Департамент образования и науки города Москвы
Государственное автономное образовательное учреждение высшего образования города Москвы
«Московский городской педагогический университет»
Институт цифрового образования
Департамент информатики, управления и технологий

Макарова Екатерина Павловна

ЛАБОРАТОРНАЯ РАБОТА 3

Интеграция данных из разных источников (баз данных)

Проектный практикум по разработке ETL-решений

Направление подготовки

38.03.05 Бизнес-информатика

Профиль подготовки

Аналитика данных и эффективное управление

Курс обучения: 4

Форма обучения: очная

Преподаватель: кандидат технических наук,
доцент Босенко Тимур Муртазович

Москва
2025

Цель работы: разработка ETL-процесса для интеграции данных между PostgreSQL и MySQL с использованием Pentaho Data Integration.

Задачи:

- Создать исходные таблицы в PostgreSQL с различными наборами данных.

- Настроить целевые таблицы в MySQL для приема данных.

- Разработать процессы трансформации данных в Pentaho.

- Реализовать механизмы обработки ошибок и валидации данных.

- Создать представления для связанных данных.

Задачи для самостоятельной работы:

Задание 1 (PostgreSQL) Создать таблицу projects (id, name, start_date, end_date, budget, status)

Задание 2 (MySQL) Создать таблицу project_tracking с полем completion_percentage

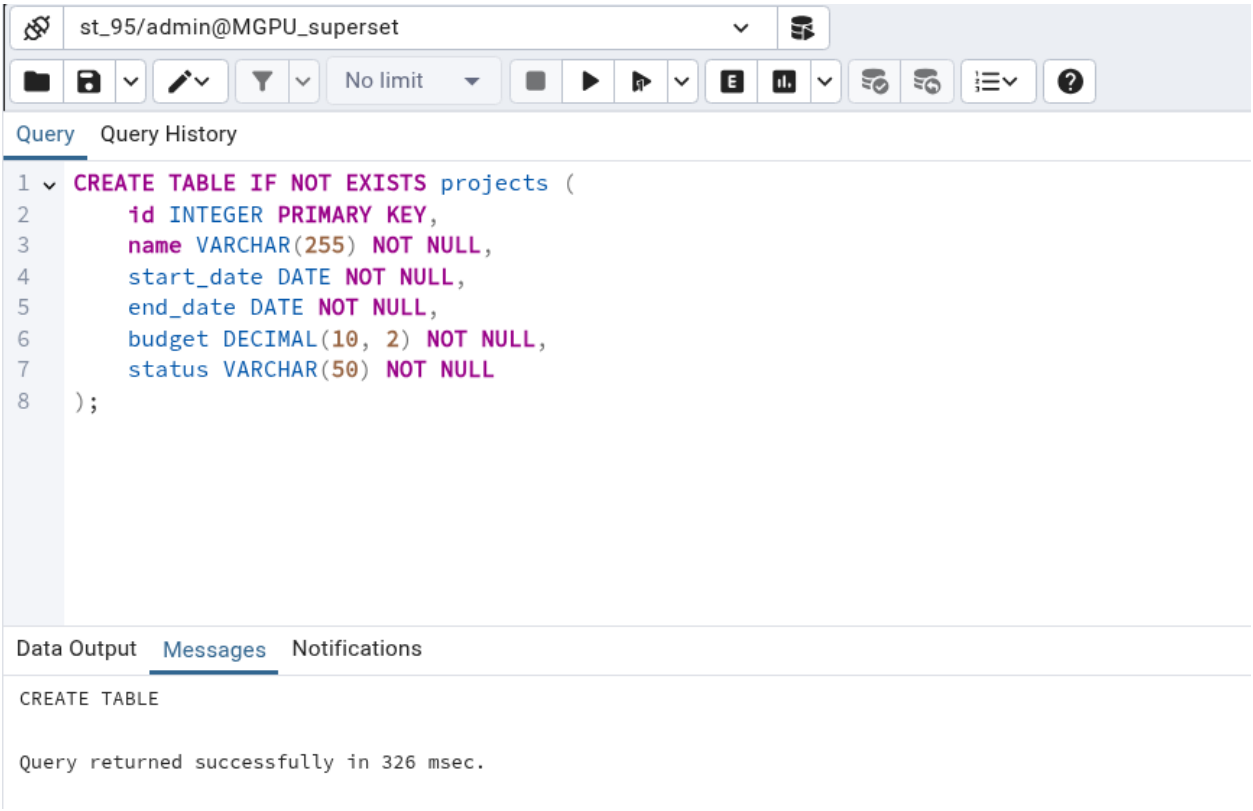
Задание 3 (Pentaho) Фильтр активных проектов

Задание 4 (Pentaho) Расчет освоения бюджета

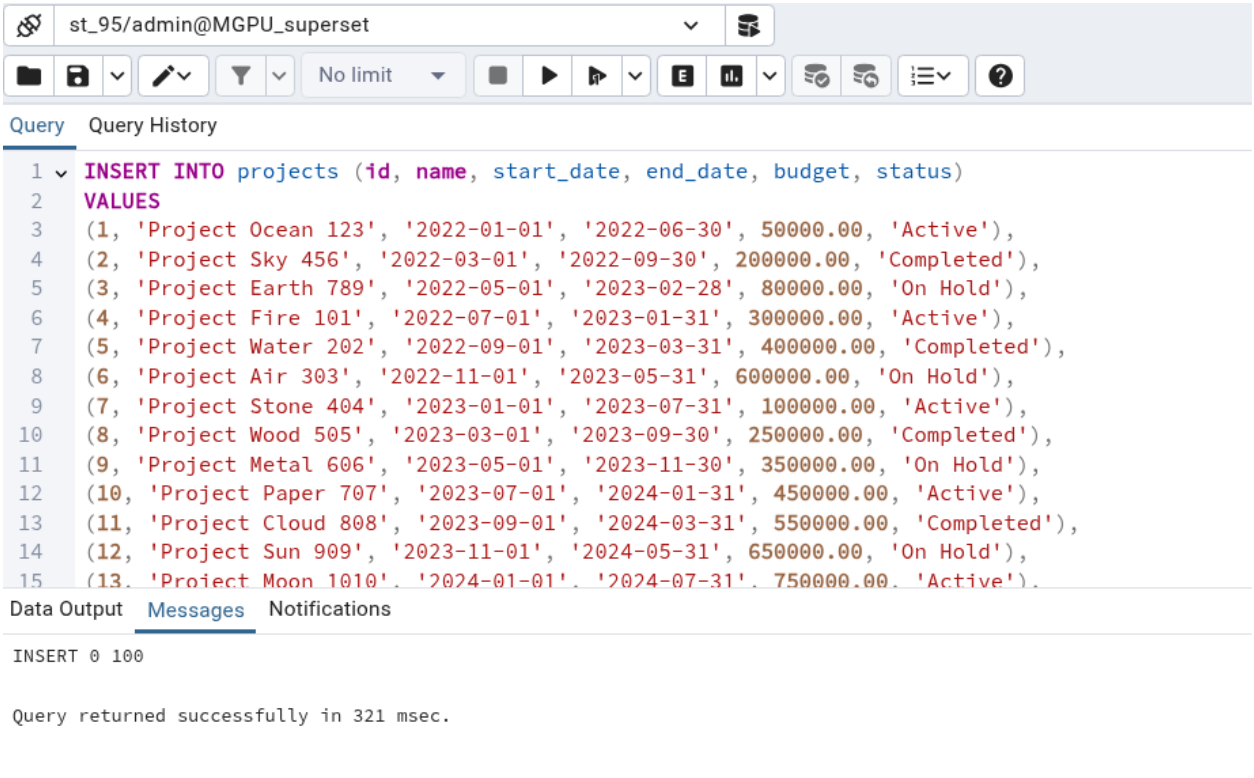
Задание 5 (Pentaho) Определение процента выполнения

Ход работы:

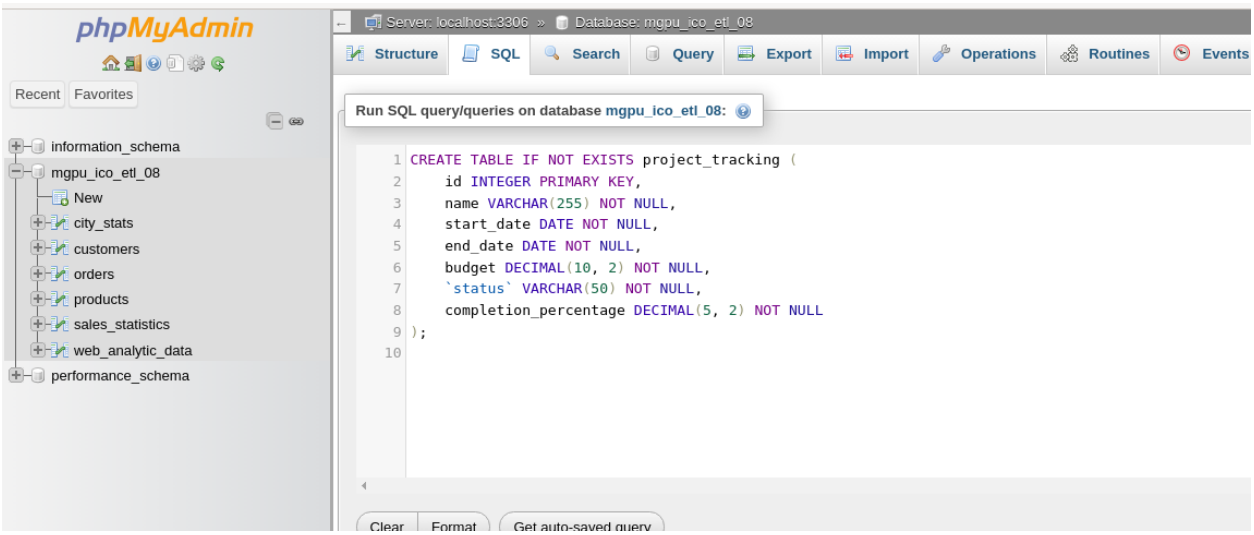
1. Создание таблицы projects в PostgreSQL



2. Вставка данных (100 записей) в таблицу projects

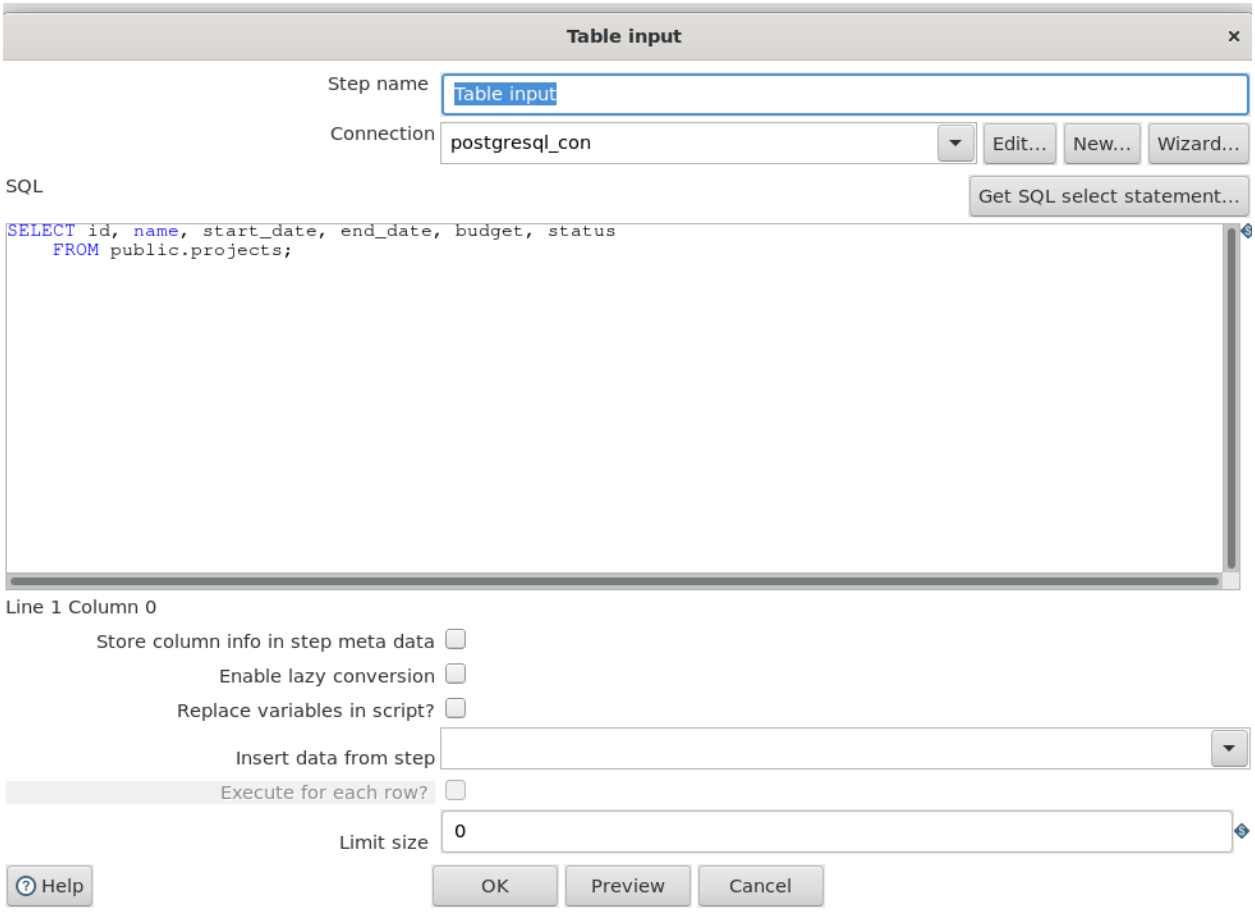


3. Создание таблицы “project_tracking”

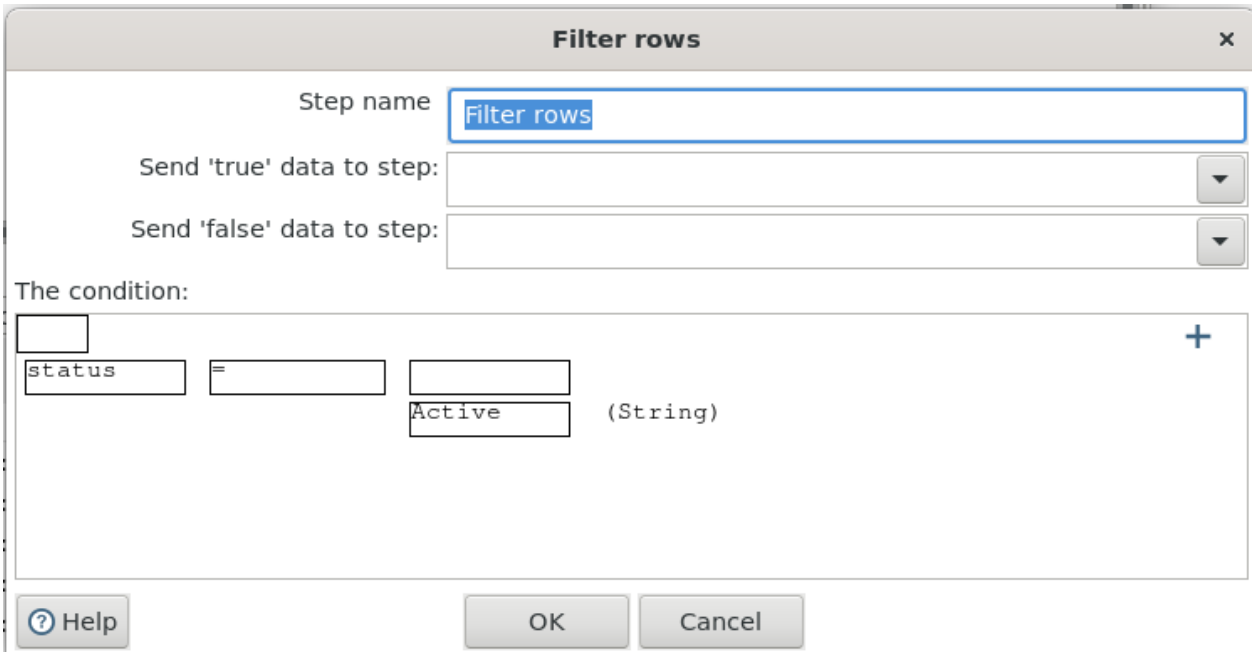


4. Трансформация

1) Создания подключения к PostgreSQL, таблица «project»



2) Фильтр проекта по статусу «Активен»



3) Для выполнения задания по расчету освоения бюджета необходимо дополнить данными по фактическим затратам:

- Создана новая таблица с данными по фактическим затратам в PostgreSQL

```sql
CREATE TABLE project_actual_costs (
    cost_id SERIAL PRIMARY KEY,          -- Уникальный ID записи о затратах
    project_id INT NOT NULL,             -- Ссылка на проект
    actual_amount DECIMAL(12, 2) NOT NULL CHECK (actual_amount > 0),  -- Сумма затрат

    -- Внешний ключ для связи с таблицей projects
    CONSTRAINT fk_project
        FOREIGN KEY (project_id)
        REFERENCES projects(id)
        ON DELETE CASCADE
);
```

- Вставляем данные по id и сумме затрат в таблицу

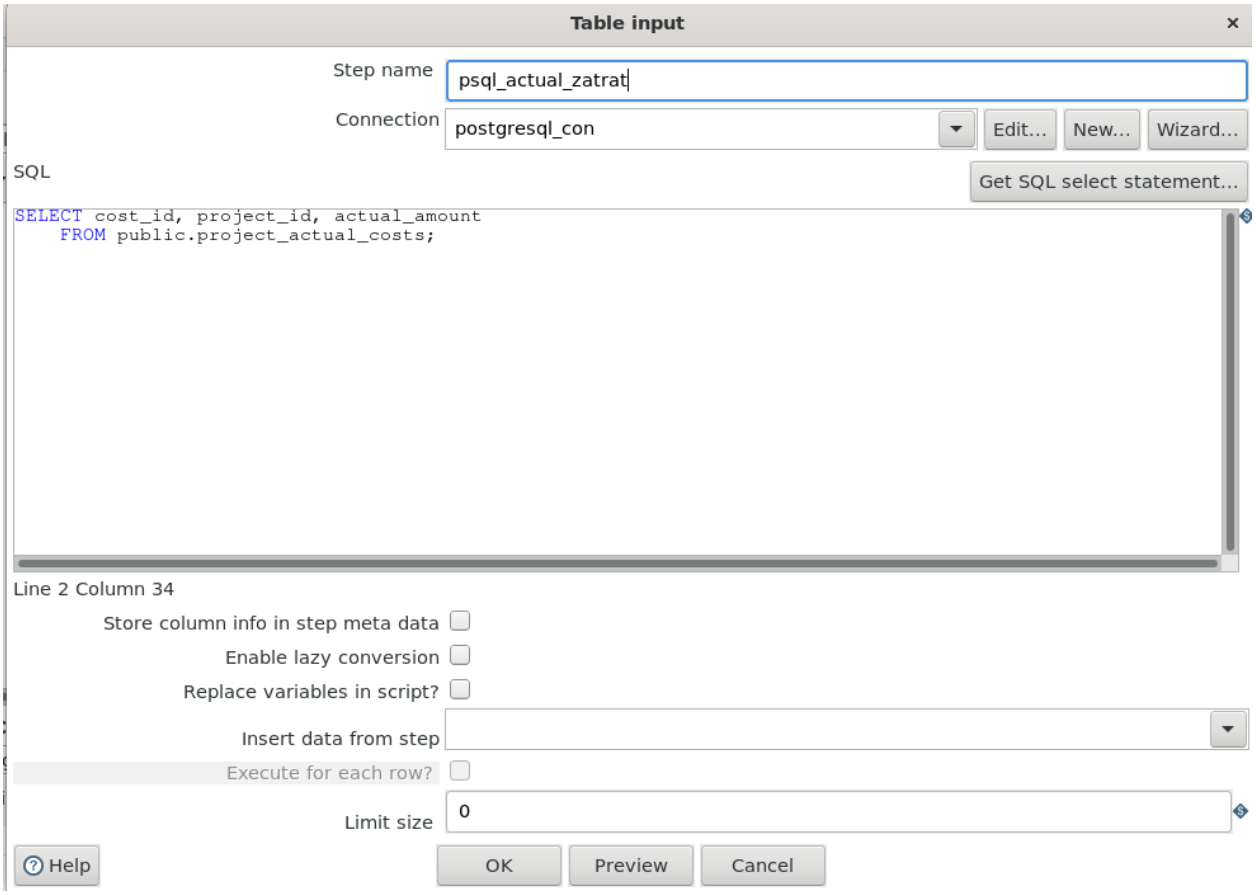
Query | Query History

```
1  INSERT INTO project_actual_costs (project_id, actual_amount)
2  VALUES
3  (1, 1500.00),
4  (2, 2500.00),
5  (3, 1200.00),
6  (4, 3000.00),
7  (5, 1800.00),
8  (6, 2200.00),
9  (7, 1600.00),
10 (8, 2800.00),
11 (9, 1400.00),
12 (10, 3200.00),
13 (11, 2000.00),
14 (12, 2400.00),
15 (13, 1700.00),
```

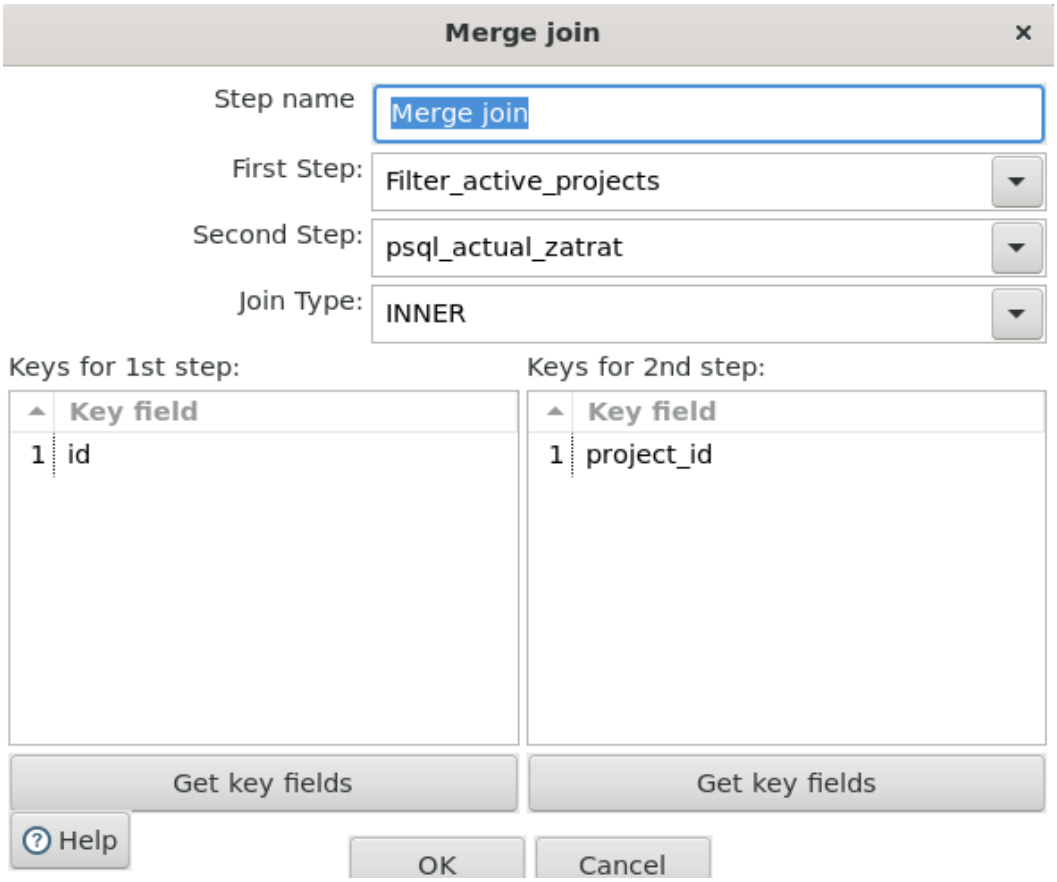
Data Output | Messages | Notifications

```
INSERT 0 100
```

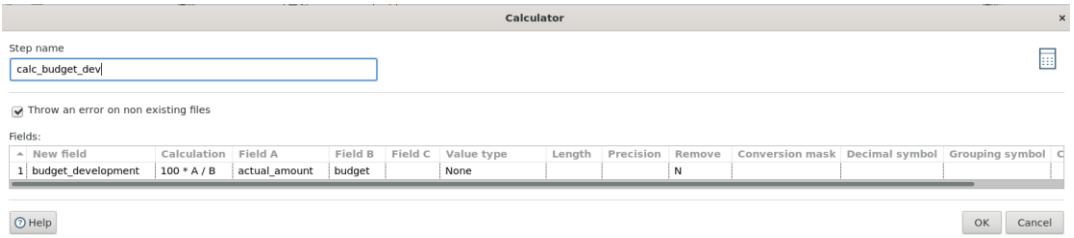
4) Создаём подключение к новой таблице в PostgreSQL



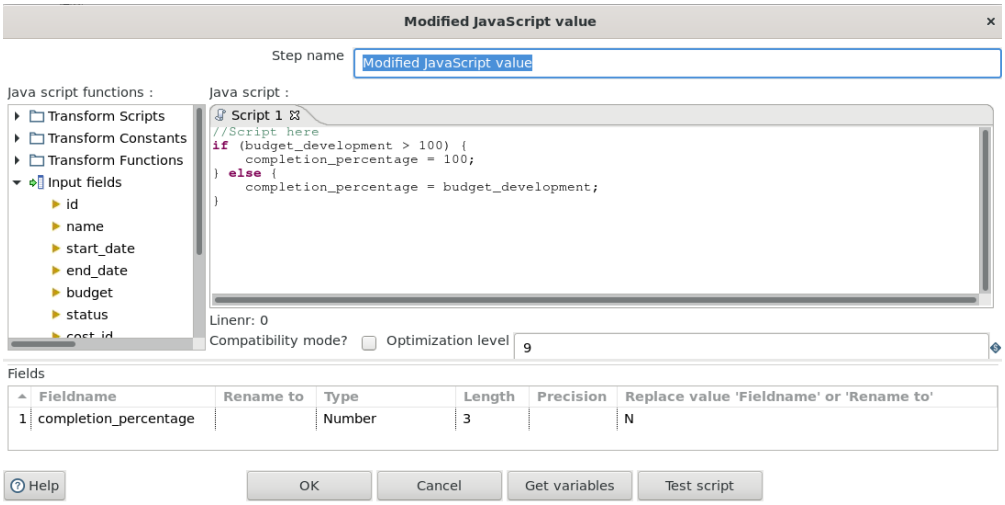
5) Объединяем данные по ключу id = project_id



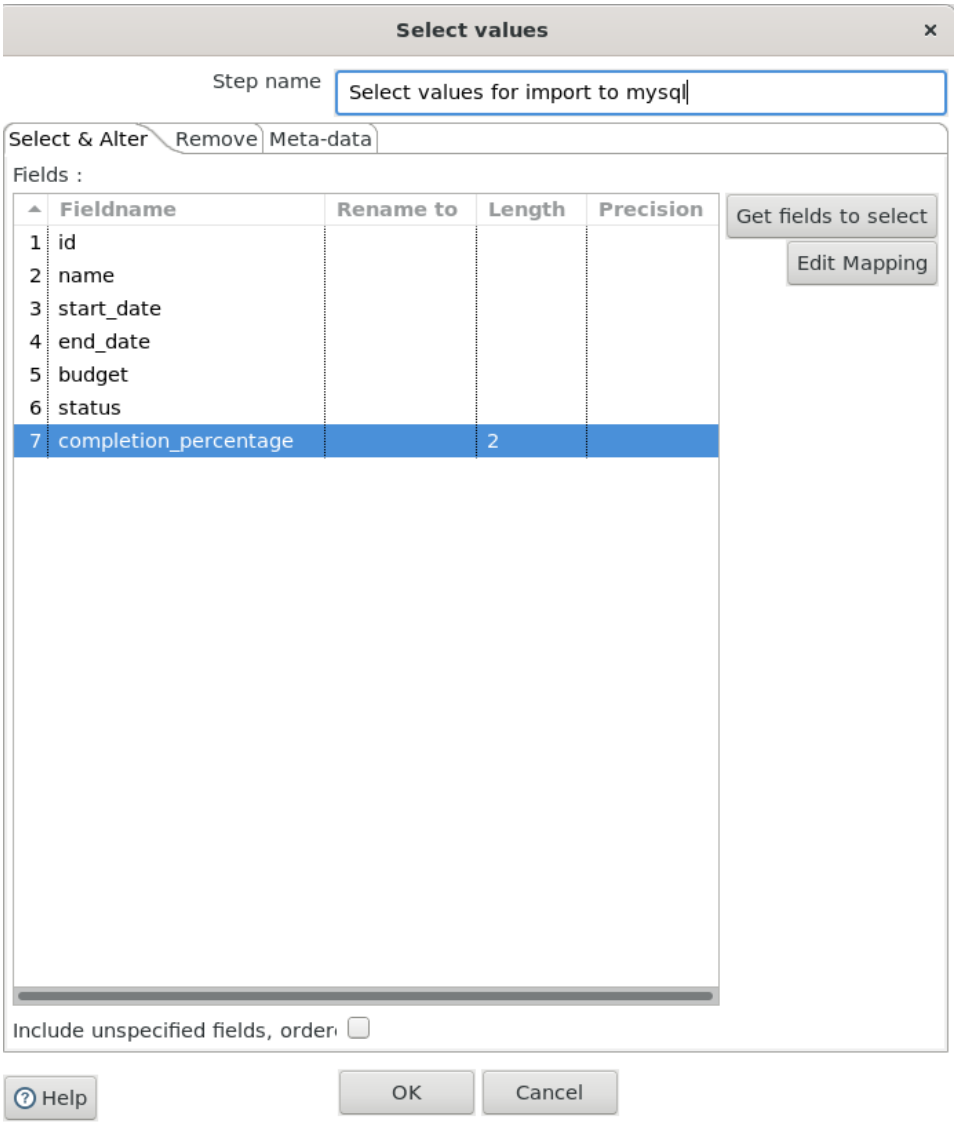
6) Расчёт процента освоения бюджета: Фактические затраты/Плановые *100



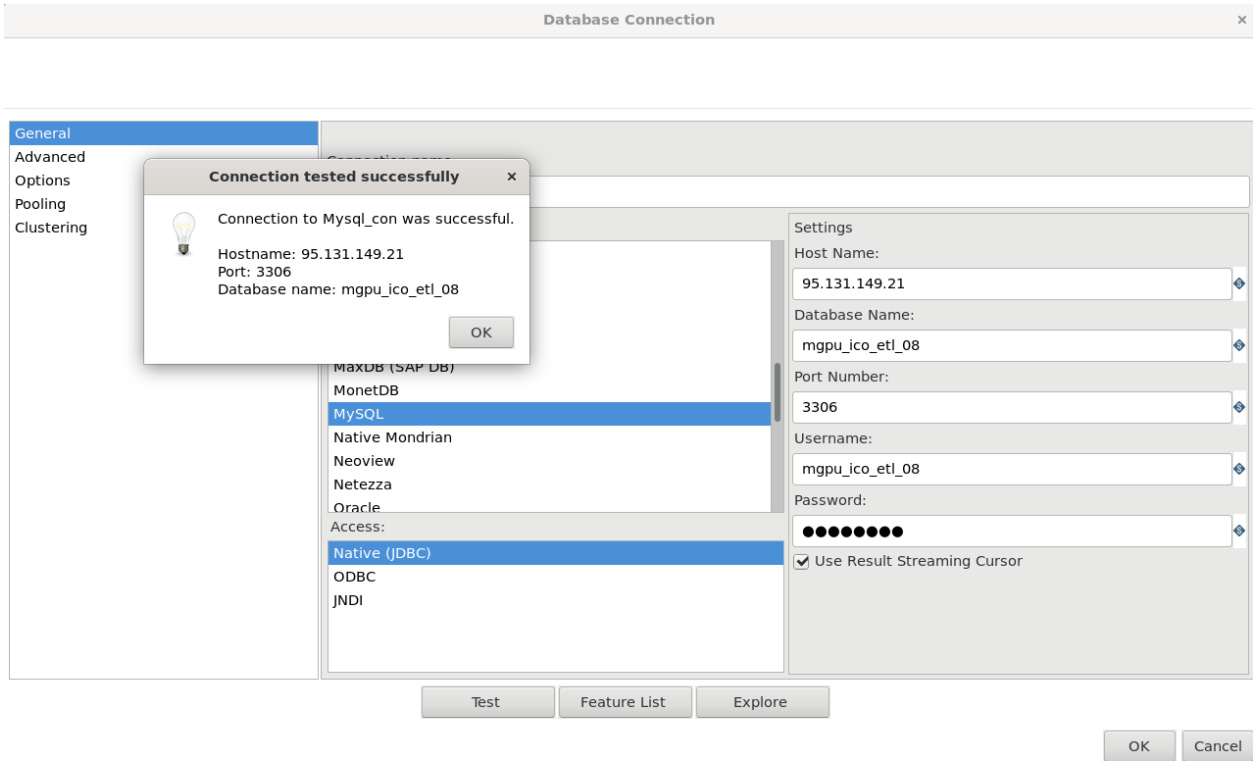
7) Рассмотрим процент завершения проекта через процент освоения бюджета, если процент освоения бюджете = 100, то проект считается завершённым, если меньше, то ставим процент освоения бюджета

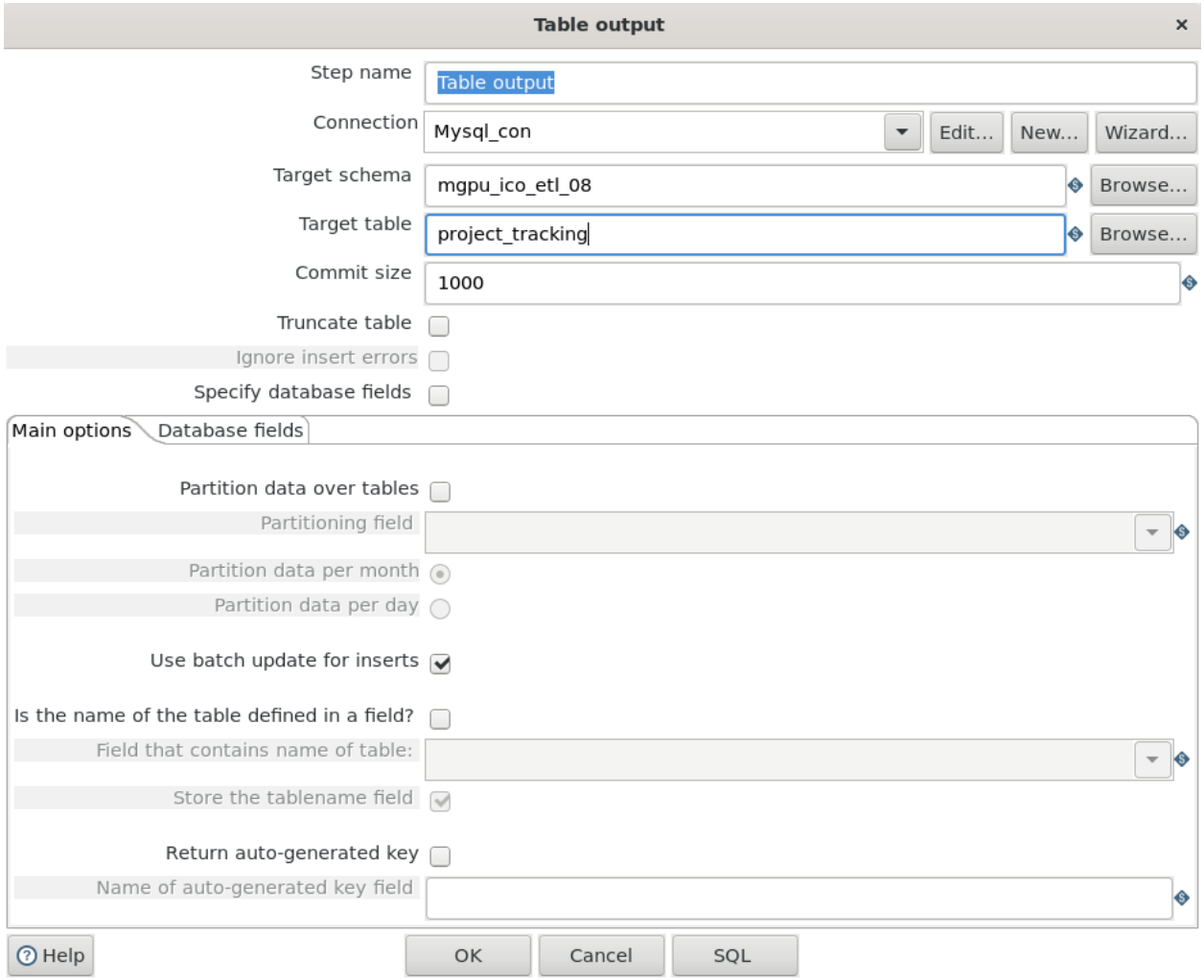


8) Выбираем атрибуты данных, которые будут импортированы в таблицу Mysql

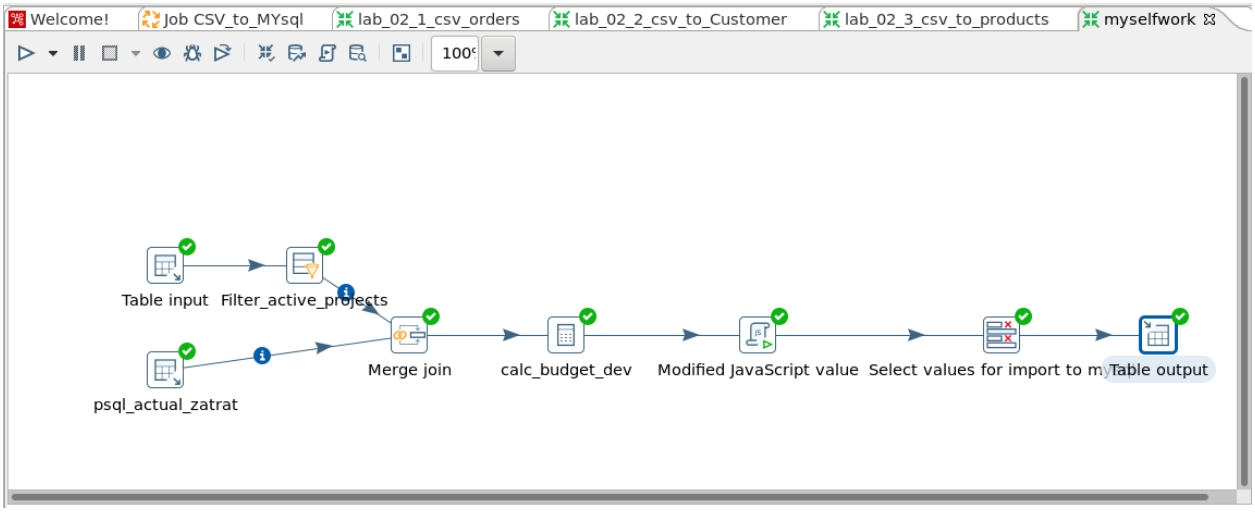


9) Настраиваем подключение к MySQL

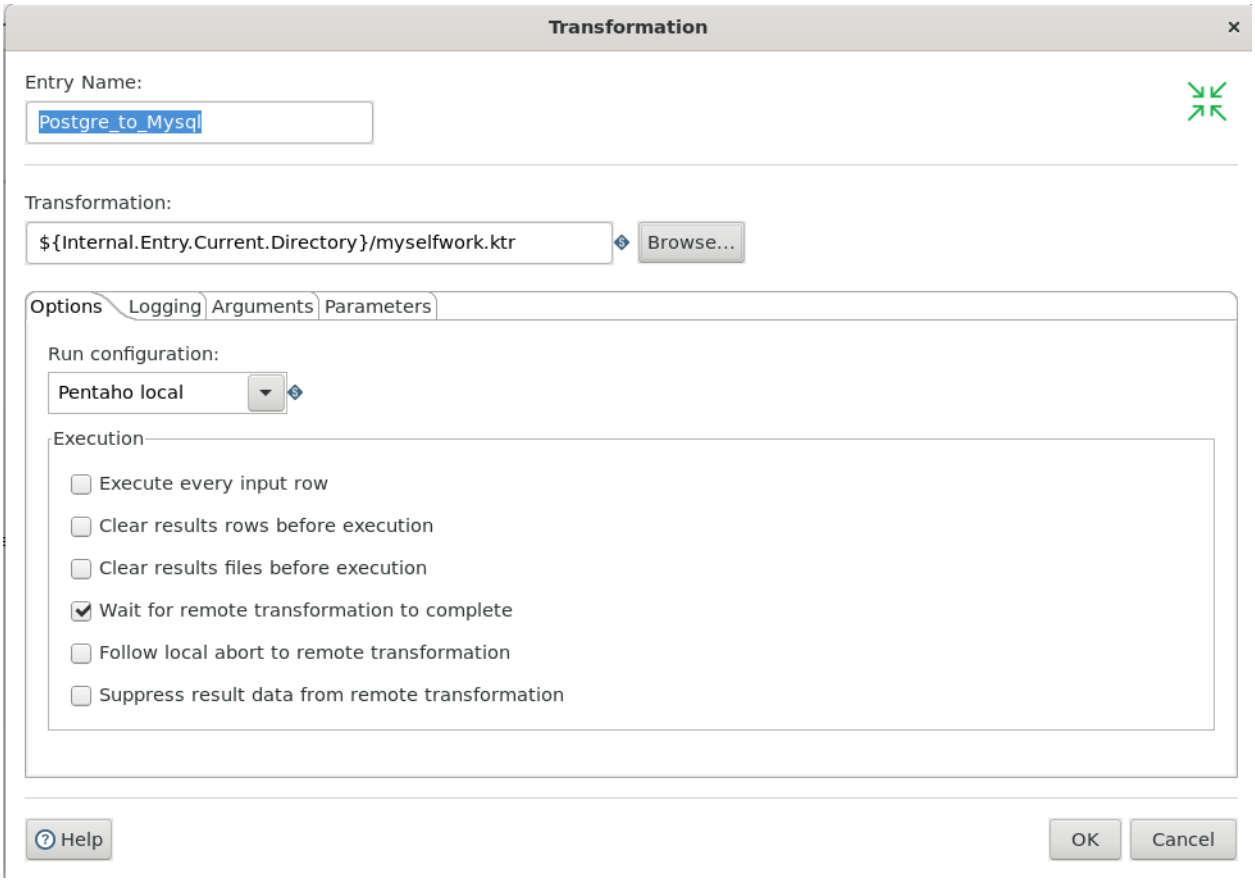




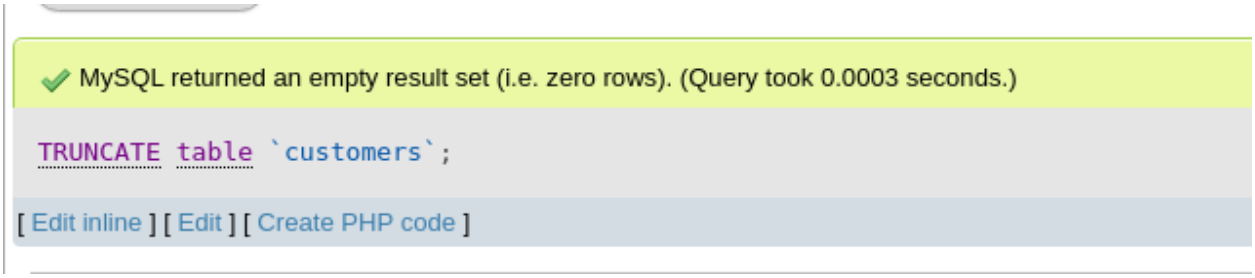
10) Трансформация имеет вид:



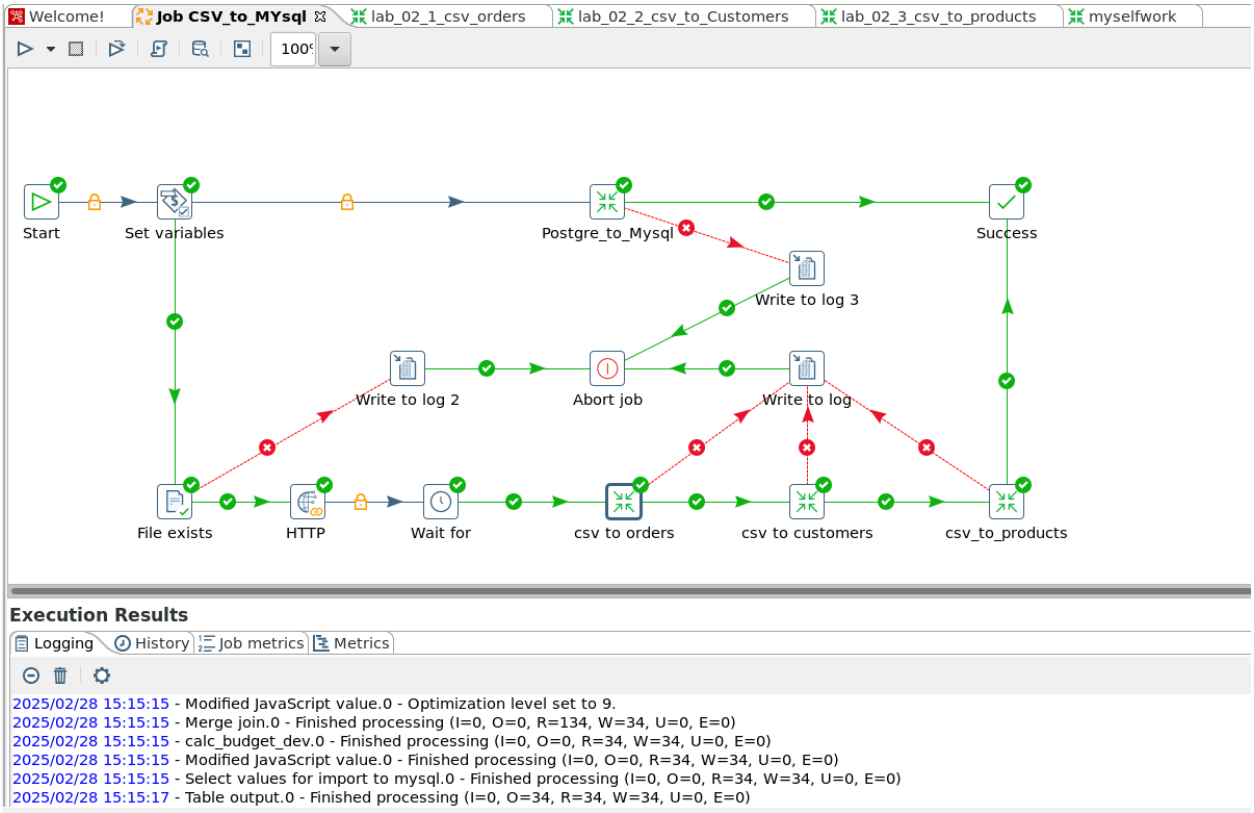
11) Настраиваем трансформацию в Job



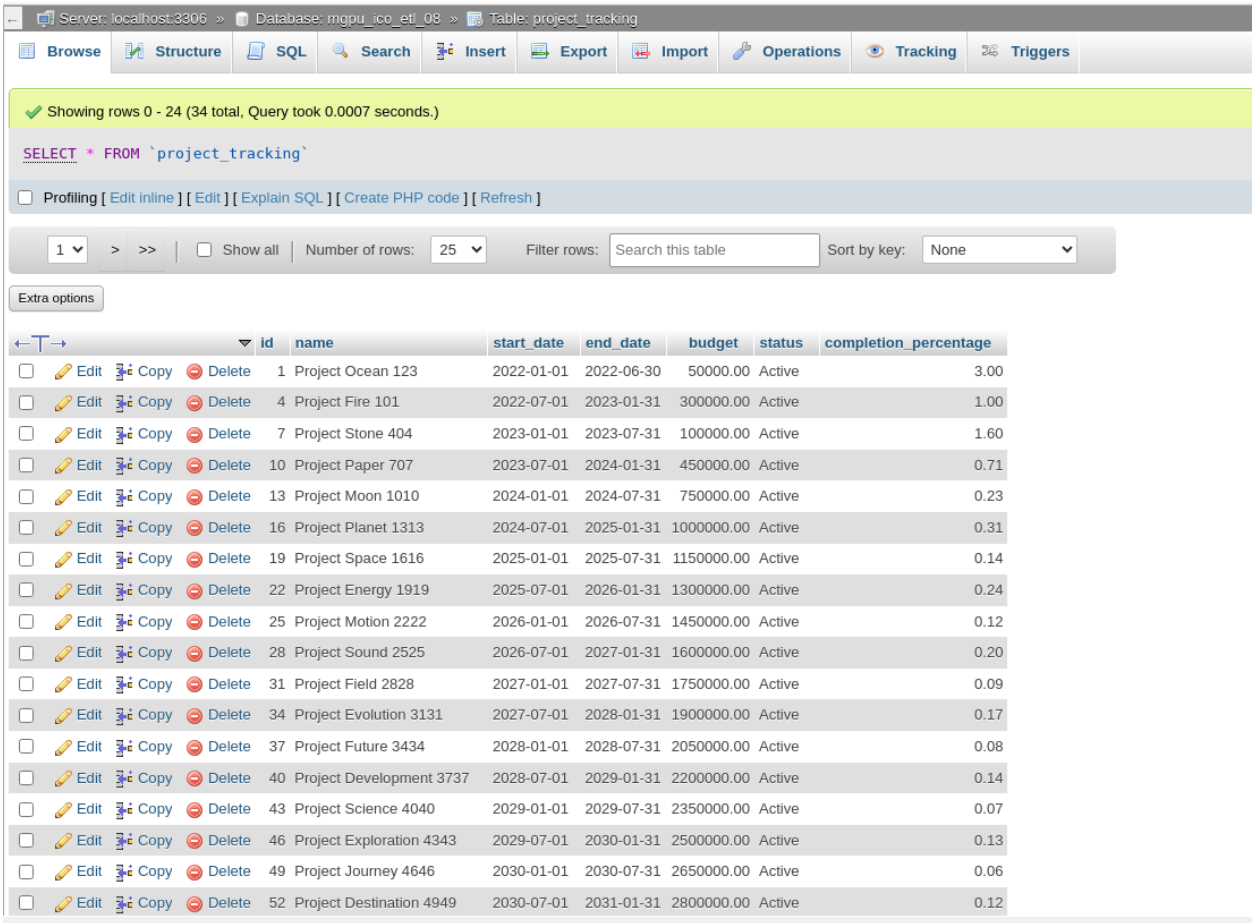
12) Очищаем таблицы customers, orders, products для импорта данных с помощью Job



13) Запускам job, он был успешно выполнен



14) Проверяем импортированные данные из PostgreSQL



	id	name	start_date	end_date	budget	status	completion_percentage
Edit Copy Delete	1	Project Ocean 123	2022-01-01	2022-06-30	50000.00	Active	3.00
Edit Copy Delete	4	Project Fire 101	2022-07-01	2023-01-31	300000.00	Active	1.00
Edit Copy Delete	7	Project Stone 404	2023-01-01	2023-07-31	100000.00	Active	1.60
Edit Copy Delete	10	Project Paper 707	2023-07-01	2024-01-31	450000.00	Active	0.71
Edit Copy Delete	13	Project Moon 1010	2024-01-01	2024-07-31	750000.00	Active	0.23
Edit Copy Delete	16	Project Planet 1313	2024-07-01	2025-01-31	1000000.00	Active	0.31
Edit Copy Delete	19	Project Space 1616	2025-01-01	2025-07-31	1150000.00	Active	0.14
Edit Copy Delete	22	Project Energy 1919	2025-07-01	2026-01-31	1300000.00	Active	0.24
Edit Copy Delete	25	Project Motion 2222	2026-01-01	2026-07-31	1450000.00	Active	0.12
Edit Copy Delete	28	Project Sound 2525	2026-07-01	2027-01-31	1600000.00	Active	0.20
Edit Copy Delete	31	Project Field 2828	2027-01-01	2027-07-31	1750000.00	Active	0.09
Edit Copy Delete	34	Project Evolution 3131	2027-07-01	2028-01-31	1900000.00	Active	0.17
Edit Copy Delete	37	Project Future 3434	2028-01-01	2028-07-31	2050000.00	Active	0.08
Edit Copy Delete	40	Project Development 3737	2028-07-01	2029-01-31	2200000.00	Active	0.14
Edit Copy Delete	43	Project Science 4040	2029-01-01	2029-07-31	2350000.00	Active	0.07
Edit Copy Delete	46	Project Exploration 4343	2029-07-01	2030-01-31	2500000.00	Active	0.13
Edit Copy Delete	49	Project Journey 4646	2030-01-01	2030-07-31	2650000.00	Active	0.06
Edit Copy Delete	52	Project Destination 4949	2030-07-01	2031-01-31	2800000.00	Active	0.12

Выводы по работе:

В ходе работы были выполнены поставленные задачи