

# Infomaximum

«Инфомаксимум»  
(Общество с ограниченной ответственностью)

## Инструкция к Видеокурсу ETL

2023 г.

## Содержание

1. Создание пространства .....	2
2. Создание таблиц .....	3
3. Создание скрипта .....	6
3.1. Настройка блока «Планировщик» .....	7
3.2. Выбор блока-действия .....	9
3.3. Создание подключения .....	9
3.4. Добавление блока «Пространство: Добавить строки».....	11
3.5. Визуальный маппинг .....	12
3.6. Удаление данных из таблицы .....	13
3.7. Тестирование скрипта .....	14
3.8. Добавление блока «Пространство: SQL-запрос» в пространство .....	16
3.9. Добавление блока «Файл: Прочитать CSV» .....	16
3.10. Замена типов данных полей в таблице .....	17
3.11. Добавление блока «HTTP: Отправить запрос».....	18
3.12. Добавление данных в таблицу с помощью скрипта.....	19
4. Дублирование скрипта .....	20
5. Переименование скрипта.....	20
6. Изменение скрипта.....	21
7. Активация скрипта .....	25
8. Журнал выполнения скрипта .....	26
9. Подключение таблицы из базы .....	29
10. Экспорт скриптов .....	29
11. Удаление и импорт скрипта .....	30

# 1. Создание пространства

Чтобы начать работу для решения задачи по извлечению данных из системы контроля и управления доступа в офис, необходимо создать **Пространство**.

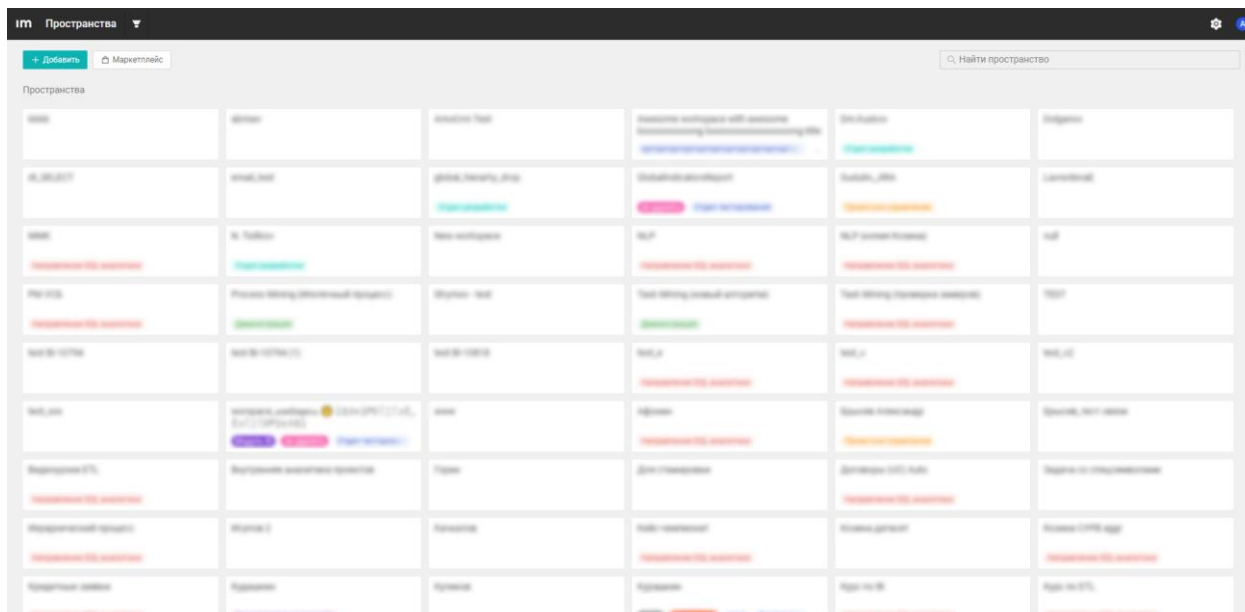


Рисунок 1 - Пространства

Чтобы создать пространство, кликните на кнопку **Добавить**. Открывается модальное окно создания пространства.

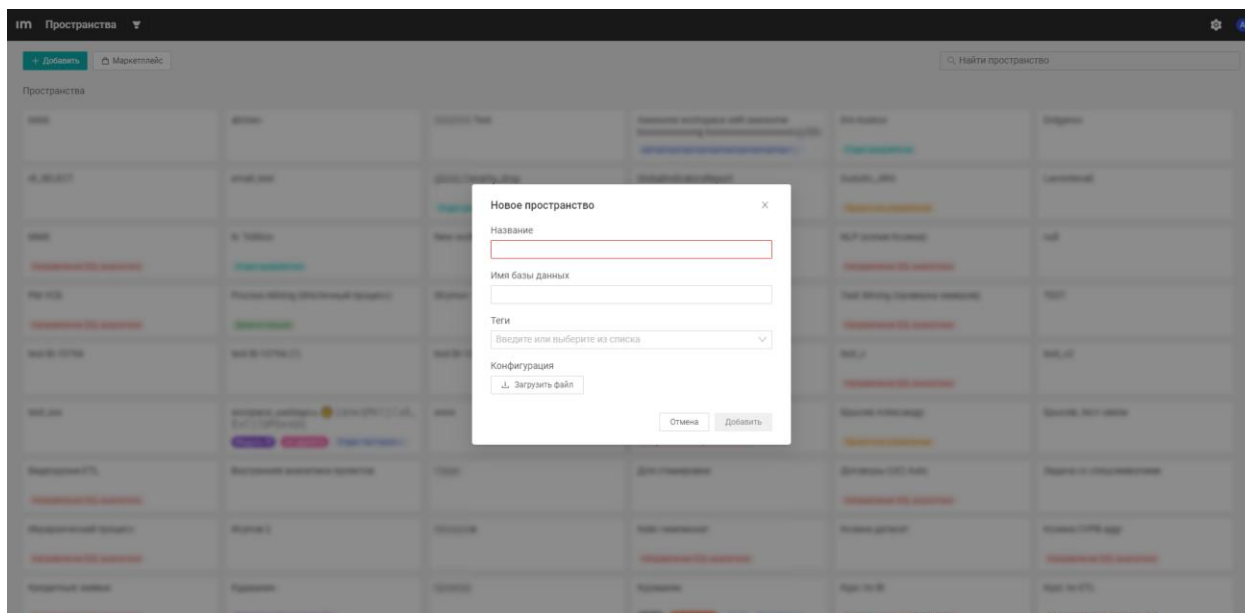


Рисунок 2 – Модальное окно создания пространства

1. Задайте:

- название;
- подключение к ClickHouse (выбор из списка доступных);
- имя базы данных;

- теги (выбор из списка доступных).
2. Нажмите **Добавить**.

## 2. Создание таблиц

Если доступен только один сервер базы данных, то поле **Подключение к ClickHouse** не отображается.

В поле **Имя базы данных** недопустимы кириллические символы. Если название базы уникальное, то система создаст новую. Если название повторяется, то система использует готовую базу данных.

После создания пространства, необходимо создать результирующие таблицы. Все доступные таблицы, как и другие элементы, отображаются в виде карточек во вкладке «*Панель управления*» в пространстве. Имя таблицы должно быть уникальным в рамках пространства.

Для создания новой таблицы нажмите **Добавить** — «*Таблицу*».

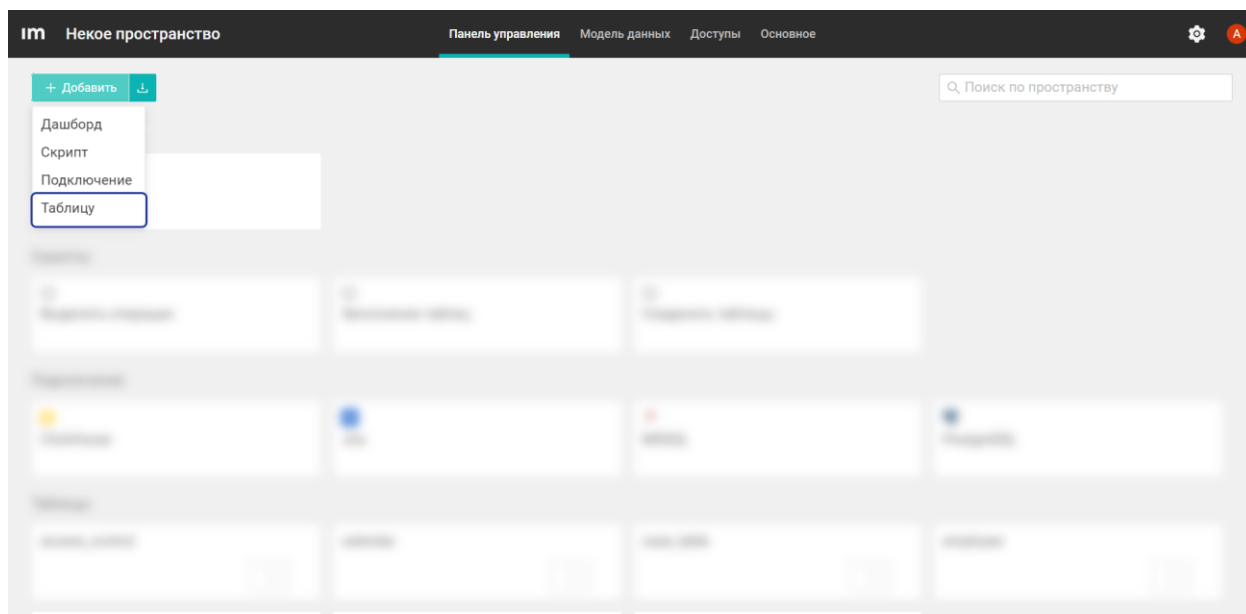


Рисунок 3 – Добавление таблицы

Формат таблиц – Csv-файл в UTF-8 кодировке без BOM. Файл обязательно должен содержать заголовок. В конце файла не должно быть пустых строк.

При создании новой таблицы открывается модальное окно, в котором можно ввести:

- Название таблицы;
- Названия колонок и тип данных;
- Партиционирование;
- Сортировку.

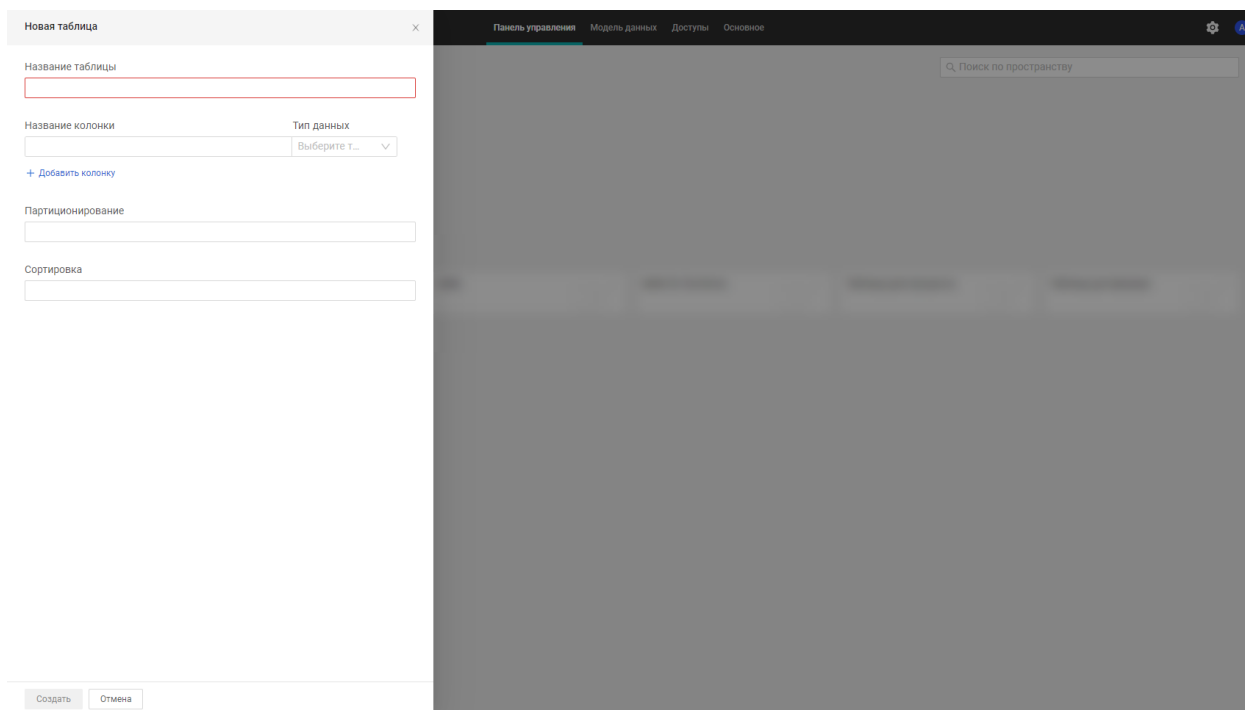


Рисунок 4 – Модальное окно создания таблицы

Создание таблицы доступно при выполнении условий:

- задано название таблицы;
- в таблице есть хотя бы одна колонка, для которой определены название и тип данных.

Первая таблица содержит сведения из СКУД. Колонки первой таблицы (access\_control):

- рассматриваемая дата (date - тип данных DateTime64);
- идентификатор сотрудника (employee\_id - тип данных Integer);
- время первого входа (start\_time - тип данных DateTime64);
- время выхода (finish\_time - тип данных DateTime64).

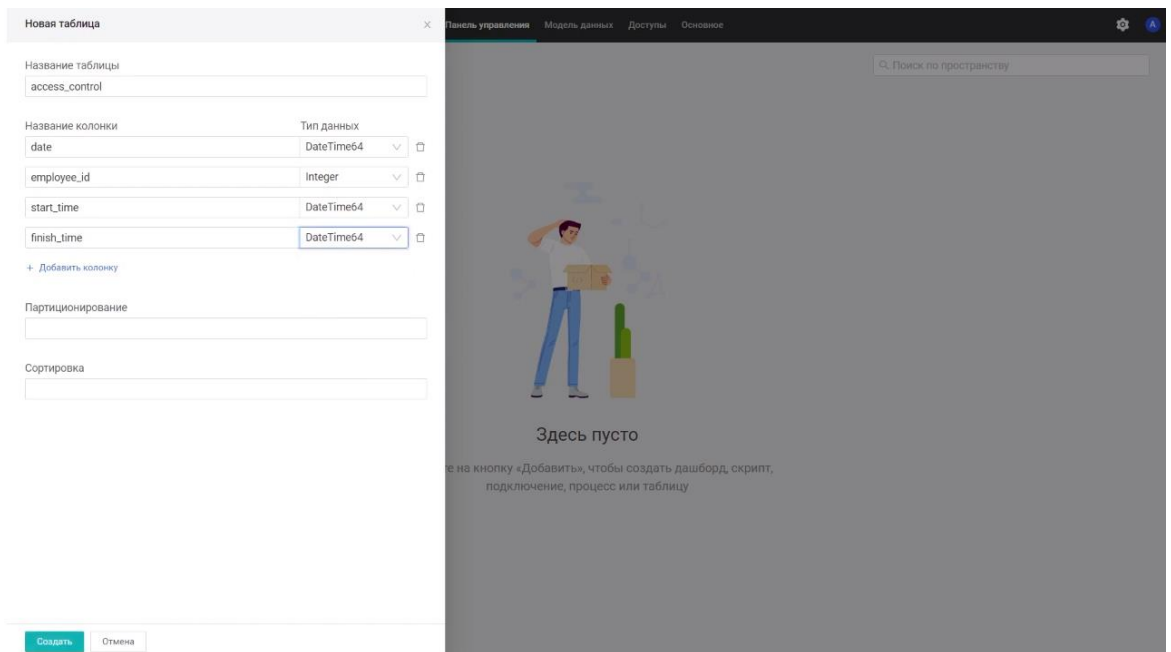


Рисунок 5 – Создание таблицы access\_control

Таблица со справочником сотрудников (employee):

- идентификатор сотрудника (id - тип данных Integer);
- ФИО сотрудника (name - тип данных String).

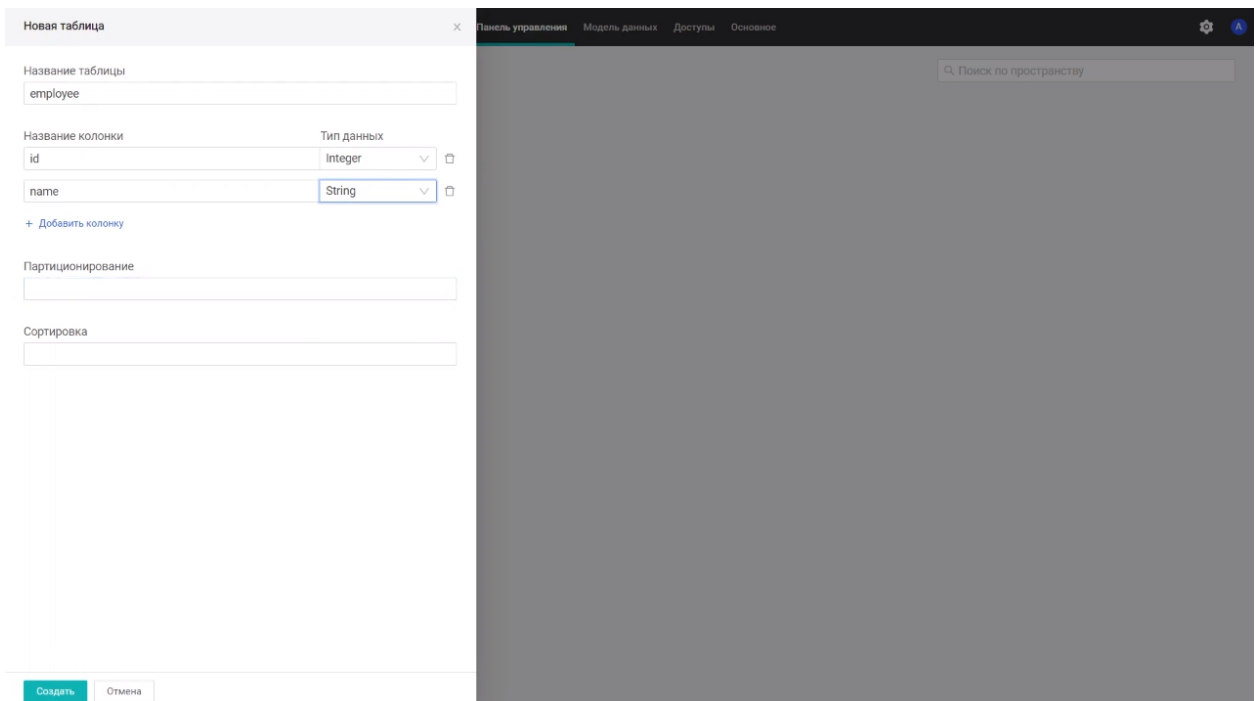


Рисунок 6 – Создание таблицы employee

Таблица для хранения производственного календаря (calendar):

- дата (date – так как источник данных возвращает дату в формате строки, тип данных String);
- тип дня (type - тип данных String).

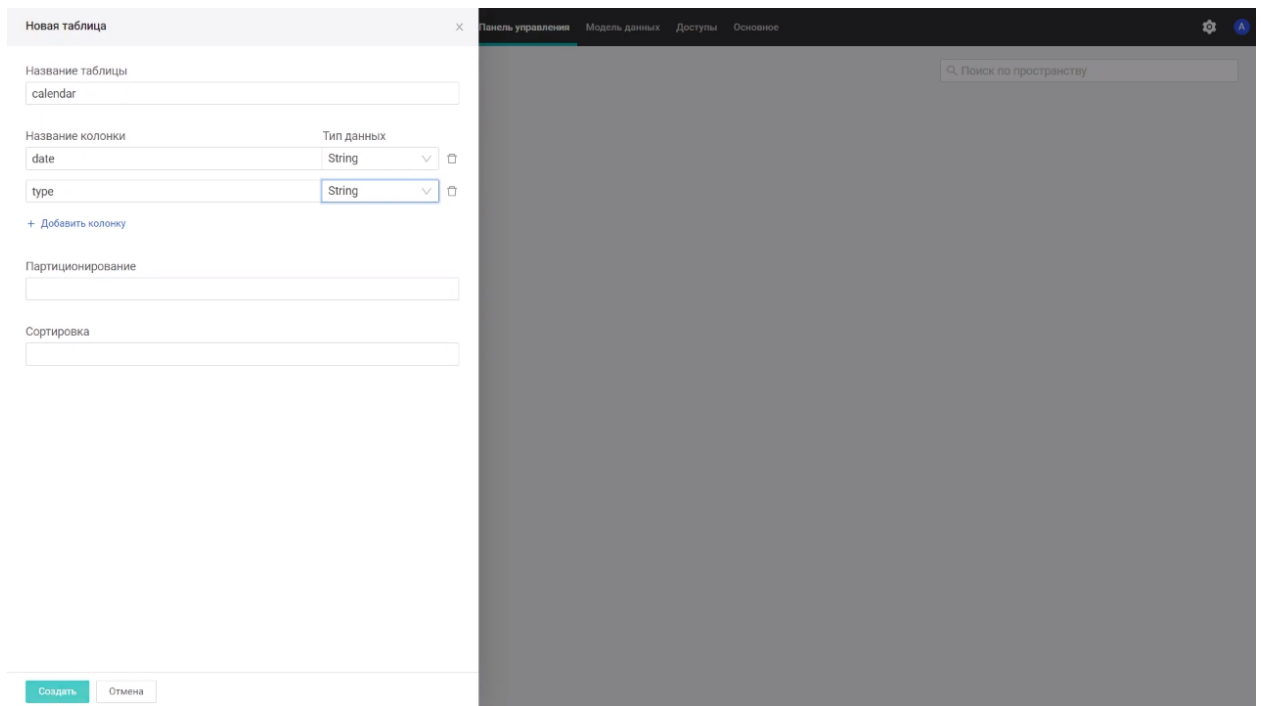


Рисунок 7 – Создание таблицы calendar

### 3. Создание скрипта

Далее необходимо создать скрипт для первичного заполнения таблиц. Чтобы создать скрипт, нажмите **Добавить**, выберите **Скрипт**.

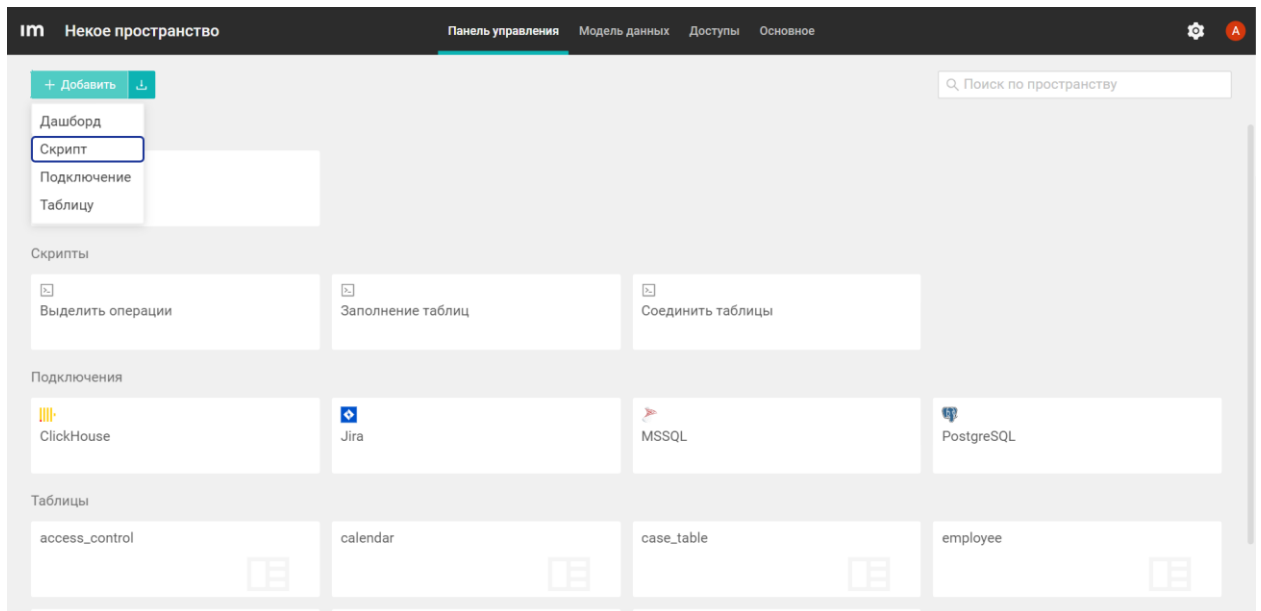


Рисунок 8 – Добавление скрипта

В открывшемся диалоговом окне укажите следующие параметры:

- название скрипта;
- описание (необязательно).

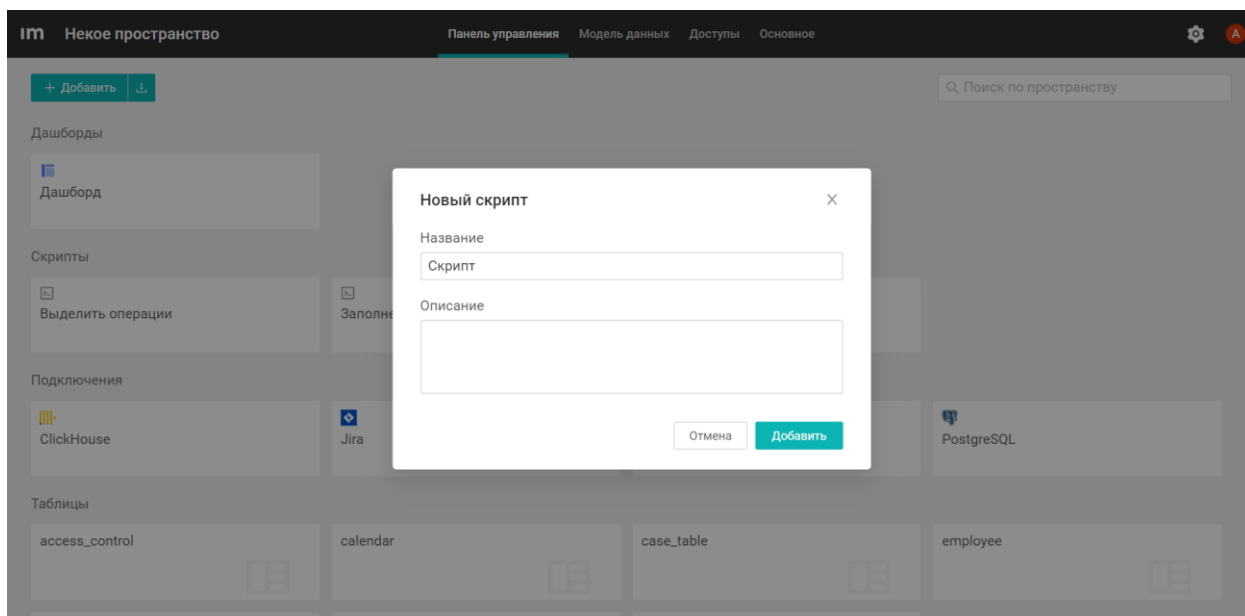


Рисунок 9 – Диалоговое окно (Новый скрипт)

### 3.1. Настройка блока «Планировщик»

После создания скрипт открывается в режиме редактирования и в нем автоматически создается блок «Выберите триггер».

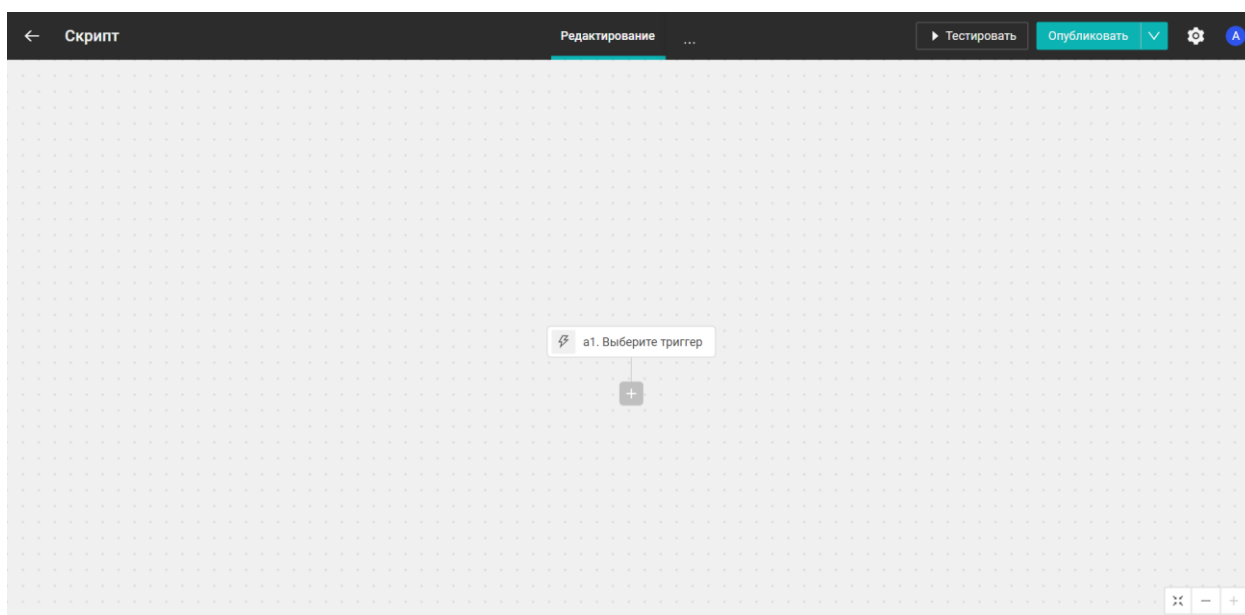


Рисунок 10 – Выберите триггер

Расписание выполнения скрипта задается через блок **Планировщик**. Этот блок может быть вставлен только в начало скрипта. Выберите вариант расписания в настройках блока:

- по минутам (по умолчанию 15 минут);
- по дням недели (нужно выбрать дни недели, часы и минуты);
- по датам (нужно выбрать месяцы, дни, часы, минуты);



- по CRON расписанию (система для автоматического запуска задач в определенное время).

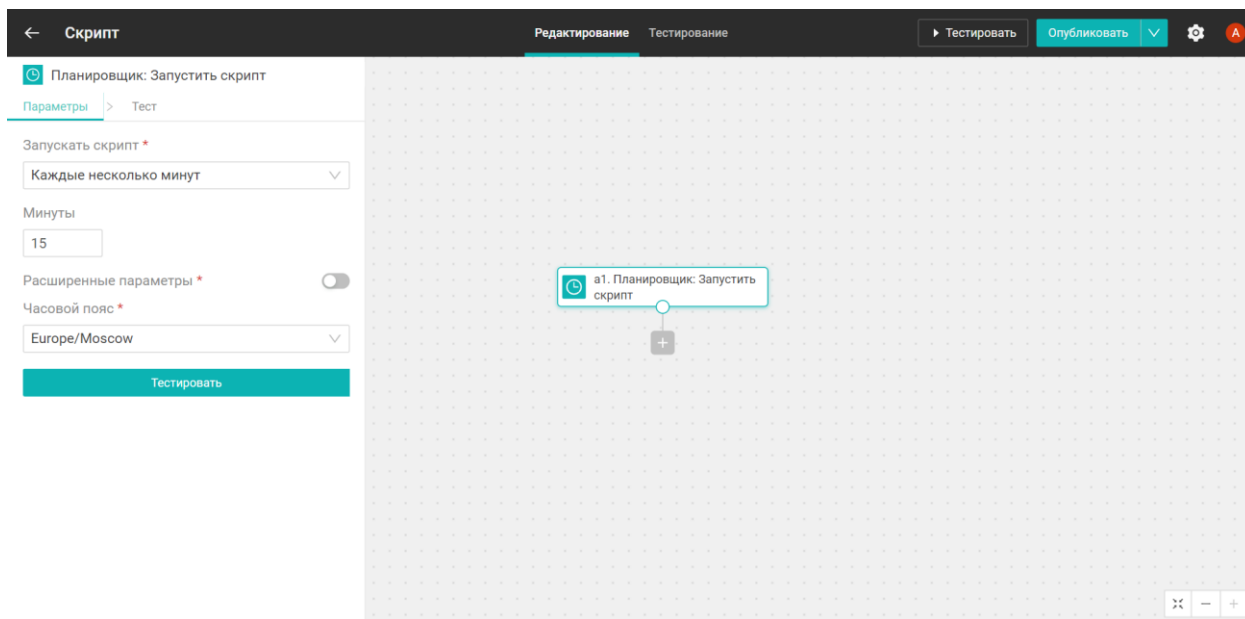


Рисунок 11 – Параметры планировщика

В расширенных параметрах можно указать даты первого и последнего запуска скрипта.

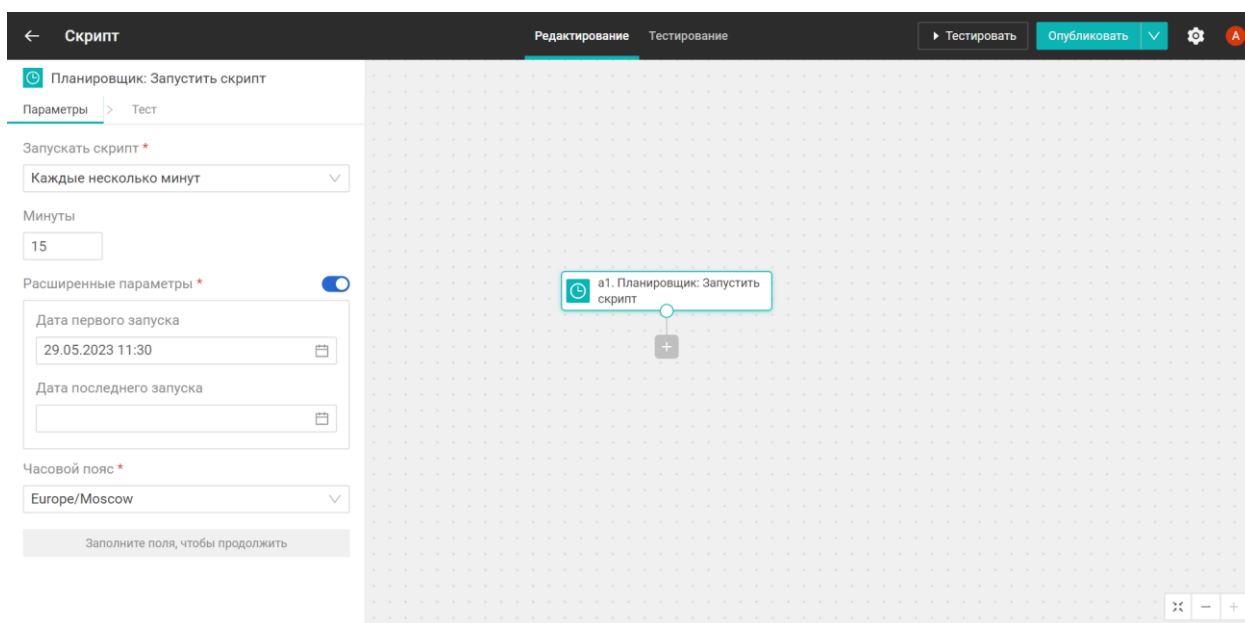


Рисунок 12 – Расширенные параметры планировщика

При создании блока часовой пояс задается автоматически на основе часового пояса браузера. При необходимости его можно изменить.

При нажатии на кнопку **Тестировать**, происходит пробный запуск блока. Во вкладке *Тест* отображаются выходные параметры блока. Например, когда был запущен скрипт.

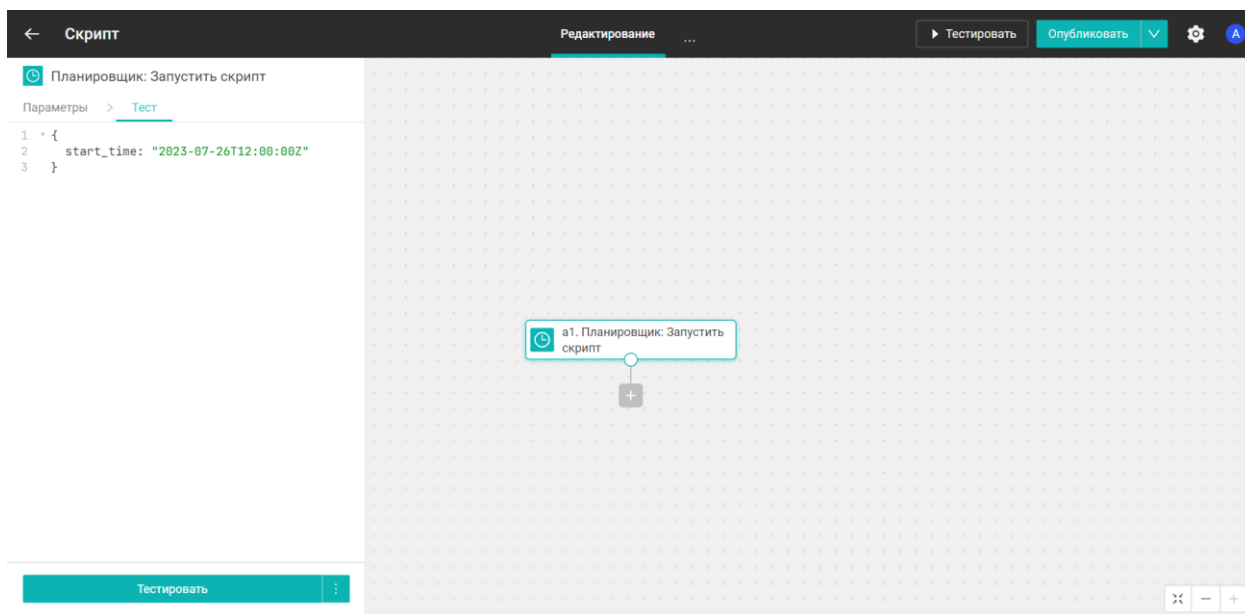


Рисунок 13 – Выходные данные после тестирования блока

## 3.2. Выбор блока-действия

Следующий блок выполнит обращение к базе данных PostgreSQL.

Добавить блок можно двумя способами:

1. Через знак + внизу существующего блока;
2. Нажать правой кнопкой мыши на линию связи или на блок и выбрать **Добавить блок**.

Необходимо выбрать пакет «PostgreSQL», затем «SQL-запрос». Далее блок необходимо настроить.

## 3.3. Создание подключения

Необходимо создать подключение. Для этого нажмите на кнопку **Новое подключение**.

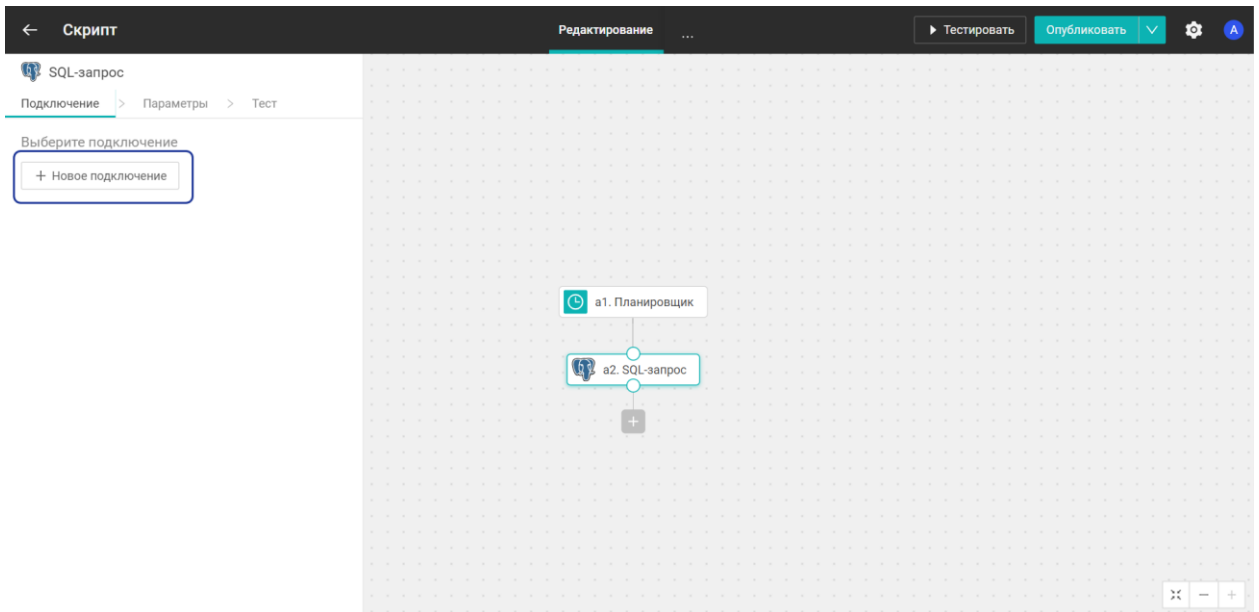


Рисунок 14 – Создание подключения

Заполните поля (они разнятся в зависимости от подключения).

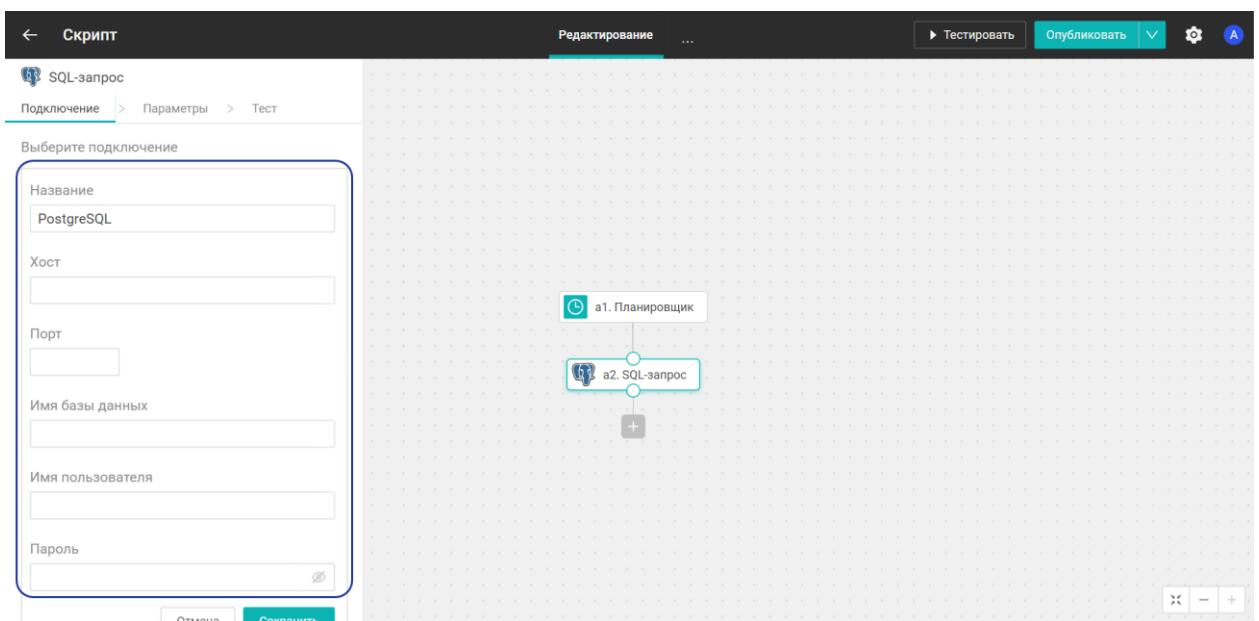


Рисунок 15 – Поля подключения

Введите произвольный SQL-запрос в поле «SQL-запрос» во вкладке «*Параметры*». Также можно открыть полноэкранный редактор, нажав на значок разворота. Пример запроса:

```
select * from название_таблицы
```

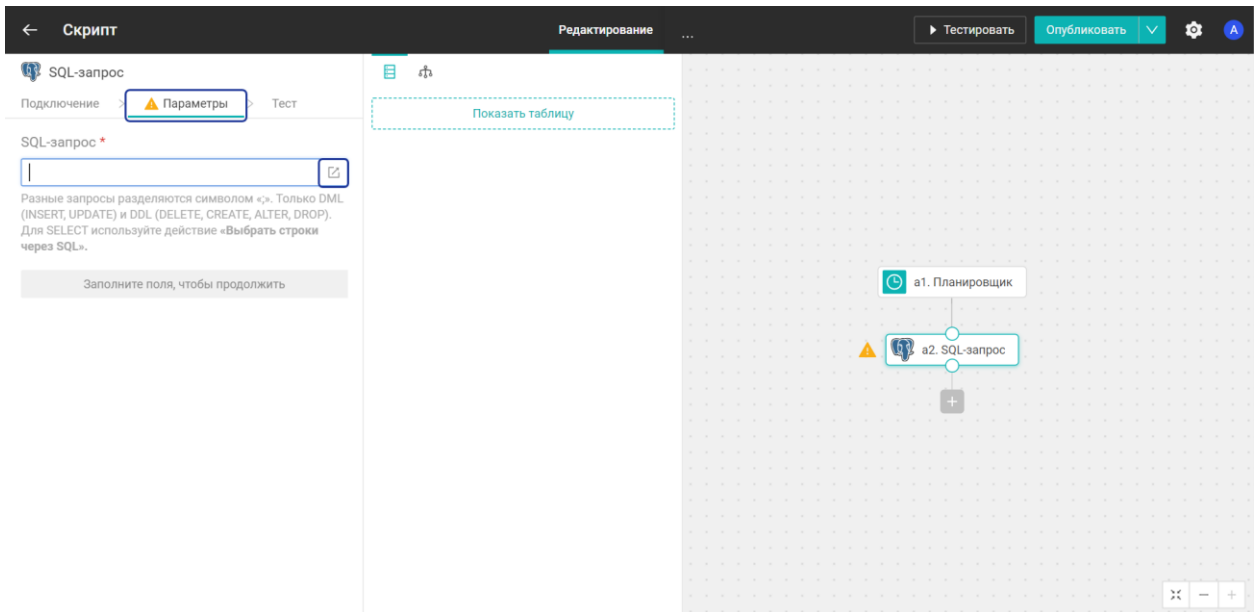


Рисунок 16 – Параметры SQL-запроса

В полноэкранном редакторе можно пользоваться автодополнением (команда *Ctrl+пробел*) и подстановкой таблиц и полей (нажать на название таблицы или поля).

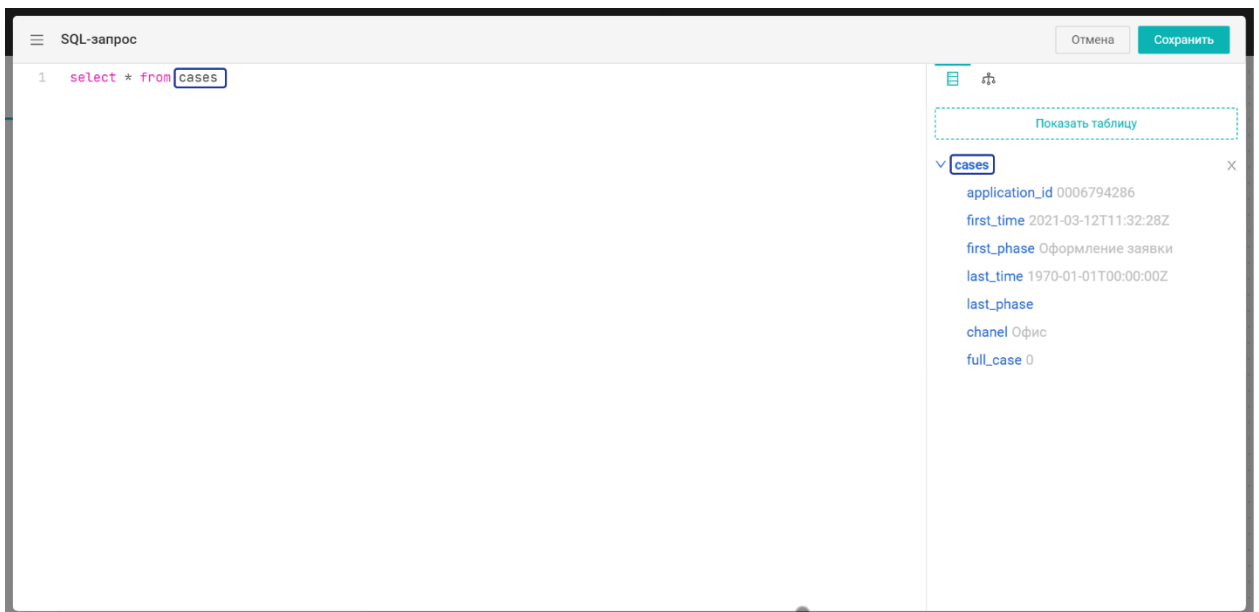


Рисунок 17 – Полноэкранный редактор

### 3.4. Добавление блока «Пространство: Добавить строки»

Этот блок выполнит вставку данных полученных в предыдущем блоке. После добавления нового блока выберите пакет «Пространство», затем «Добавить строки». В блоке во вкладке *Параметры* выберите таблицу-получатель.

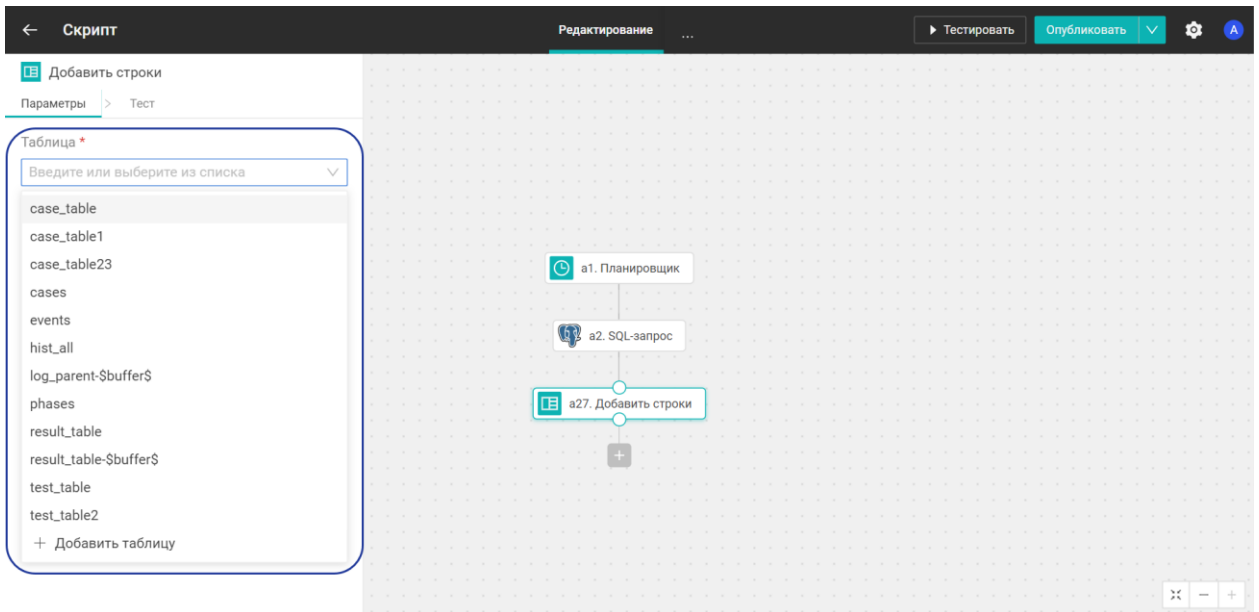


Рисунок 18 – Таблица-получатель

### 3.5. Визуальный маппинг

Далее необходимо сопоставить поля таблицы получателя с полями, полученными из предыдущего блока. Сопоставления полей происходит через визуальный маппинг. Нажмите на поле и выберите значение из маппинга.

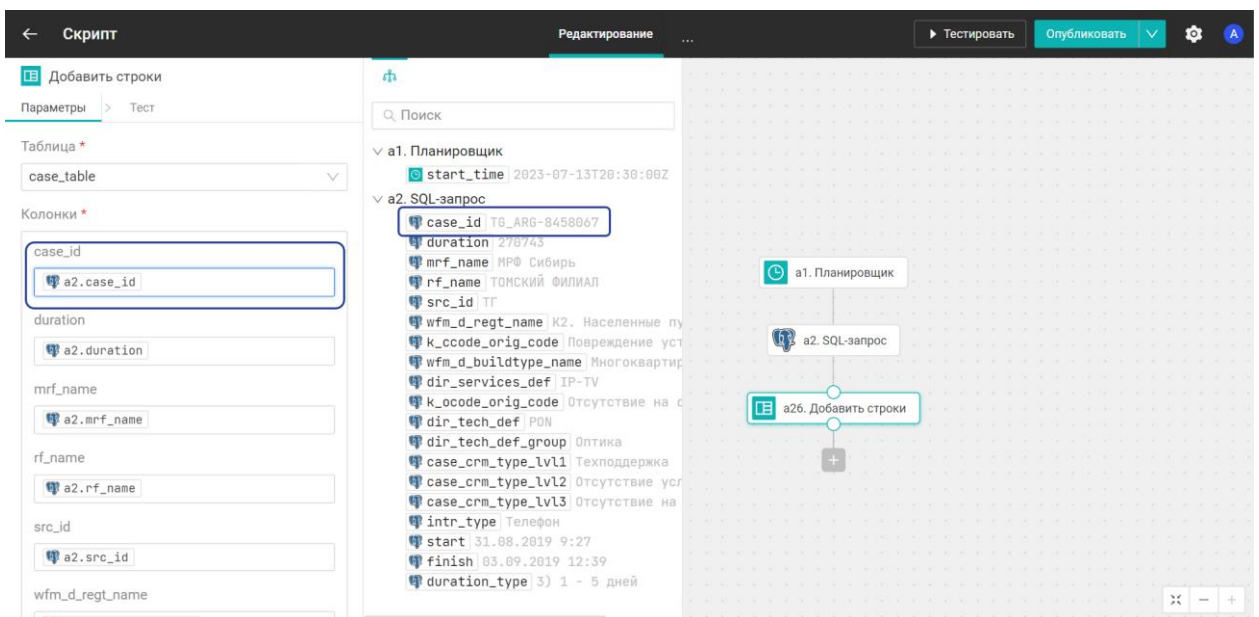


Рисунок 19 – Визуальный маппинг

Недостающие колонки можно добавить, нажав на кнопку **Добавить колонку**.

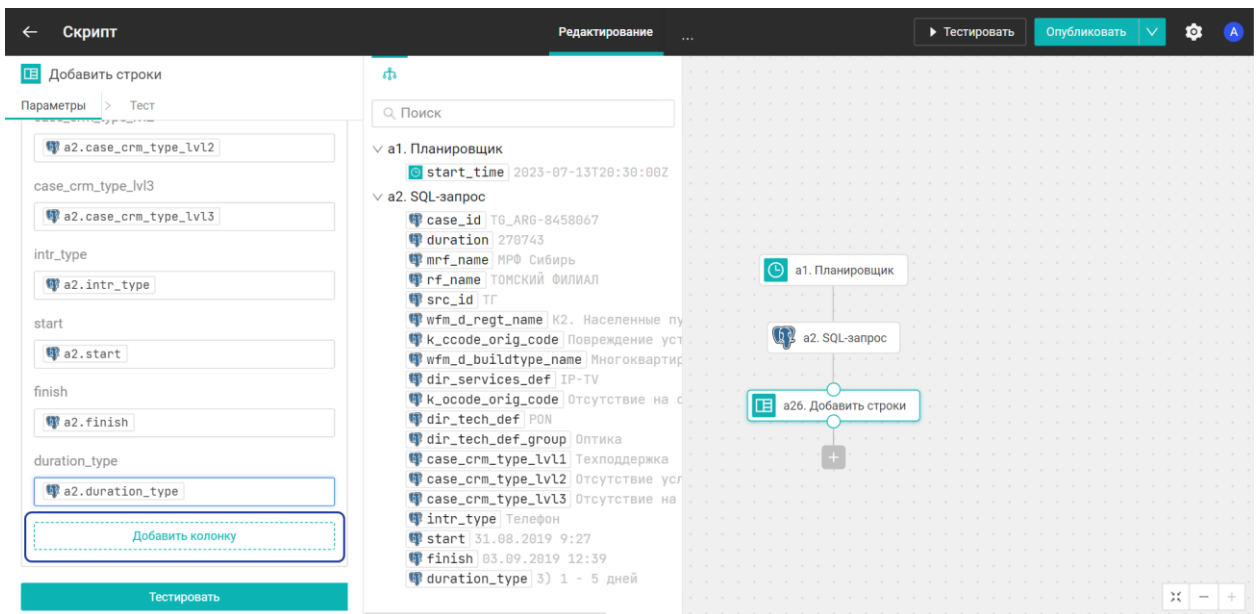


Рисунок 20 – Добавление колонки

Укажите:

- название колонки;
- тип данных.

Нажмите на кнопку **Добавить**.

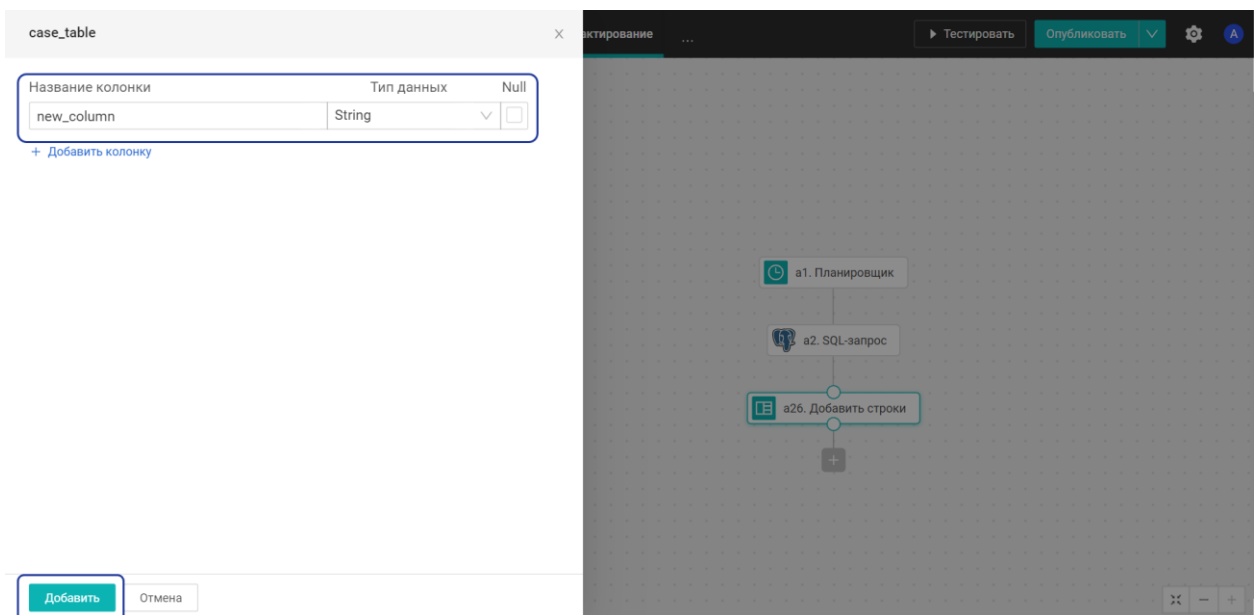


Рисунок 21 – Новая колонка

### 3.6. Удаление данных из таблицы

При тестировании блоков действие происходит на самом деле, поэтому необходимо создать блок для удаления данных. Нажмите правой кнопкой мыши на связь между блоками и выберите **Добавить**.

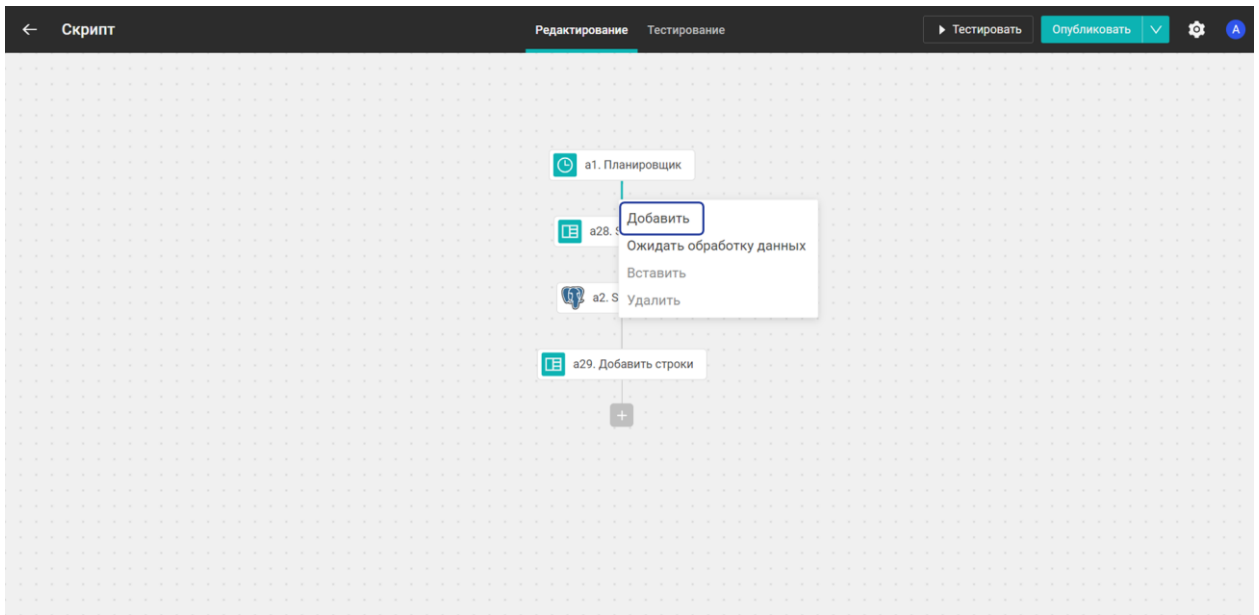


Рисунок 22 – Добавить блок

Выберите пакет «Пространство», затем «SQL-запрос». Укажите выражение и протестируйте блок. Пример выражения:

```
truncate table название_таблицы
```

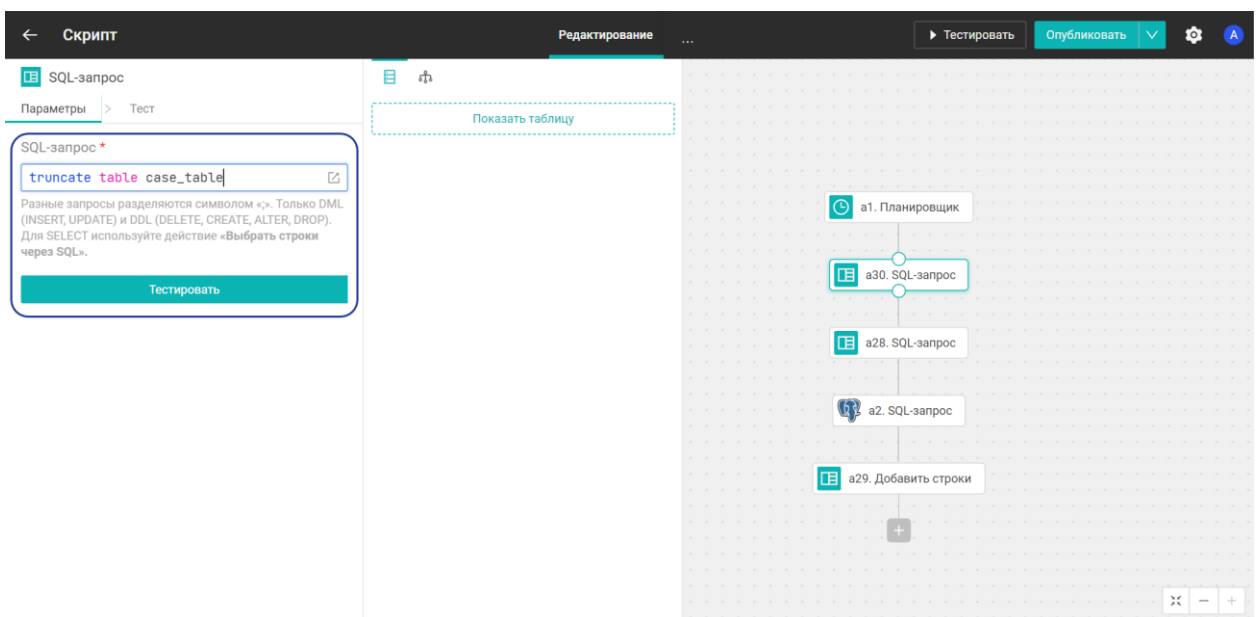


Рисунок 23 – SQL-запрос

### 3.7. Тестирование скрипта

Чтобы проверить, как работает скрипт, в режиме редактирования нажмите на кнопку **Тестировать** в правом верхнем углу.

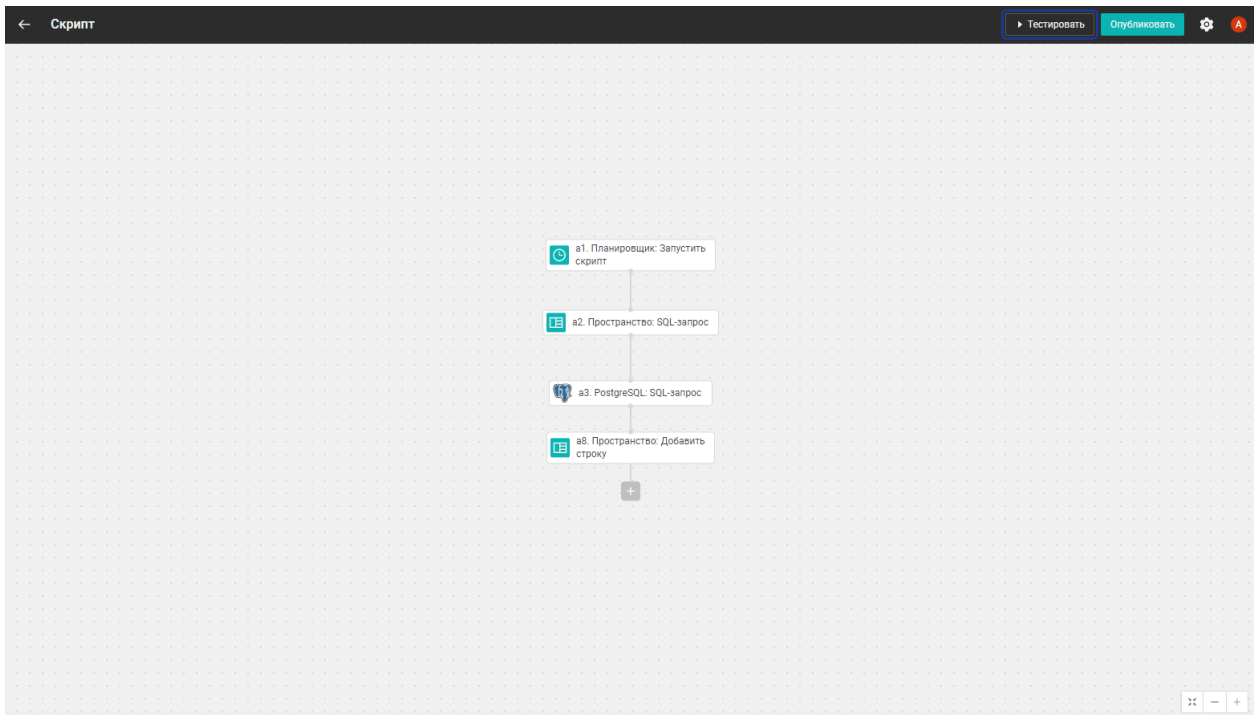


Рисунок 24 – Тестирование скрипта

Если скрипт работает корректно, его необходимо опубликовать. Для этого нажмите кнопку **Опубликовать**.

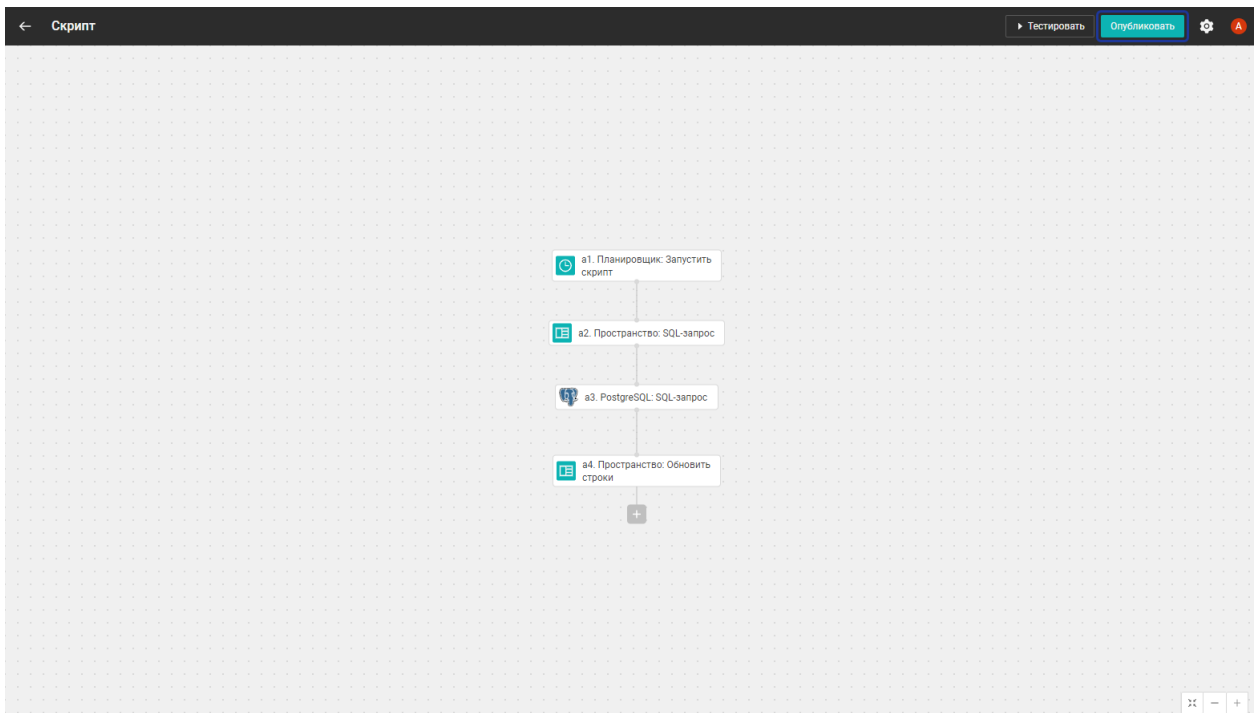


Рисунок 25 – Публикация скрипта

При публикации скрипта создается новая версия. Невозможно опубликовать скрипт, если в блоках есть незаполненные обязательные поля.



### 3.8. Добавление блока «Пространство: SQL-запрос» в пространство

Перед тем как получить справочник сотрудников, необходимо очистить таблицу. По аналогии удаления данных из таблицы, нажмите правой кнопкой мыши на блок «Планировщик» и выберите **Добавить**.

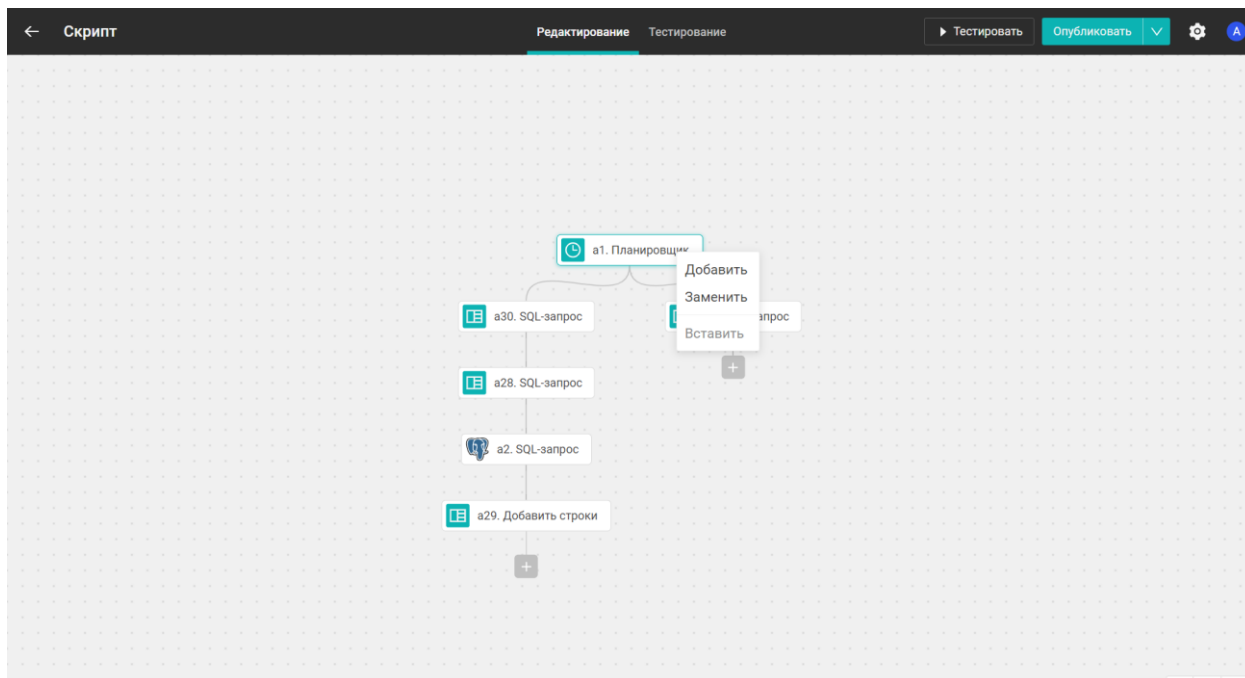


Рисунок 26 – Добавление блока

Выберите пакет «Пространство» затем «SQL-запрос». Укажите выражение и протестируйте блок.

### 3.9. Добавление блока «Файл: Прочитать CSV»

Список сотрудников хранится в виде CSV таблиц на файловом сервере. Добавьте блок и выберите пакет «Файл», затем «Прочитать CSV». Заполните поля: «Путь к файлу» (на сервере) и «Разделитель» (запятая или точка с запятой). Протестируйте блок.

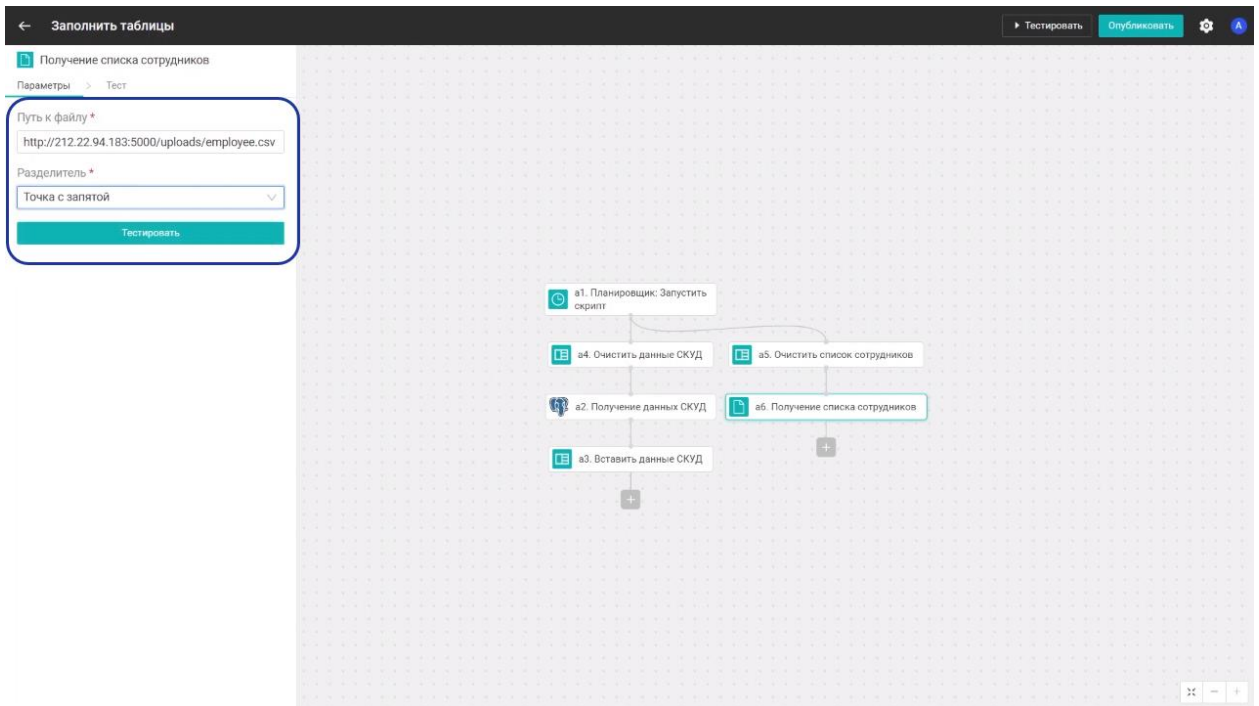


Рисунок 27 – Параметры блока «Файл: прочитайте CSV»

### 3.10. Замена типов данных полей в таблице

Далее необходимо вставить список сотрудников в таблицу-получатель. Это можно сделать с помощью блока «Пространство: Добавить строки».

Если при тестировании произошла ошибка, указанная на рисунке 28, необходимо поменять тип данных в строке.

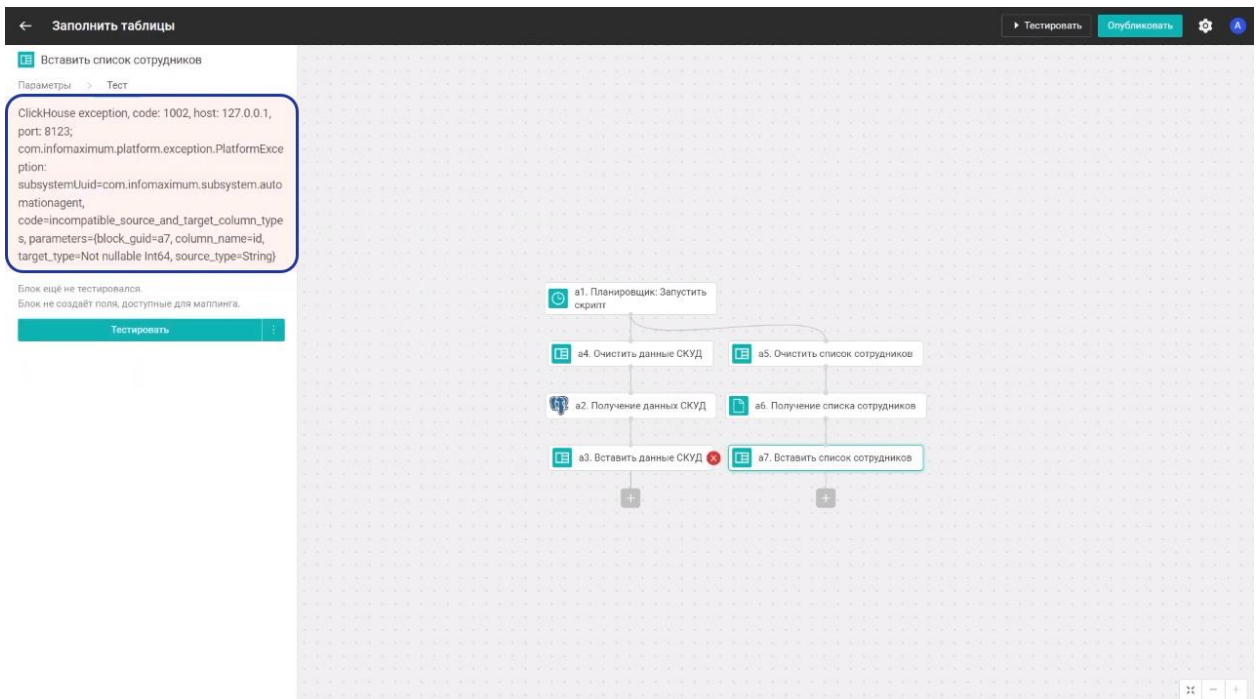


Рисунок 28 – Ошибка при тестировании

Для этого откройте нужную таблицу и нажмите **Изменить** в правом верхнем углу.

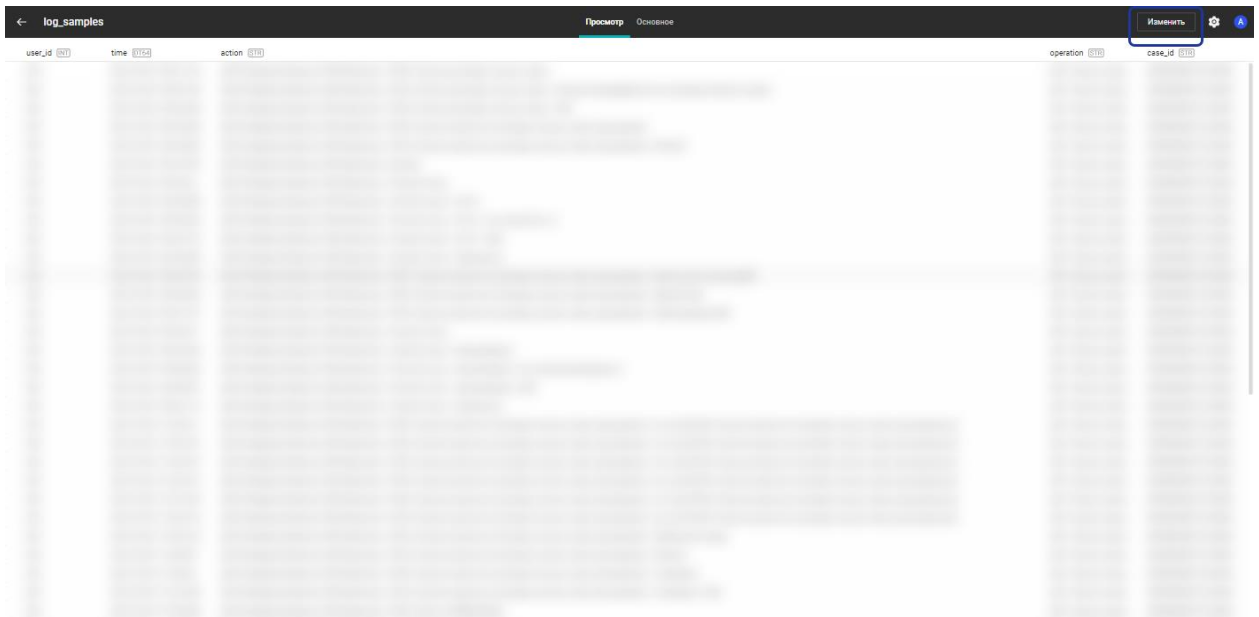


Рисунок 29 – Изменение таблицы

Нажмите на поле «Тип данных» напротив нужной колонки и выберите новый тип из списка.



Рисунок 30 – Тип данных

### 3.11. Добавление блока «HTTP: Отправить запрос»

Необходимо добавить ветку в скрипт для извлечения данных о производственном календаре. Добавьте блок для очистки («Пространство: SQL-запрос»).

Получение календаря происходит через обращение к REST API, поэтому следующий блок «HTTP: Отправить запрос». Добавьте блок и выберите «HTTP», затем «Отправить запрос». Настройте подключение в соответствующей вкладке блока.

Во вкладке *Параметры* укажите:

- URL;

- метод (GET, POST);
- заголовки:
  - ключ;
  - значение.
- тело запроса.

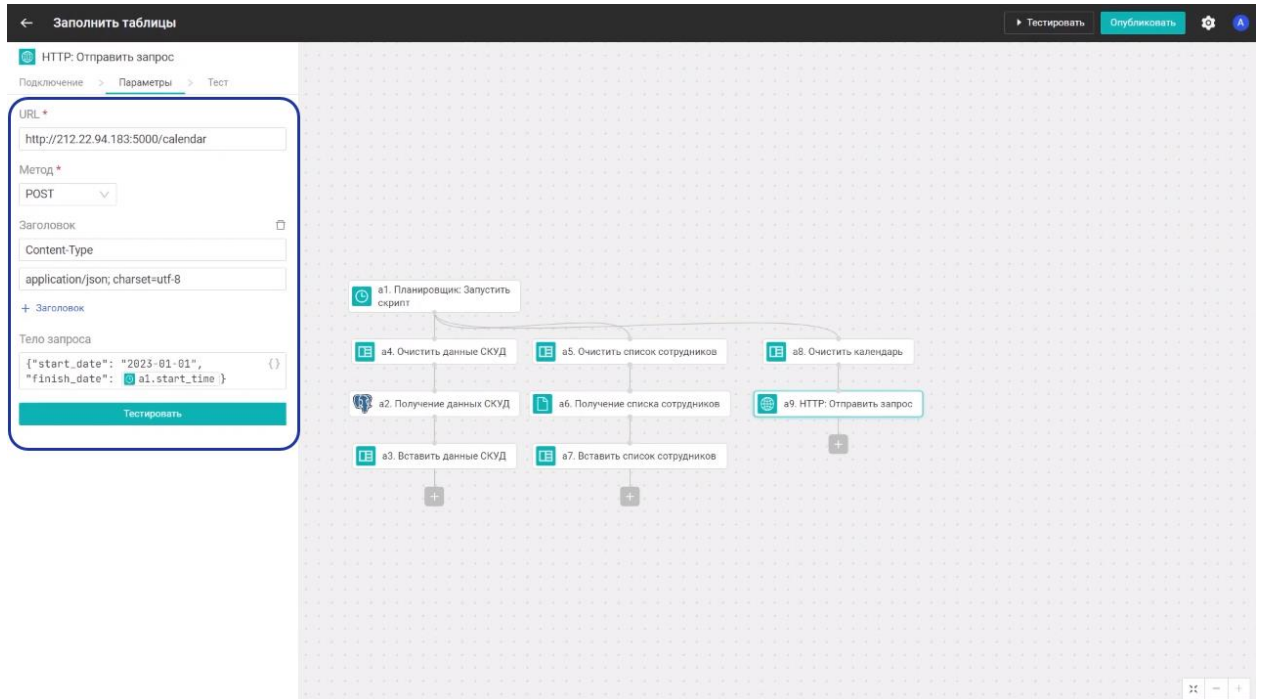


Рисунок 31 – Параметры блока «HTTP: Отправить запрос»

### 3.12. Добавление данных в таблицу с помощью скрипта

Далее необходимо выполнить вставку данных производственного календаря в таблицу-получатель. Добавьте блок «Пространство: Добавить строки», выберите таблицу получатель и сопоставьте поля с помощью визуального маппинга. Протестируйте и опубликуйте скрипт.

#### Примечание:

Увеличить/уменьшить масштаб холста с блоками можно с помощью кнопок **подогнать размер**, + и - в правом нижнем углу.

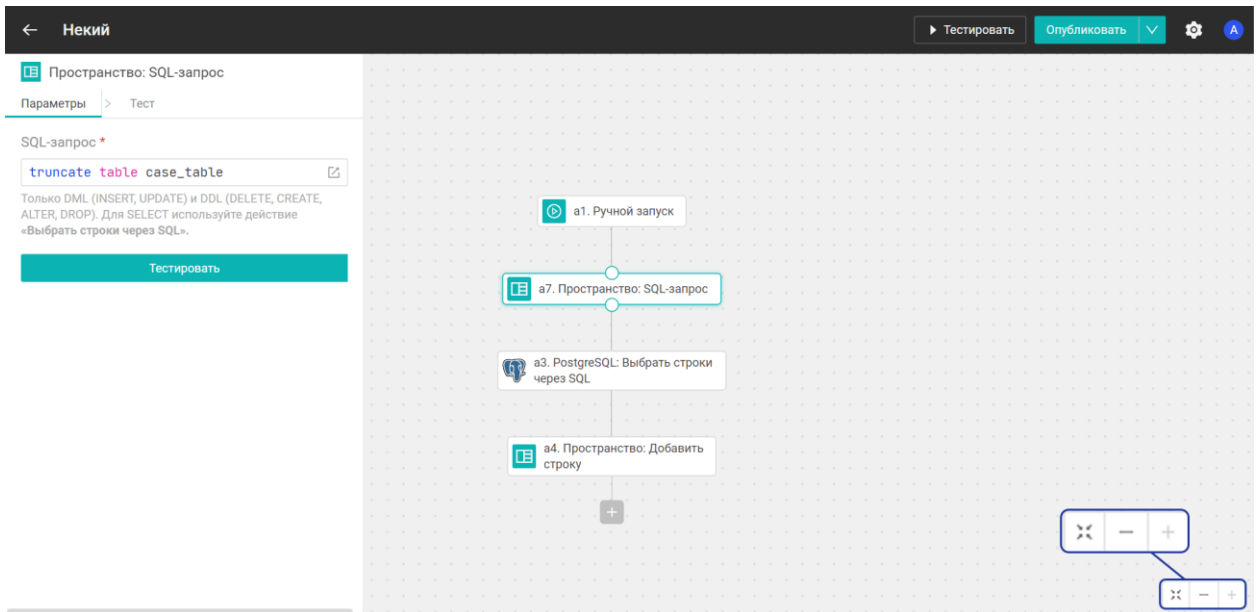


Рисунок 32 – Масштаб скрипта

## 4. Дублирование скрипта

Первичное заполнение выполнено, необходимо перейти к инкрементальной загрузке. Это можно сделать через дублирование. Чтобы дублировать скрипт, нажмите на него правой кнопкой мыши и выберите **Дублировать**.

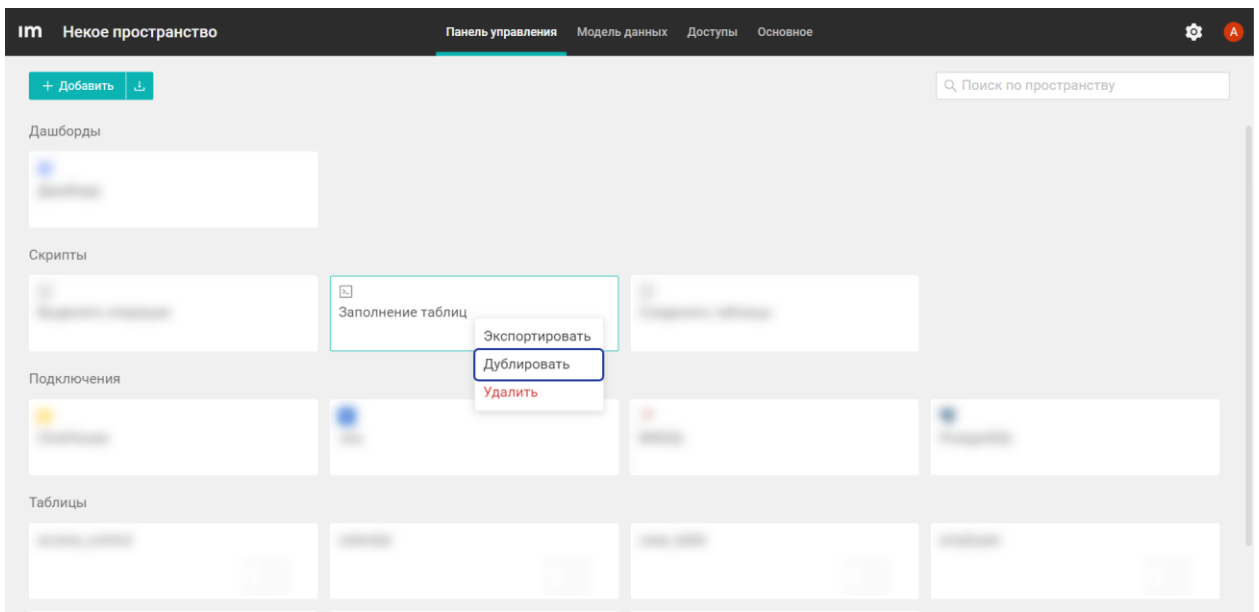


Рисунок 33 – Дублирование скрипта

## 5. Переименование скрипта

Дублированный скрипт необходимо переименовать. Для этого перейдите во вкладку *Основное* и отредактируйте поле «Название скрипта».

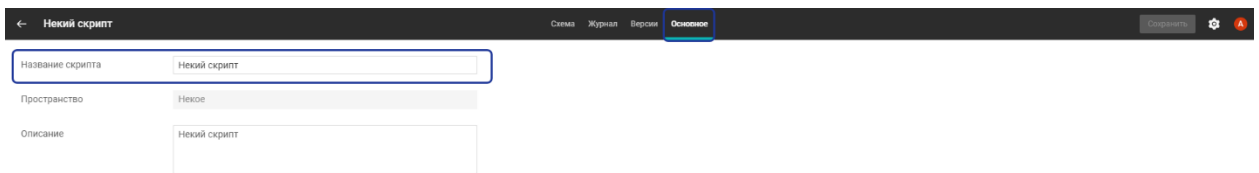


Рисунок 34 – Переименование скрипта

Чтобы сохранить изменения, нажмите кнопку **Сохранить**.

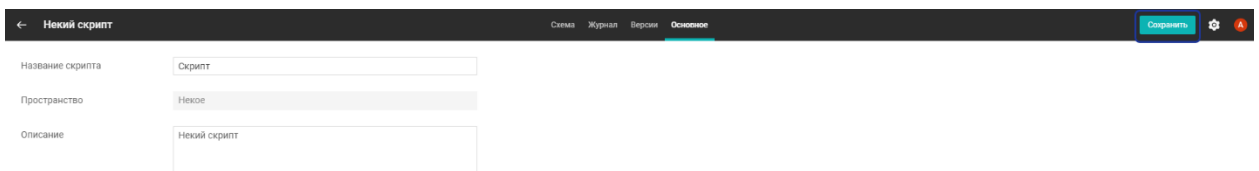


Рисунок 35 – Сохранение изменений

## 6. Изменение скрипта

Далее необходимо отредактировать скрипт. Чтобы перейти к редактированию скрипта, нажмите **Изменить**.

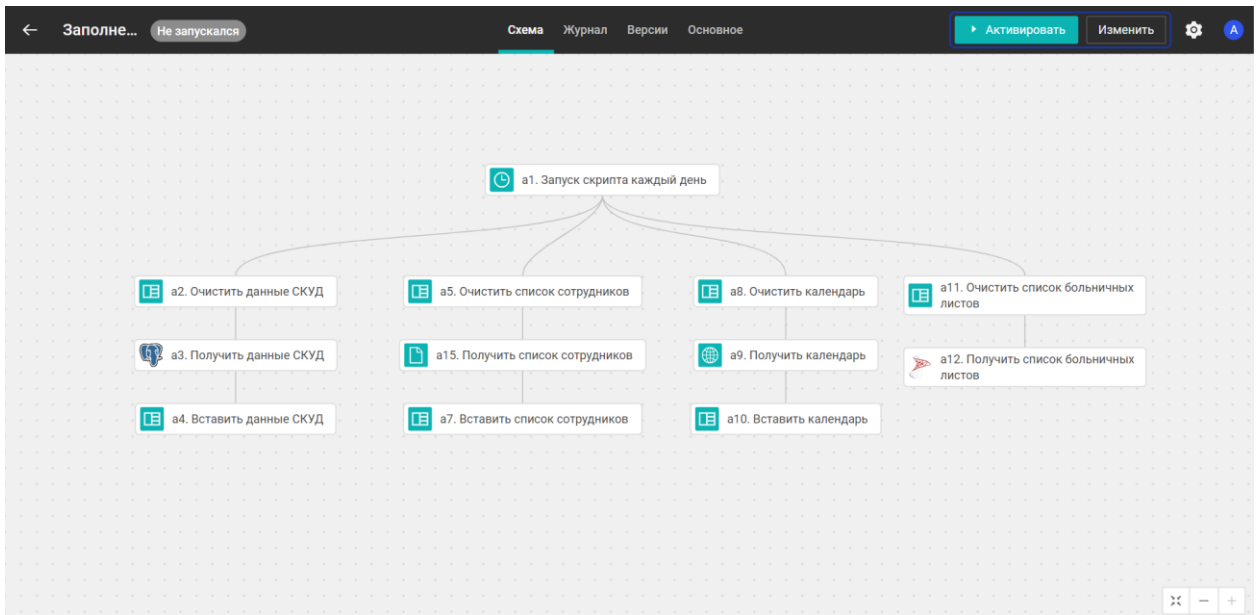


Рисунок 36 – Изменение скрипта

Порядок выполнения скрипта определяется через создание соединений между блоками. Чтобы добавить связь, нажмите на блок, выберите точку входа или выхода и протяните связь к нужному действию.

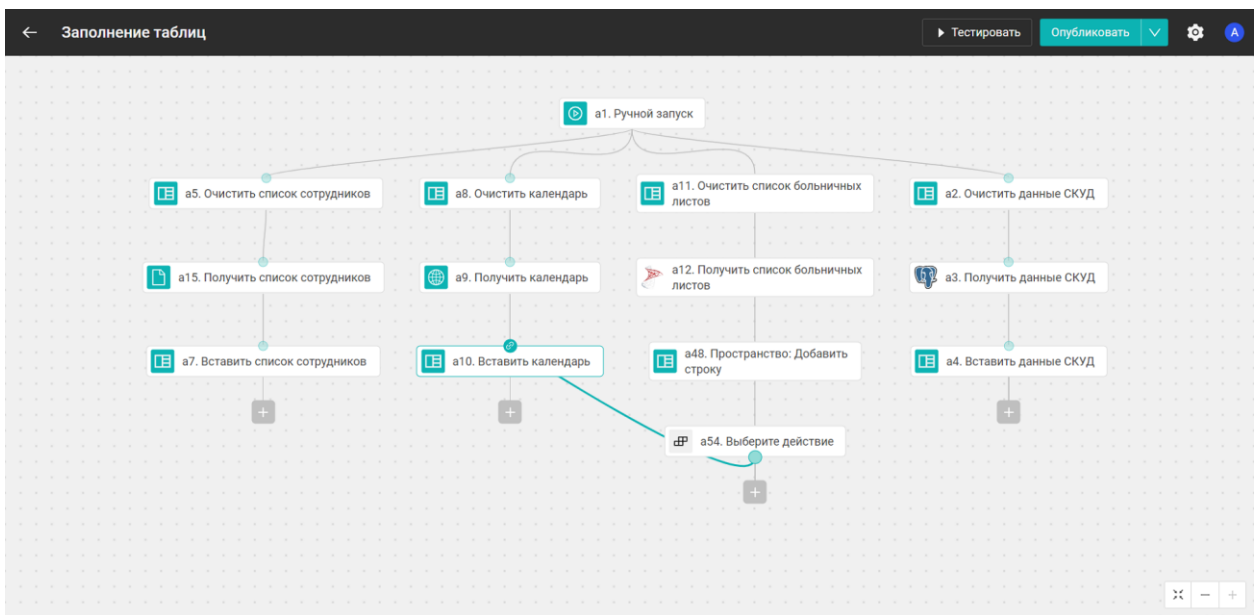


Рисунок 37 – Добавление связи

Если перетащить блок поверх другого, между ними также построится связь.

**Примечания:**

- У блока может быть несколько выходов и один вход.
- У блока обязательно должна быть хотя бы одна связь.

Чтобы удалить связь, нажмите по ней правой кнопкой мыши и выберите соответствующее действие.

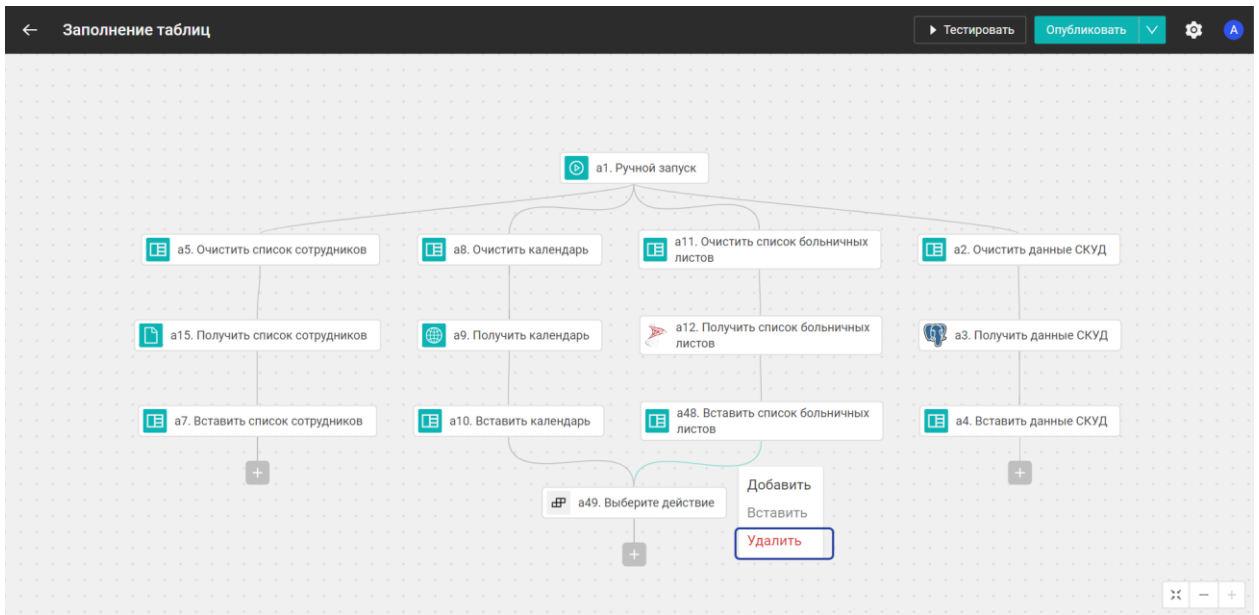


Рисунок 38 – Удаление связи

При нажатии правой кнопкой мыши на блок открываются возможные действия:

- добавить;
- дублировать;
- заменить;
- вырезать;
- копировать;
- вставить;
- удалить.

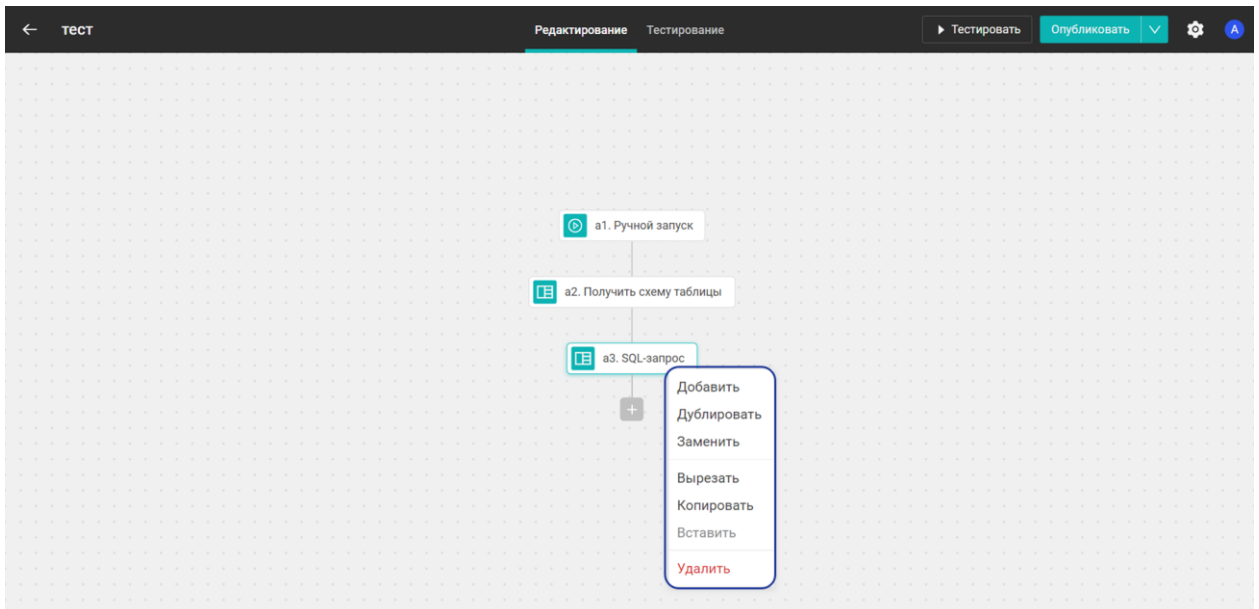


Рисунок 39 – Действия с блоком

На левой панели расположено меню настроек блока. Здесь вы можете:

1. Переименовать блок;
2. Изменить настройки блока;



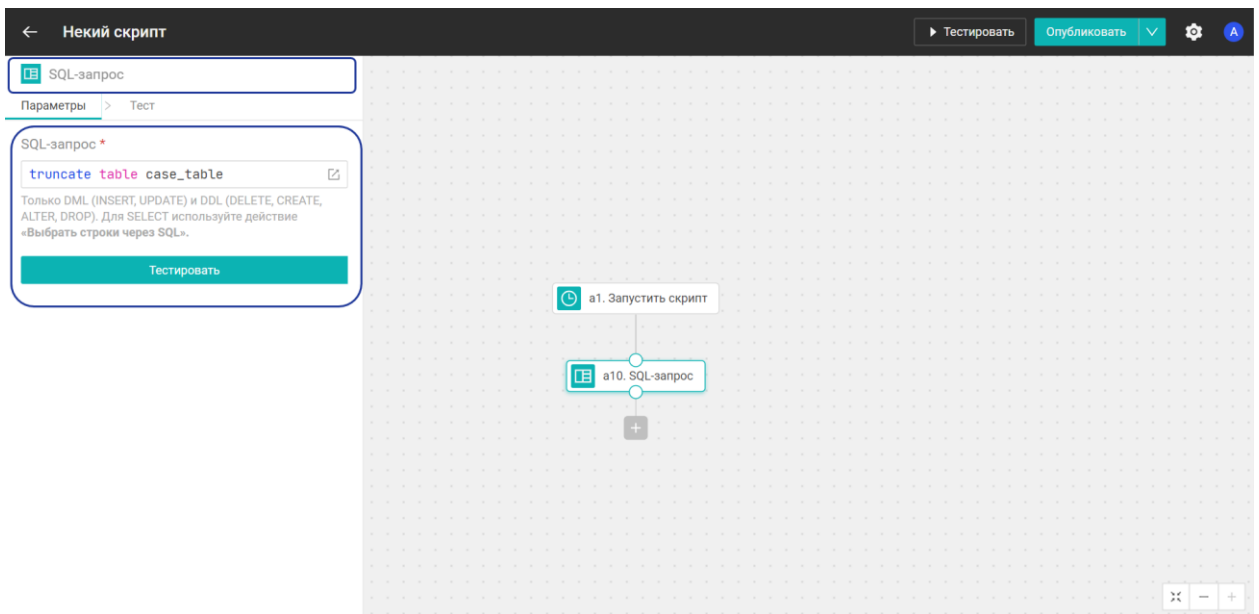


Рисунок 40 – Настройки блока

### 3. Просмотреть выходные данные блока и запустить блок.

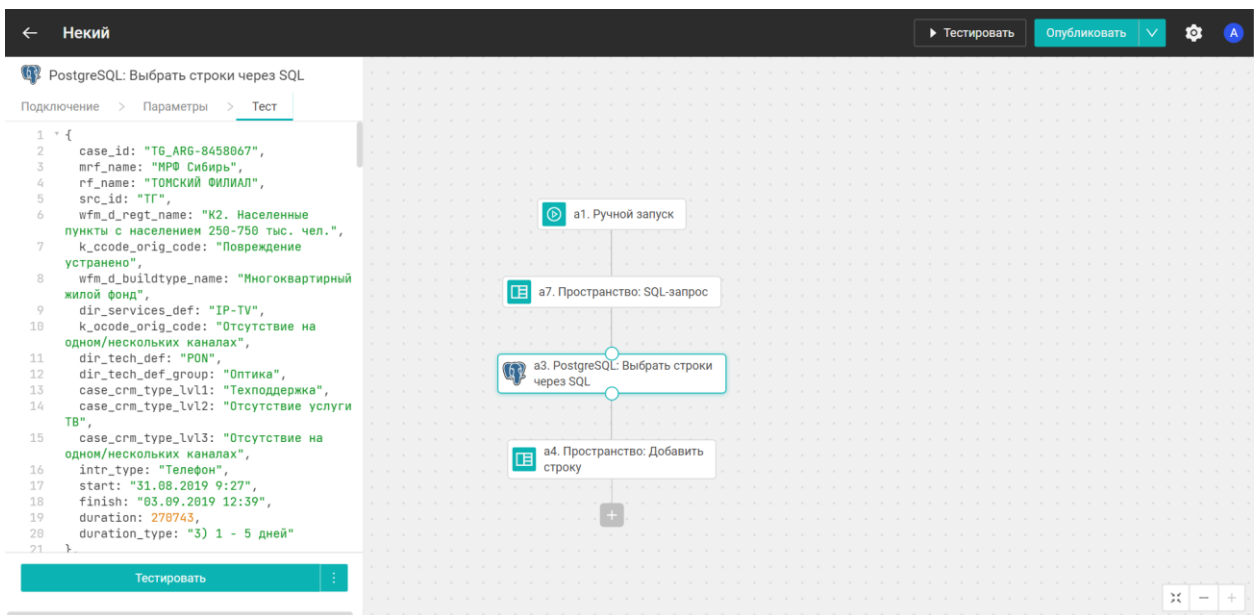


Рисунок 41 – Выходные данные

Если выбранный блок в скрипте обращается к данным из предыдущего блока, вы можете ввести входные значения вручную. Для этого сначала протестируйте блок, далее нажмите на многоточие и кликните на **Заменить входные значения**.

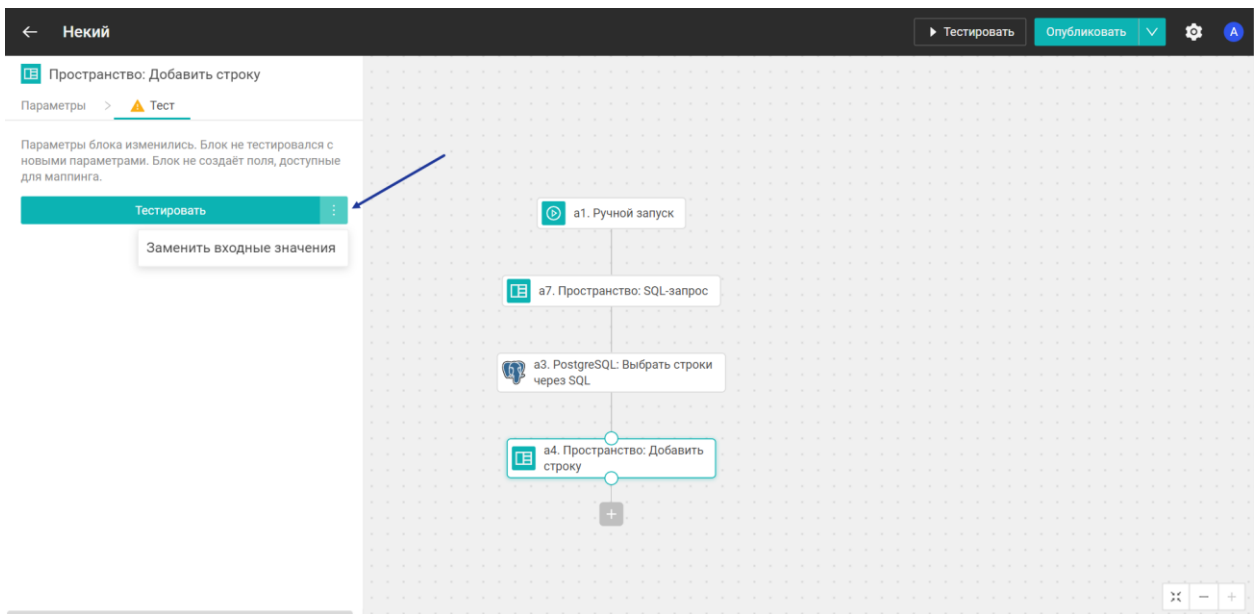


Рисунок 42 – Заменить входные данные

Введите значения и снова протестируйте блок. При следующем открытии отладочного запуска блок сохранит ваши значения. Если нужно вернуть оригинальные значения, нажмите **Восстановить**.

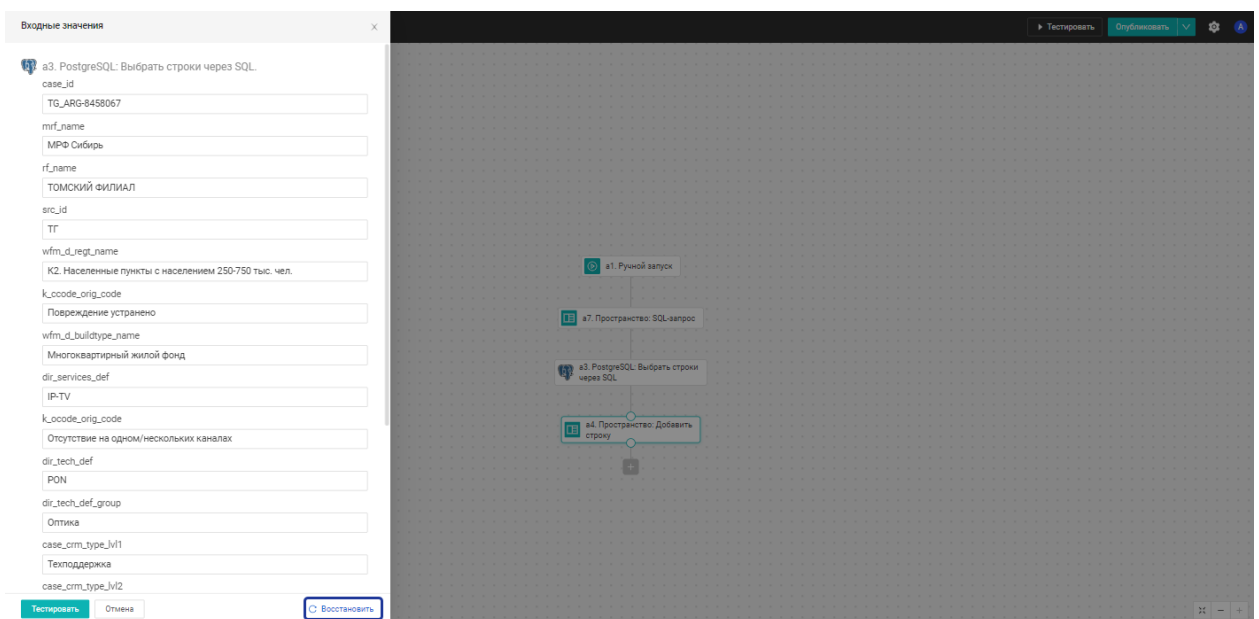


Рисунок 43 – Восстановление значений

## 7. Активация скрипта

Чтобы включить скрипт, нажмите **Активировать**.

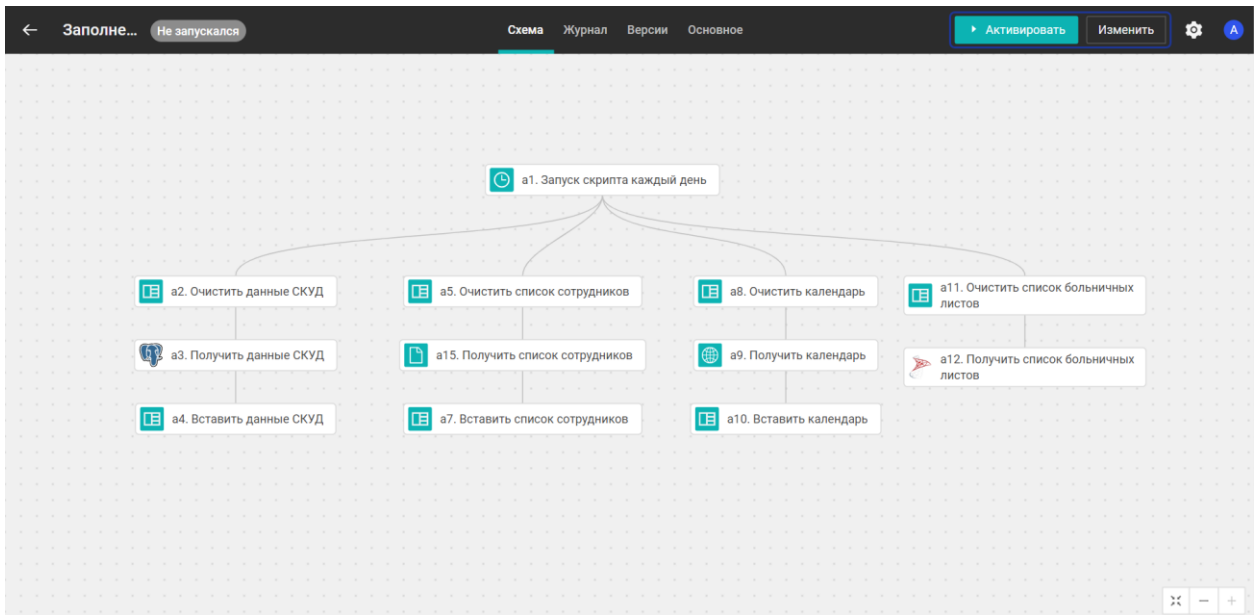


Рисунок 44 – Активация скрипта

## 8. Журнал выполнения скрипта

Журнал – это история выполнения скрипта. В нем отображены:

- время начала выполнения скрипта;
- версия скрипта;
- блок (название триггера запуска скрипта или первый блок, в котором произошла ошибка);
- длительность (время от запуска триггера до присваивания статуса).

Время	Версия	Блок	Длительность
15:08, 20 июля 2023	2	a3. Ручной запуск	63мс
15:08, 20 июля 2023	2	a3. Ручной запуск	56мс
15:06, 20 июля 2023	2	a3. Ручной запуск	89мс
15:06, 20 июля 2023	2	a3. Ручной запуск	90мс
15:06, 20 июля 2023	2	a3. Ручной запуск	90мс
15:06, 20 июля 2023	2	a3. Ручной запуск	151мс
14:43, 17 июля 2023	1	a3. Ручной запуск	50мс
14:42, 17 июля 2023	1	a3. Ручной запуск	48мс
14:41, 17 июля 2023	1	a3. Ручной запуск	61мс
14:41, 17 июля 2023	1	a3. Ручной запуск	47мс
14:41, 17 июля 2023	1	a3. Ручной запуск	115мс
21:00, 02 июля 2023	1	a3. Ручной запуск	67мс
21:00, 02 июля 2023	1	a3. Ручной запуск	70мс
21:00, 02 июля 2023	1	a3. Ручной запуск	65мс
21:00, 02 июля 2023	1	a3. Ручной запуск	66мс
21:00, 02 июля 2023	1	a3. Ручной запуск	57мс
20:59, 02 июля 2023	1	a3. Ручной запуск	123мс
15:30, 22 июля 2023	1	a3. Ручной запуск	69мс
15:30, 22 июля 2023	1	a3. Ручной запуск	50мс
15:30, 22 июля 2023	1	a3. Ручной запуск	70мс
15:29, 22 июля 2023	1	a3. Ручной запуск	159мс
14:56, 14 июля 2023	1	a3. Ручной запуск	67мс

Рисунок 45 – Журнал выполнения скрипта

Активированный скрипт попадает в Журнал выполнения, если обрабатывает триггер – первый блок в скрипте.

Выполнение может быть завершено:

- когда отработали блоки всех веток;
- когда отработали блоки всех веток до следующего триггера;
- если возникла ошибка при обработке всех веток триггера;
- если пользователь остановил скрипт.

### Примечания:

1. Статус «Ошибка» регистрируется в 2 случаях:
  - блок с ошибкой блокирует дальнейшую работу;
  - блок с ошибкой не блокирует дальнейшую работу, обрабатываются все блоки.
2. Остановить скрипт можно только в статусе «Выполняется».

Записи можно отсортировать:

- по актуальности (колонки «Время» и «Версия»);
- по алфавитному порядку (колонка «Блок»);
- по длительности (колонка «Длительность»).

При нажатии на запись выполнения скрипта открывается его схема. В верхней панели отображены:

- название;
- дата и время;
- длительность;
- статус;
- версия.



Рисунок 46 – Схема скрипта

Схема скрипта представлена блоками действий и подключений. У блока возможно одно из состояний:

- блок выполнен (доступен просмотр входных и выходных данных выбранного блока, просмотр входных данных триггера доступен всегда);

- ошибка в блоке;
- блок относится к конкретному выполнению, но его невозможно выбрать для просмотра данных;
- блок не относится к конкретному выполнению и его невозможно выбрать для просмотра данных.

Данные между блоками отображаются в формате JSON.

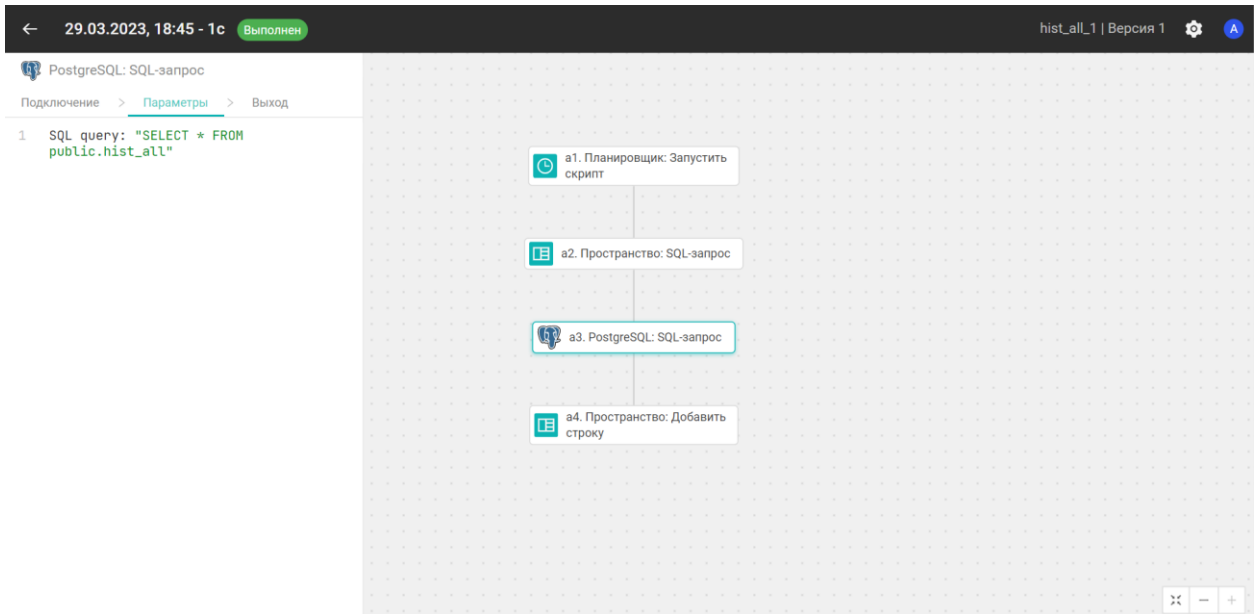


Рисунок 47 – Данные между блоками

При обработке пакетных данных из БД на входе и на выходе в панели слева отображаются первые 10 записей и общее количество записей пакетов.

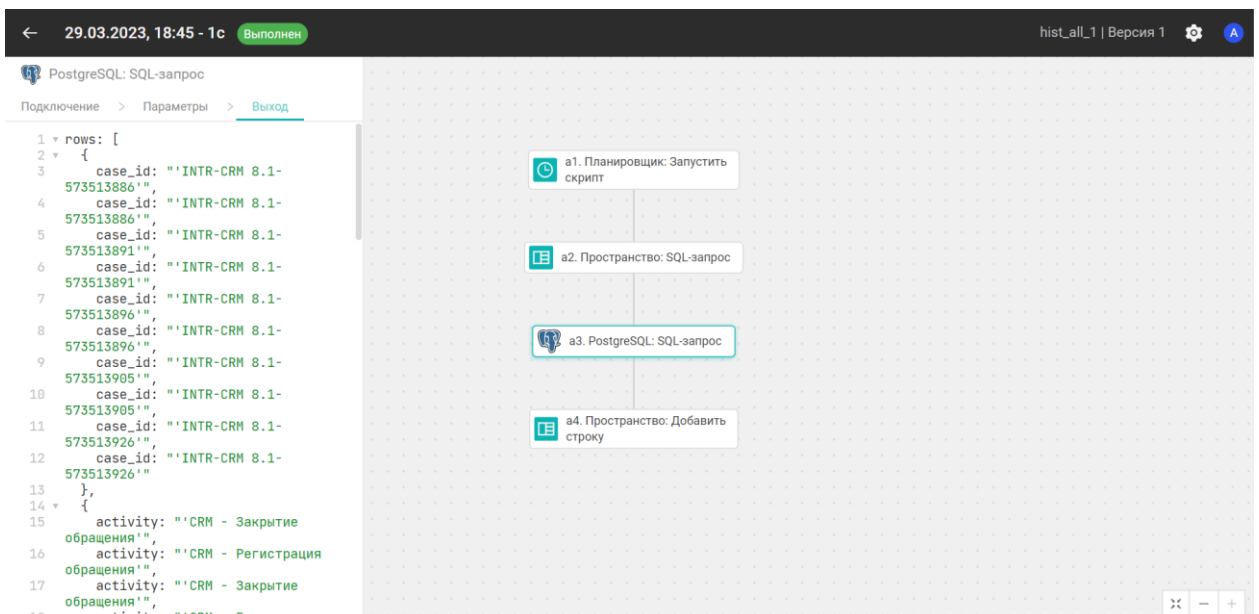


Рисунок 48 – Записи пакетов

## 9. Подключение таблицы из базы

Таблицы в пространство подключаются из базы данных ClickHouse, сервер которой добавлен в разделе «Хранилища данных». Чтобы добавить таблицу, нажмите на пункт **Подключить из хранилища**. Если в ClickHouse есть таблицы, появится модальное окно с их списком. Невозможно добавить повторно таблицы, которые уже подключены к пространству.

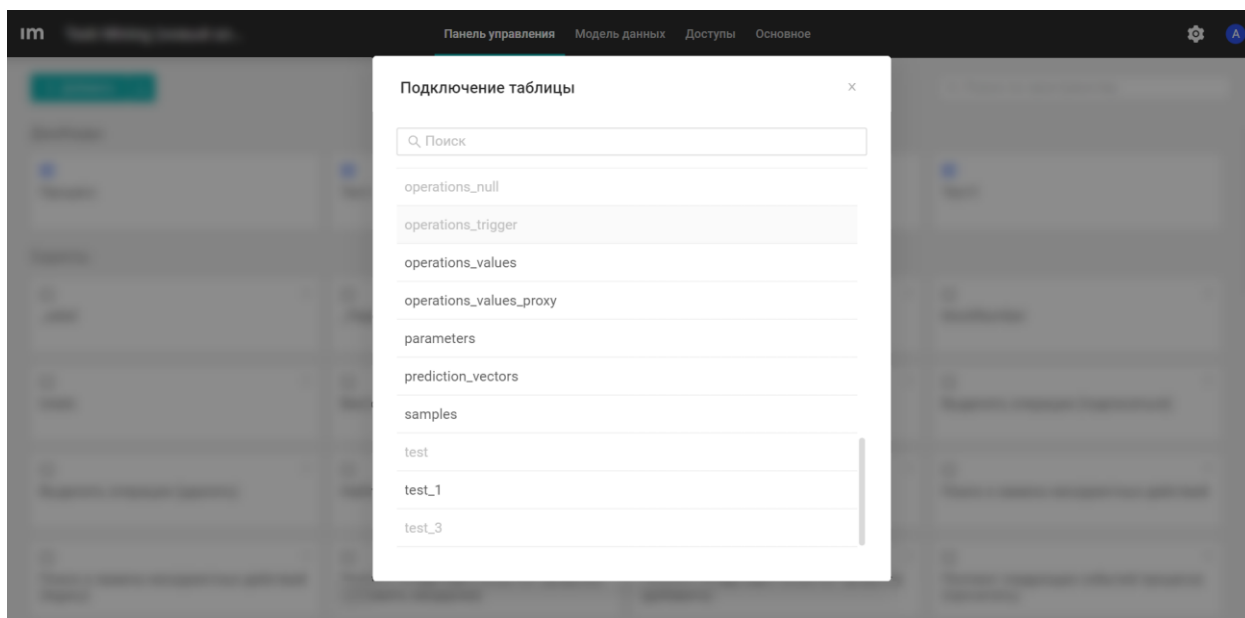


Рисунок 49 – Подключение таблицы

Выберите нужную и в режиме предпросмотра таблицы нажмите **Подключить**.

## 10. Экспорт скриптов

Чтобы экспортировать скрипт, откройте *Панель управления* и нажмите на скрипт правой кнопкой мыши. Выберите соответствующий пункт.

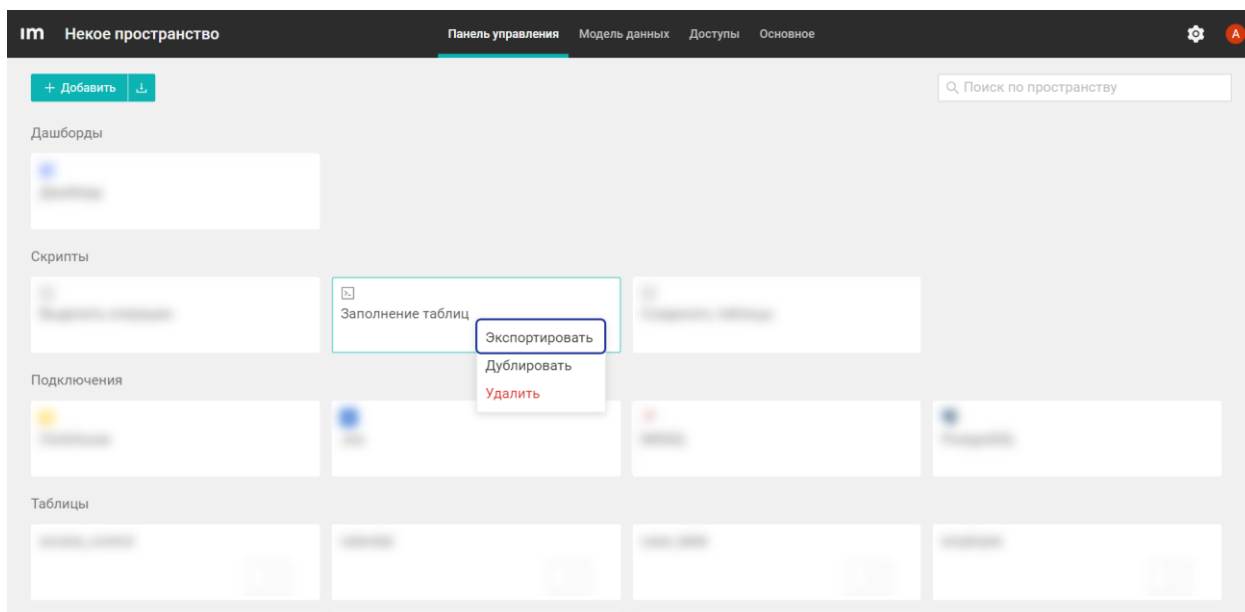


Рисунок 50 – Экспортирование скрипта

Экспортируется опубликованная версия скрипта и список используемых подключений. При экспорте скрипта с блоком-триггером «Трекер», в нем сохраняется имя веб-приложения, генерируются новые UUID и встраиваемый код.

## 11. Удаление и импорт скрипта

Удалить скрипт можно в *Панели управления*. Активный скрипт удалить невозможно.

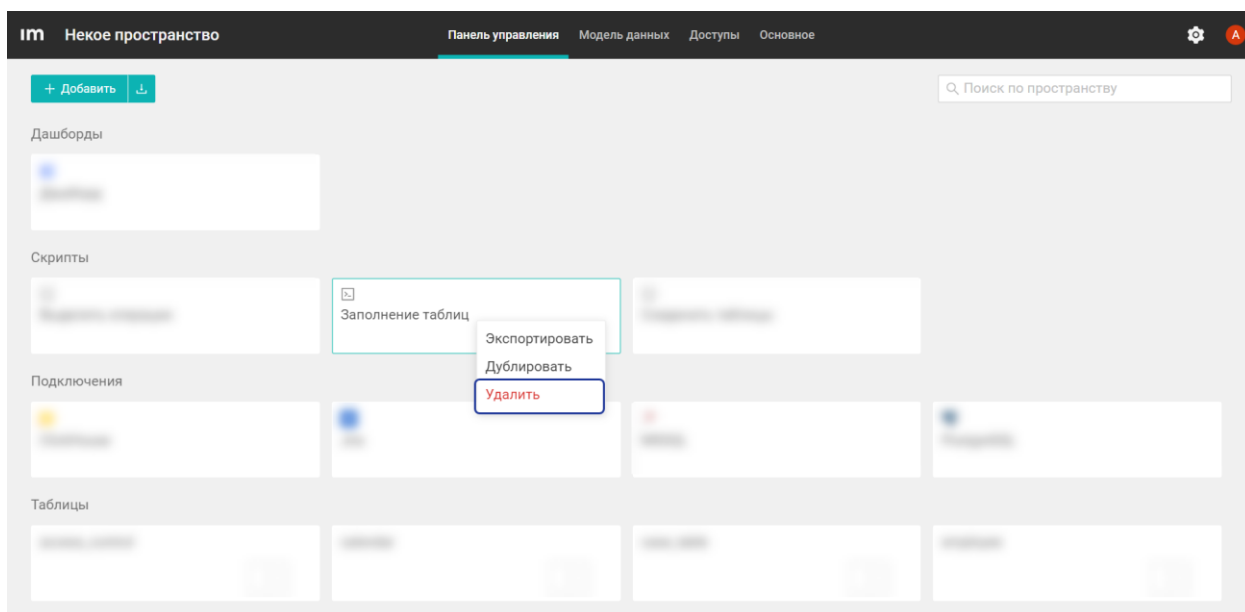


Рисунок 51 – Удаление скрипта

После нажатия **Удалить** открывается модальное окно для подтверждения удаления. Чтобы импортировать уже существующий скрипт, нажмите на соответствующую кнопку рядом с **Добавить**, выберите **Скрипт** и загрузите файл в формате *zip*.

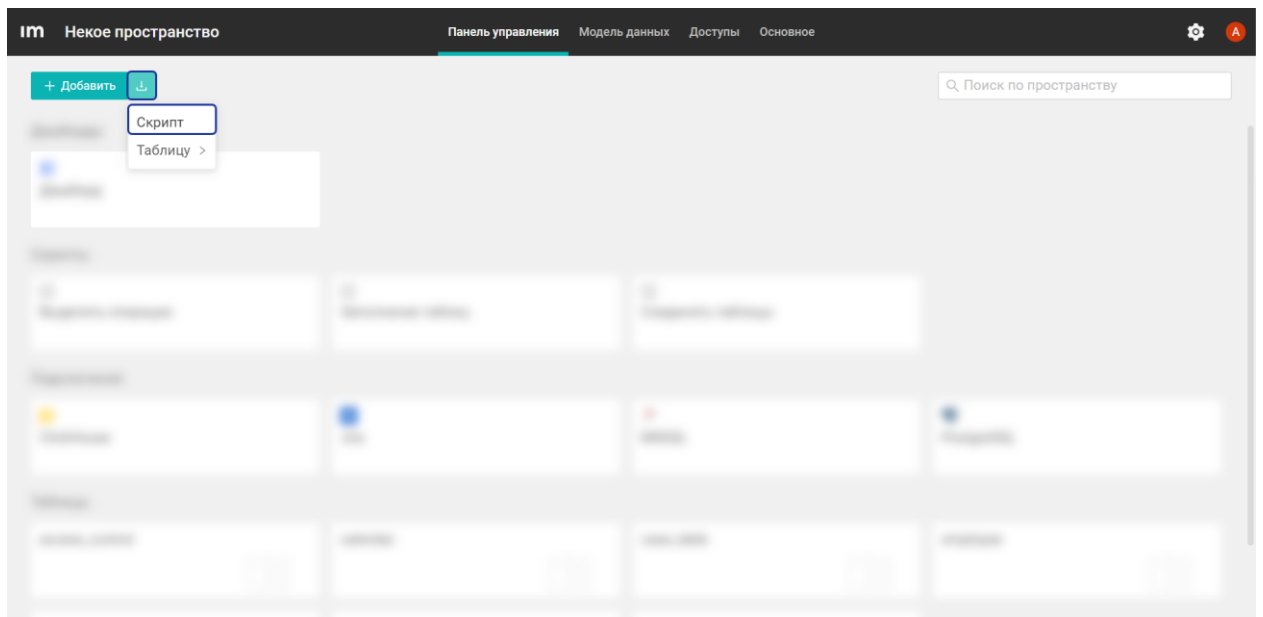


Рисунок 52 – Импорт скрипта