (Replace) – удалить и заново создать таблицу или Добавить (Append) - добавить данные;
    - **Пропуск начального пробела** – пропустить пробелы после разделителя;
    - **Пропуск пустых строк** – пропустите пустые строки, а не интерпретировать их как значения NaN;
    - **Список столбцов, которые должны быть интерпретированы как даты** – разделённый запятыми список столбцов, которые должен быть интерпретированы как даты;
    - **Автоматически интерпретировать форматы даты и времени** – автоматическая интерпретация формата даты и времени;
    - **Day First** – формат даты, где сначала день, потом месяц (международный и Европейский формат);
    - **Десятичный разделитель** - символ, который интерпретируется как десятичная точка;
    - **Пустые значения** – JSON-список значений, который нужно интерпретировать как Пусто (null);
- **Columns**:
    - **Индексный столбец** – столбец для использования в качестве меток строк данных. Оставьте

пустым, если столбец индекса отсутствует;
    - **Индекс датафрейма** – сделать индекс датафрейма столбцом;
    - **Метка(и) столбца(ов)** – метка для индексного(ых) столбца(ов). Если не задано и задан индекс датафрейма, будут использованы имена индексов;
    - **Столбцы для чтения** – JSON-список имен столбцов, которые будут использоваться;
    - **Перезаписать повторяющиеся столбцы** – если повторяющиеся столбцы не перезаписываются, они будут представлены в формате «X.0, X.1»;
    - **Расширенный тип данных** – словарь с именами столбцов и их тип данных, на который нужно изменить.

      ```
      Например, {‘user_id’: ‘integer’};
      ```

- **Rows**:
    - **Строка заголовка** – строка, содержащая заголовки для использования в качестве имен столбцов (0 - первая строка данных). Оставьте пустым, если строка заголовка отсутствует;
    - **Строки для чтения** – количество строк файла для чтения;
    - **Пропуск строк** – количество первых строк, которые нужно пропустить.

После заполнения необходимых параметров, нажмите внизу кнопку **Сохранить**.

Импорт данных из Excel в базу данных

В разделе **Базы данных**, нажимаем справа на выпадающий список **Загрузить файл в базу данных** (рядом с кнопкой **+ БАЗА ДАННЫХ**). Из списка выбираем **Загрузить файл Excel**.

Далее необходимо заполнить следующие

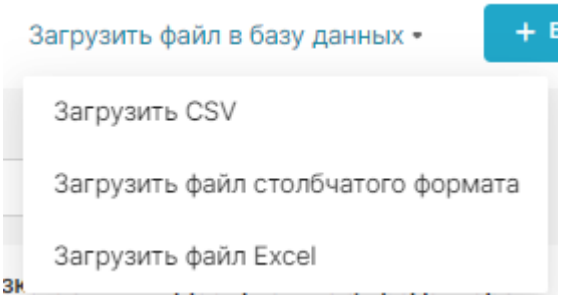
параметры:

- **Имя таблицы** – имя таблицы, которая будет сформирована из данных Excel;
- **Excel-файл** – прикрепите сюда Excel-файл, который будет загружен в БД;
- **Имя листа** – имя листа (по умолчанию первый лист);
- **База данных** – база данных, в которую будет добавляться таблица;
- **Схема** – схема, в которую будет добавлена таблица (если это поддерживается базой данных);
- **Таблица существует** – что должно произойти, если таблица уже существует: Ошибка (Fail) – ничего не делать, Заменить (Replace) – удалить и заново создать таблицу или Добавить (Append) – добавить данные;
- **Строка заголовка** – строка, содержащая заголовки для использования в качестве имен столбцов (0 - первая строка данных); оставьте пустым, если строка заголовка отсутствует;
- **Индексный столбец** – столбец для использования в качестве меток строк данных; оставьте пустым, если столбец индекса отсутствует;
- **Управление повторяющимися столбцами** – обозначить повторяющиеся столбцы как «X.0, X.1»;
- **Пропуск строк** – количество первых строк, которые нужно пропустить;
- **Строки для чтения** - количество строк файла для чтения;
- **Парсинг дат** – разделённый запятыми список столбцов, которые должен быть интерпретированы как даты;
- **Десятичный символ** – символ, который интерпретируется как десятичная точка;
- **Индекс датафрейма** – записать индекс датафрейма, как отдельный столбец;
- **Метка(и) столбца(ов)** – обозначение столбца для столбцов с индексами. Если поле пустое, а настройка [Индекс] включена, то используются имена индексов;
- **Пустые значения** – JSON-список

значений, который нужно интерпретировать как Пусто (null).

После заполнения необходимых параметров, нажмите внизу кнопку **Сохранить**.

Импорт данных столбчатого формата



В разделе **Базы данных**, нажимаем справа на выпадающий список **Загрузить файл в базу данных** (рядом с кнопкой **+ БАЗА ДАННЫХ**). Из списка выбираем **Загрузить файл столбчатого формата**.

Далее необходимо заполнить следующие:

- **Имя таблицы** – имя таблицы, которая будет сформирована из данных;
- **Файл столбчатого формата** – прикрепите сюда файл, который будет загружен в базу данных;
- **База данных** – база данных, в которую будет добавляться таблица;
- **Схема** – схема, в которую будет добавлена таблица (если это поддерживается базой данных);
- **Таблица существует** – что должно произойти, если таблица уже существует: Ошибка (Fail) – ничего не делать, Заменить (Replace) – удалить и заново создать таблицу или Добавить (Append) – добавить данные;
- **Используемые столбцы** – JSON-список имен столбцов, которые будут использоваться. Например, [«id», «name», «gender», «age»]. Если ничего не указано, то все столбцы из файла будут добавлены;
- **Индекс датафрейма** – записать индекс датафрейма, как отдельный столбец;
- **Метка(и) столбца(ов)** – обозначение столбца для столбцов с индексами. Если поле пустое, а настройка [Индекс] включена, то используются имена индексов.

После заполнения необходимых параметров, нажмите внизу кнопку **Сохранить**.

From:
https://kb.nstu.ru/ - **База знаний НГТУ НЭТИ**

Permanent link:
https://kb.nstu.ru/superset:work_in_superset:databases:databases

Last update: **2023/11/15 11:30**