

Импорт в базу данных(PostgreSQL, MySQL или ClickHouse)

Базы данных как приемник данных являются альтернативой Google BigQuery.

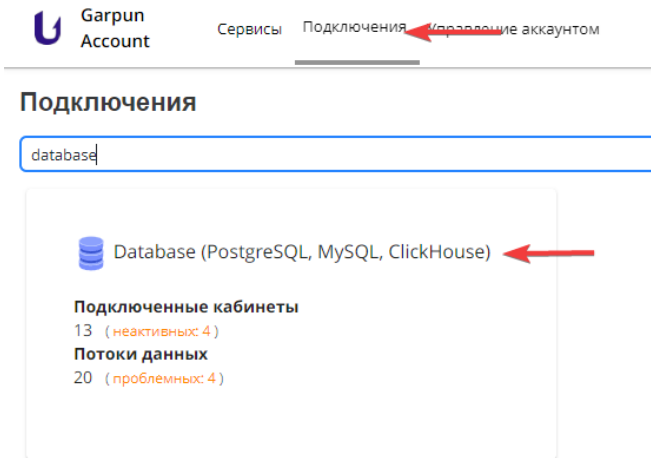
Сейчас в базы данных можно провести загрузку из большинства систем, из которых доступен импорт в Google BigQuery.

Чтобы система могла забирать и загружать данные из базы, её необходимо корректно подключить.

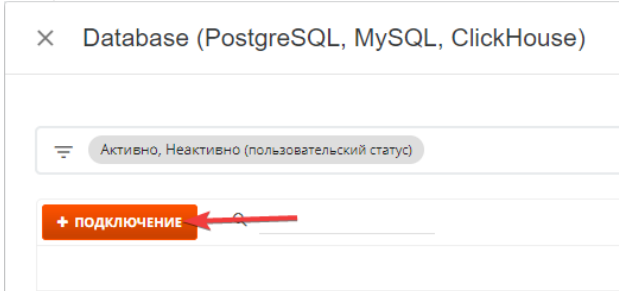
Процесс подключения расписан отдельно, поскольку сама технология организации связи с базой универсальна. Но, поскольку есть множество разных видов баз, некоторые настройки указываются не в потоках (как например при работе с Google BigQuery) а в самом подключении.

Настройка подключения едина как на импорт так и на экспорт данных.

1. Переходим в раздел "Подключения" в вашем профиле Garpun. Можно так же воспользоваться [ссылкой](#)
2. Ищем в списке систем пункт Database (PostgreSQL, MySQL, ClickHouse), кликаем на него.



3. В открывшемся окне кликаем "+ подключение"



4. Далее приступаем к вводу необходимых параметров:

- а. Имя хоста или IP-адрес - адрес обращения к базе, обязательный параметр. Эту информацию можно уточнить у администратора вашей базы, если вы им не являетесь.
- б. Порт - необязательный параметр. В зависимости от того, как организована база, для подключения к ней может быть необходим определенный порт
- в. Имя базы данных - указывать обязательно. Без этого параметра система не будет знать к какому объекту обращаться при загрузке

/выгрузке данных

г. Имя пользователя и пароль - необходимо указывать если они нужны для доступа к данным и работе с базой

д. Тип БД - указывать обязательно. У каждого типа БД свои особенности подключения и отправки запросов. В данный момент на сервисе есть поддержка PostgreSQL, MySQL и ClickHouse

е. SSL подключения для ClickHouse используется только для ClickHouse если в этом есть необходимость

5. Нажимаем "Сохранить"

i Отличия облачного хранения и сервера во внутреннем контуре:

1. Облачное хранение позволяет использовать ClickHouse как сервис в облаке, что значительно упрощает управление и масштабирование инфраструктуры. Облачные поставщики предоставляют готовые образы ClickHouse, которые можно развернуть в несколько кликов.
2. Сервер во внутреннем контуре предполагает использование собственных вычислительных ресурсов для установки и настройки ClickHouse. Это требует больше времени и усилий, но также позволяет настроить систему под конкретные потребности.
3. Облачное хранение может быть более экономичным в плане затрат на оборудование и поддержку инфраструктуры. С другой стороны, сервер во внутреннем контуре обеспечивает большую гибкость и контроль над системой.

i Если у вас защищенная база и доступ осуществляется только с разрешенных IP-адресов, просьба обратиться в поддержку Garpun за получением актуального списка наших адресов, с которых ведется подключение к базам.

i При выгрузке большого объема данных из BQ в ClickHouse(>10 гб в одной партии) используется мультипоток:

1. BQ -> S3 GCS (Google Cloud Storage). Тут мы сохраняем данные из BQ в Google Cloud Storage в формате Parquet
2. S3 GCS -> S3 (Yandex Object Storage). Передаем данные между S3 хранилищами
3. S3 (Object Storage) -> Clickhouse. Финальная передача готовых данных из S3 Yandex в ClickHouse.

Рассмотрим создание потока по передаче данных в базу на примере передачи статистики из Google BigQuery в ClickHouse

Первоначально нам необходимо войти в систему <https://feeds.garpun.com/> и нажать на

1) Источник данных > Приемник данных

В качестве источника данных выбираем **Google BigQuery**, в качестве приемника - **Database (PostgreSQL, MySQL, ClickHouse)**.

После выбора источника и приемника появится выпадающий список с возможными наборами данных. Для каждой системы они могут отличаться в зависимости от метрик, которые передаются.

Чтобы посмотреть, какие параметры будут передаваться, необходимо нажать на значок лупы справа от набора данных.

1 Источник данных > Приемник данных

Не нашли нужную систему?

Источник данных *

Google BigQuery

>

Приемник данных *

Database (PostgreSQL, MySQL, ClickHouse)

Набор данных *

Стандартная выгрузка данных

+

Добавить

Далее

2) Настройка источника данных

+ Добавить

На втором этапе выбираем аккаунт, или нажимаем кнопку , для того, чтоб добавить новое подключение, указываем Project ID и используем SQL-запрос, чтоб определить данные, которые будут выгружены

2) Настройка источника данных

Выберите аккаунт для получения данных или добавьте новый *

Q

+ Добавить

[Редактировать подключения](#)

Project ID в BigQuery *

Standard SQL Query *

1

SELECT * FROM 'test'

Используйте плейсхолдеры \$(feed_date_from) и \$(feed_date_to), чтобы подставить в запрос период получения данных

[Далее](#)

3) Настройка приемника данных

- Выбираем существующее подключение в списке, либо добавляем новое с помощью соответствующей кнопки
- Выбираем название базы данных. Это необходимо для корректной передачи информации и создания таблиц.
- Называем схему базы данных. Необходимо заполнять только если ваша база работает на PostgreSQL
- Указываем способ записи данных в таблицу. По умолчанию установлен способ "обновить".
- При выгрузке в ClickHouse можно включить доп.опцию, которая будет создавать реплицированные таблицы в разных узлах кластера для обеспечения сохранности полученных данных

4) Общие настройки

- В графе "Название потока" ввести название либо оставить сгенерированное автоматически
- В графе "Период сбора при автоматическом запуске" можно выбрать за какой период фид будет осуществлять пересбор статистики.
- В графе "Расписание" - выбрать например 8:00 утра, в это время фид будет запускаться ежедневно. При нажатии на кнопку

[+ Добавить](#)

можно добавить дополнительную строку, таким образом фид будет отрабатывать по более гибкому график

Расписание

6
7
8
9

0
5
10
15

Каждый (ые/ую) День в

След. работа: в 08:00 каждый день

Итого: Завтра, в 8:00

17
18
19
20

0
5
10
15

Каждый (ые/ую) День в

След. работа: в 20:00 каждый день

Итого: Сегодня, в 20:00

- Нажать "Готово"