

# Airbyte 部署及使用

---

AUTHOR: 彭玲 TIME: 2022/11/10

---

## Airbyte 部署及使用

[Airbyte 本地部署](#)

[Airbyte 登录](#)

[通过 Airbyte 同步 PG 至 CH](#)

[配置 Source](#)

[配置 Destination](#)

[配置 Connection](#)

[同步数据](#)

[Airbyte 同步方式](#)

---

## Airbyte 本地部署

---

Airbyte 本地部署较为简单，提前准备好 Docker 和 Docker Compose 环境即可开始：

```
1 ! [airbyte-connection-source] (assets/airbyte-connection-source.jpg) $ git clone
  https://github.com/airbytehq/airbyte.git
2 $ cd airbyte
3 $ docker-compose up
```

上述命令执行成功后，打开浏览器，输入：<http://localhost:8000> 即可访问 Airbyte 服务。

## Airbyte 登录

---

访问 <http://localhost:8000> 时，界面提示输入用户名和密码，初始 用户名/密码 为 airbyte/password。

## 通过 Airbyte 同步 PG 至 CH

---

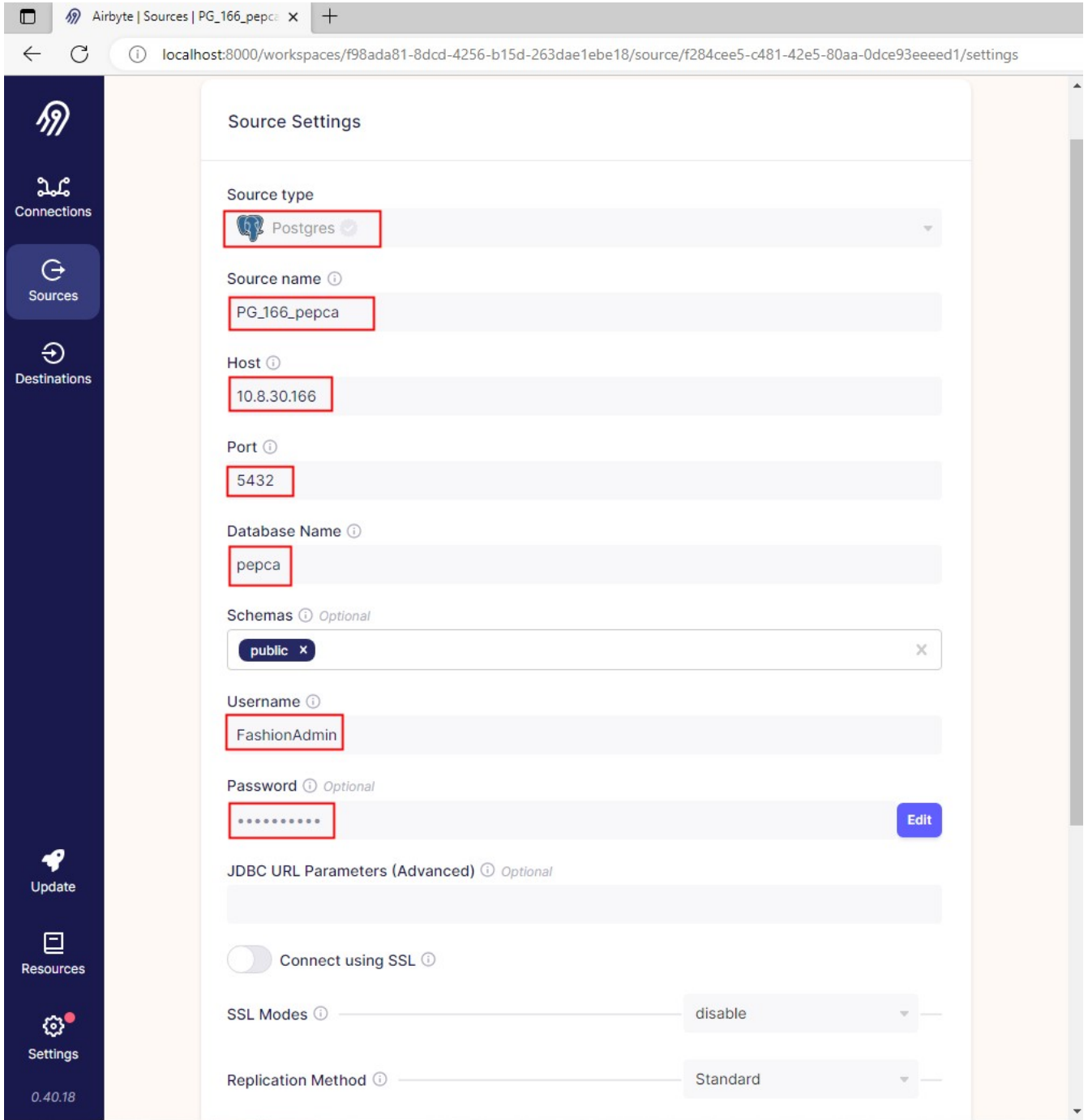
这里介绍下如何通过 Airbyte 界面将 PostgreSQL 数据库 (Source) 同步至 ClickHouse 数据库 (Destination)。3 个步骤：

1. 创建 源数据库 (Source) ；
2. 创建 目标数据库 (Destination) ；
3. 建立 源数据库 至 目标数据库 之间的连接 (Connection) 。

说明：以下步骤可以通过 Onboarding 菜单逐步进行，也可以通过 Sources、Destinations 和 Connections 分步进行。

## 配置 Source

进入 Sources 菜单页，点击 + New source 按钮，选择 Source type 为 Postgres，然后依次填写下面的信息：



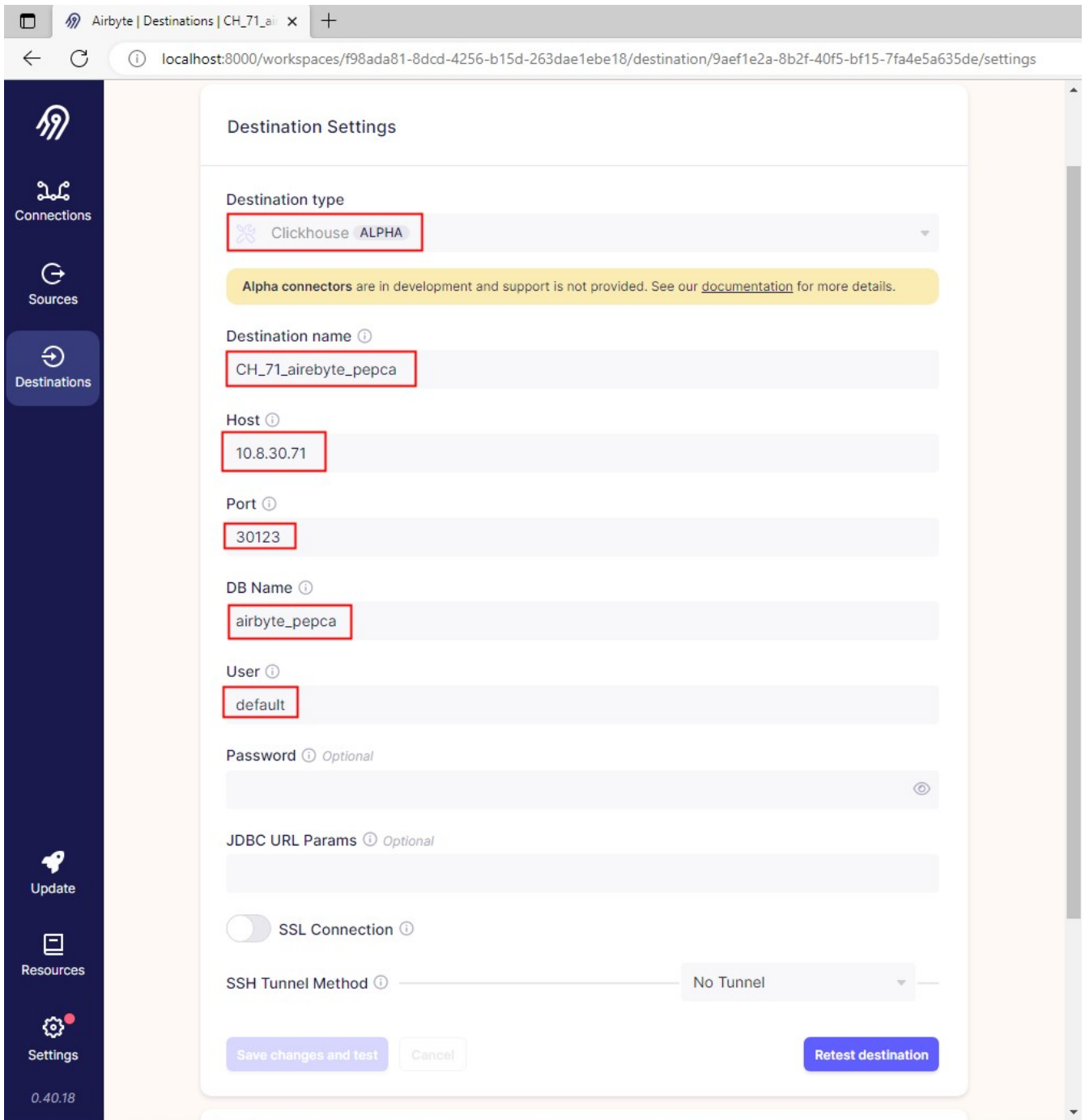
The screenshot shows the 'Source Settings' page in the Airbyte interface. The left sidebar contains navigation options: Connections, Sources, Destinations, Update, Resources, and Settings (0.40.18). The main content area is titled 'Source Settings' and contains the following fields:

- Source type: Postgres
- Source name: PG\_166\_pepca
- Host: 10.8.30.166
- Port: 5432
- Database Name: pepca
- Schemas (Optional): public
- Username: FashionAdmin
- Password (Optional): [masked]
- JDBC URL Parameters (Advanced) (Optional): [empty]
- Connect using SSL: [unchecked]
- SSL Modes: disable
- Replication Method: Standard

填写完成后，点击 Set up source 完成源数据库的配置。

## 配置 Destination

进入 Destinations 菜单页，点击 + New destination 按钮，选择 Destination type 为 Clickhouse，然后依次填写下面的信息：



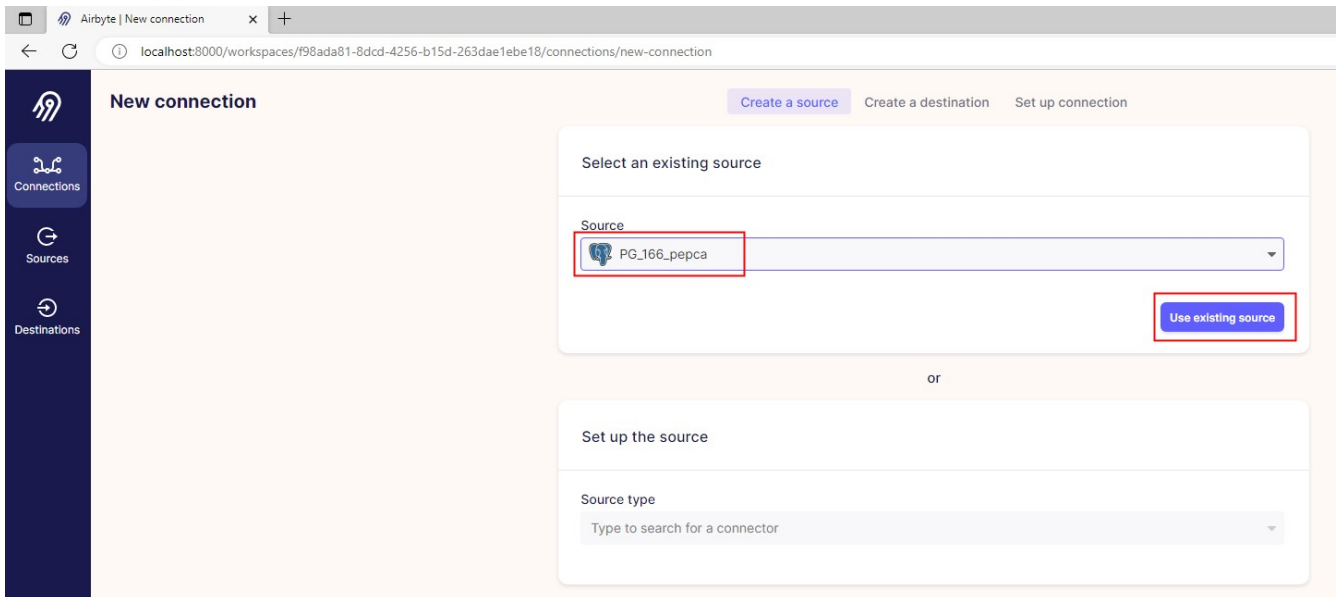
填写完成后，点击 `Set up destination` 完成目标数据库的配置。

注意：需要提前准备好 CH 数据库（如，提前创建好这里的 `airbyte_pepca` 数据库），否则，连接测试会失败。

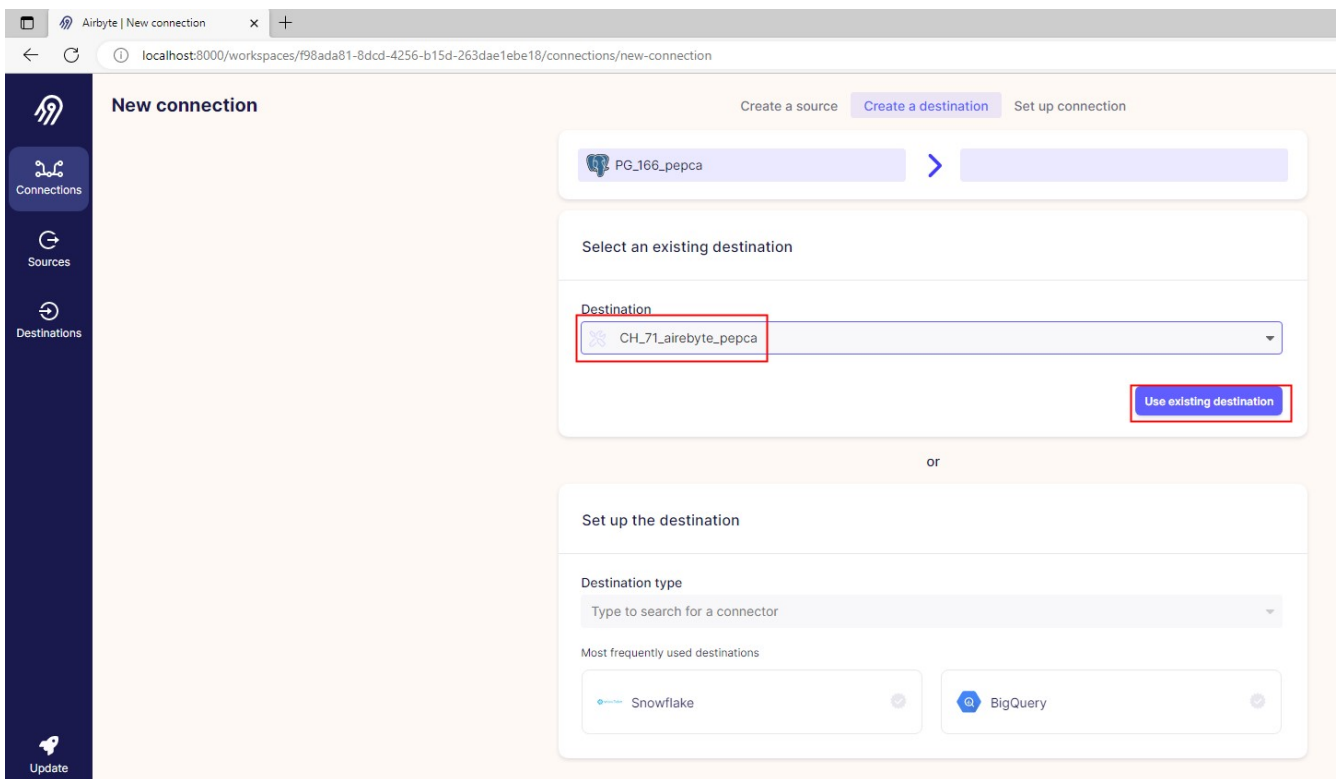
## 配置 Connection

配置好 Source 和 Destination 后，接下来为两者建立起连接即可。

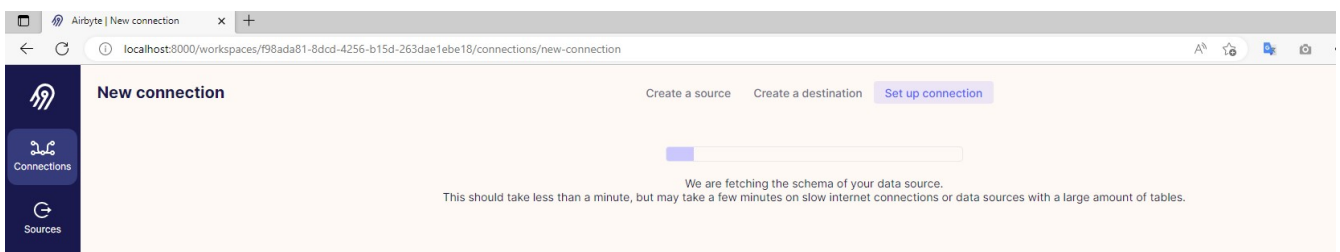
进入 Connections 菜单页，点击 `+ New connection` 按钮，选择 `select an existing source`，从 `source` 下拉列表中选择前面已经配置好的源数据库：



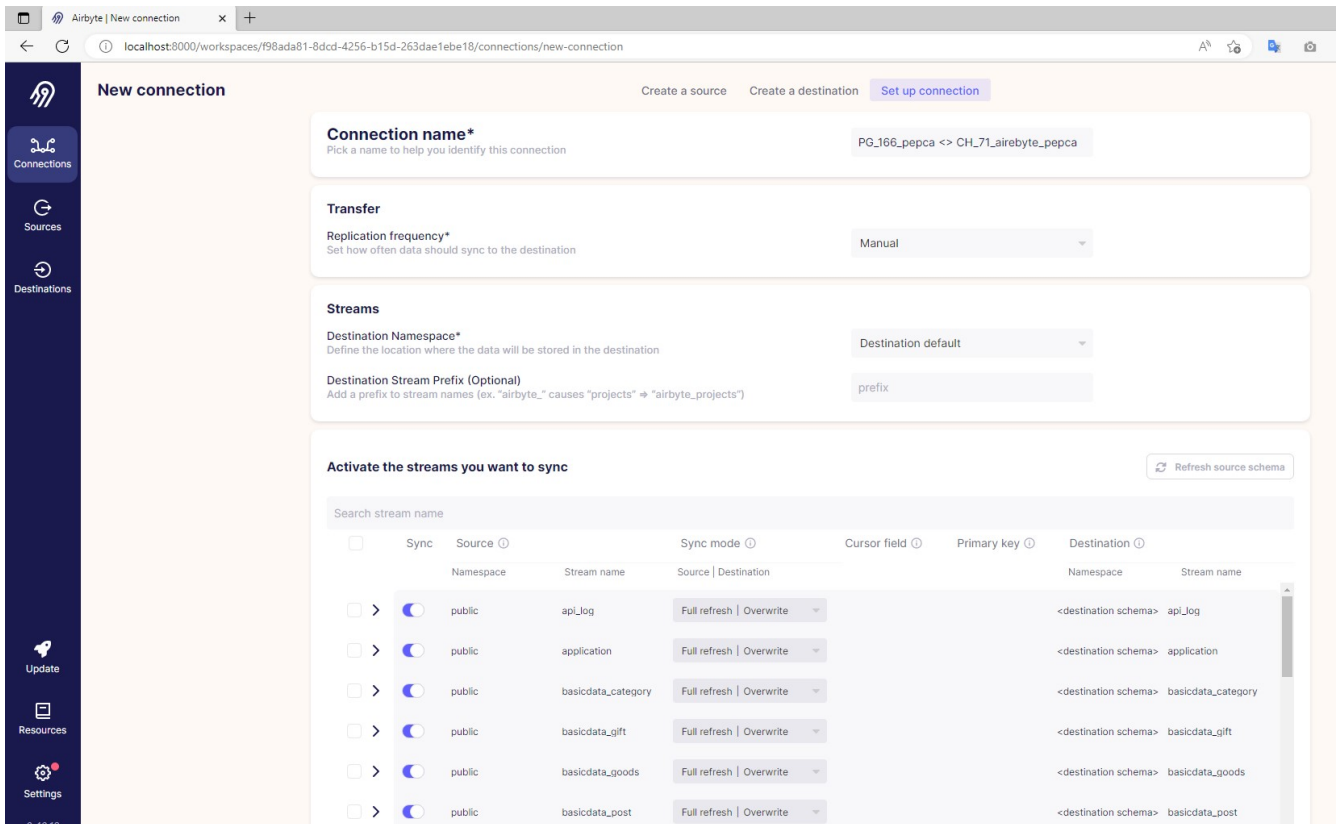
接下来选择 `select an existing destination`，从 `Destination` 下拉列表中选择已经配置好的目标数据库：



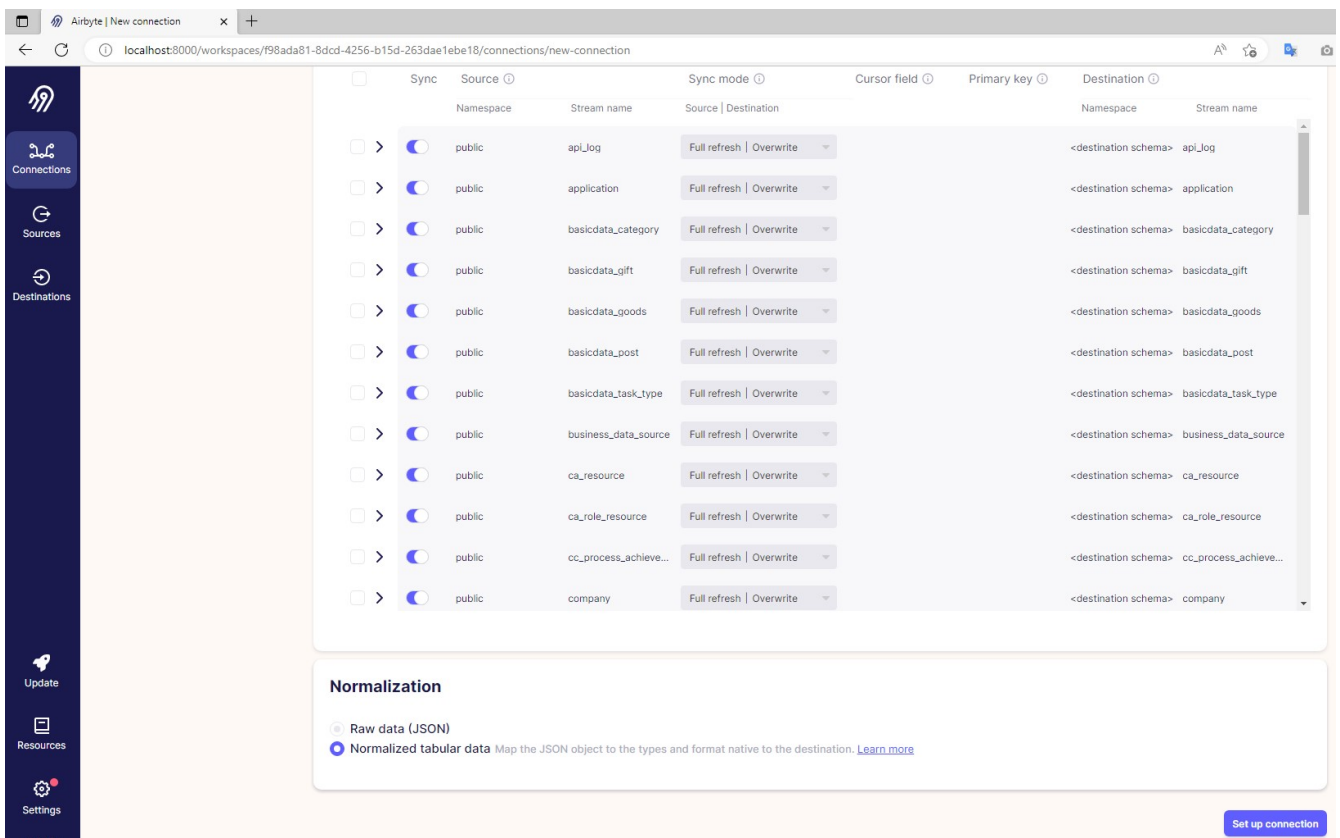
然后，系统开始建立 源数据库 和 目标数据库 之间的连接：



接下来配置 连接名称、复制频率、同步模式 等信息：

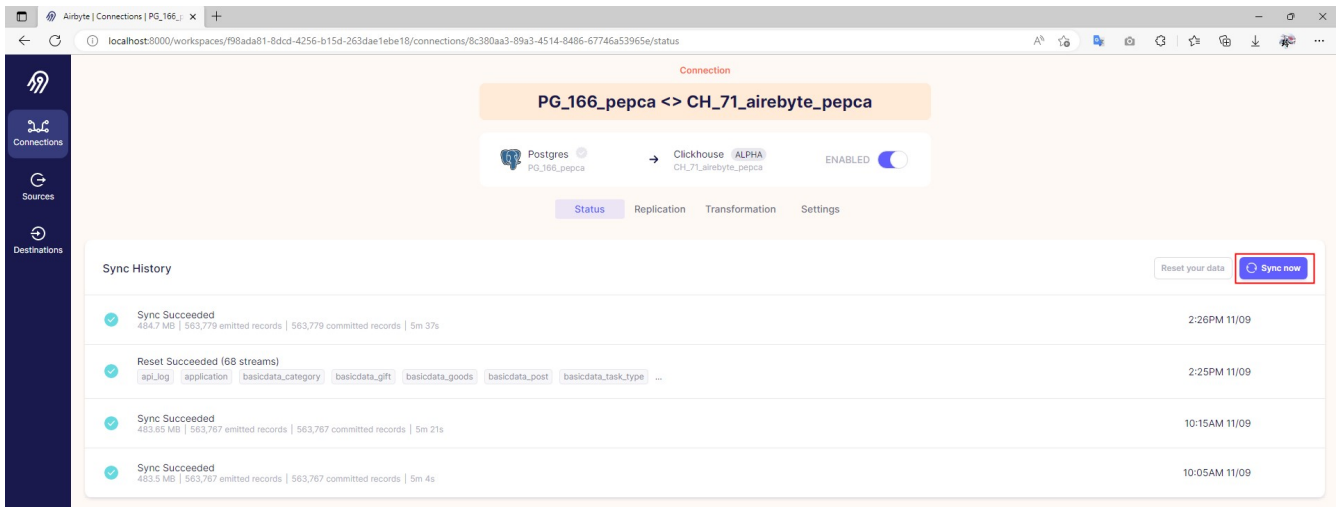


选择 标准化方式 Normalized tabular data:



## 同步数据

完成 Connection 配置后, 即可同步数据:



数据同步完成后，可以查询 ClickHouse 数据库，验证同步情况。

## Airbyte 同步方式

Airbyte 平台中给出了 4 中数据同步方式：

- Source: Full refresh | Dest: Overwrite
- Source: Full refresh | Dest: Append
- Source: Incremental | Dest: Deduped + history
- Source: Incremental | Dest: Append

经过测试，4 种方式的应用效果如下：

- **Source: Full refresh | Dest: Overwrite** 同步方式：Airbyte 从源数据库表中获取所有数据，并以覆盖的方式将数据同步至目标数据库表中。
- **Source: Full refresh | Dest: Append** 同步方式：Airbyte 从源数据库表中获取所有数据，并以追加的方式将数据同步至目标数据库表中。
- **Source: Incremental | Dest: Append** 同步方式：Airbyte 从源数据库表中获取增量数据，并以追加的方式将数据同步至目标数据库表中。
- **Source: Incremental | Dest: Deduped + history** 同步方式：这种方式与 **Source: Incremental | Dest: Append** 的差别在于，针对有数据修改的记录，Airbyte 以更新而非追加的方式同步至目标数据库表中。

**Incremental** 同步方式，需要设置游标（cursor），Airbyte 通过游标去判断有无增量数据（新增的数据、修改的数据）。通常，会使用 `updated_at` 这样的时间字段作为游标。

另外，**Incremental** 同步方式下，当源数据表有数据删除时，不影响目标数据库表中的数据，即不会同步删除目标数据库表数据。