

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

Кузьмина Дарья Юрьевна БД-241м

### **Практическая работа**

**Лабораторная работа 4.2. Динамические соединения с базами данных**

### **Вариант 11**

Направление подготовки/специальность  
38.04.05 - Бизнес-информатика  
Бизнес-аналитика и большие данные  
(очная форма обучения)

Руководитель дисциплины:  
Босенко Т.М., доцент департамента  
информатики, управления и технологий,  
доктор экономических наук

Москва  
2025

## Содержание

Введение .....	2
Основная часть .....	2
Заключение.....	4

## Введение

**Цель работы:** получить практические навыки создания ETL-процесса для загрузки данных из CSV-файла в базу данных MySQL с использованием Pentaho [Data](#) Integration.

### Задачи:

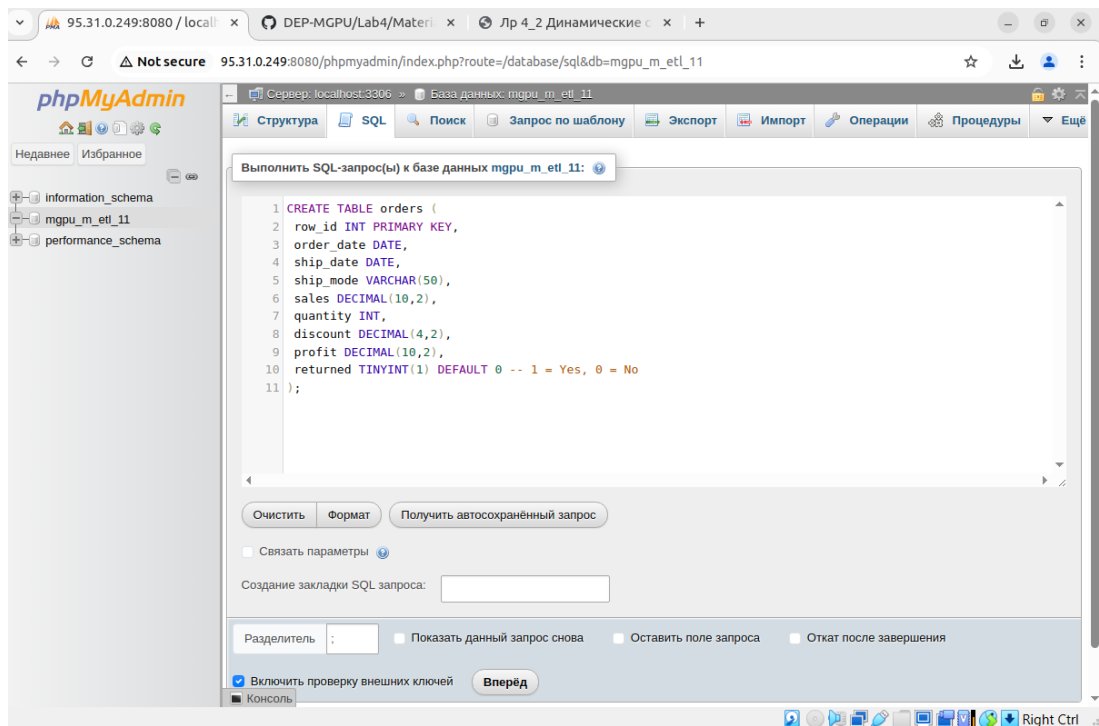
- Создать динамические подключения к различным источникам данных.
- Разработать процесс выявления и обработки дублирующихся записей.
- Реализовать механизм объединения данных в единое хранилище.
- Настроить обработку ошибок при выполнении трансформации.

## Основная часть

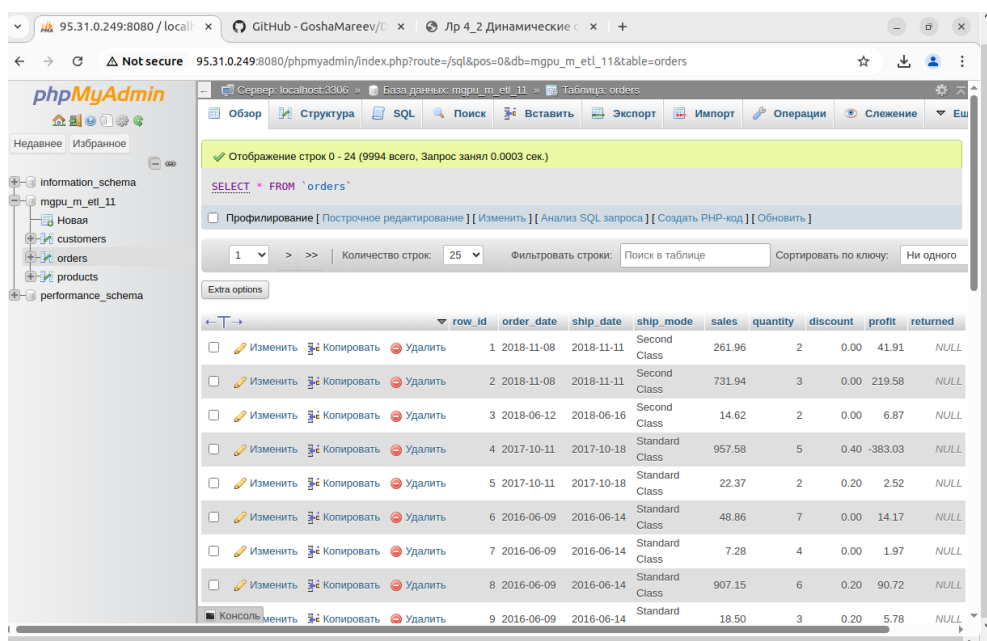
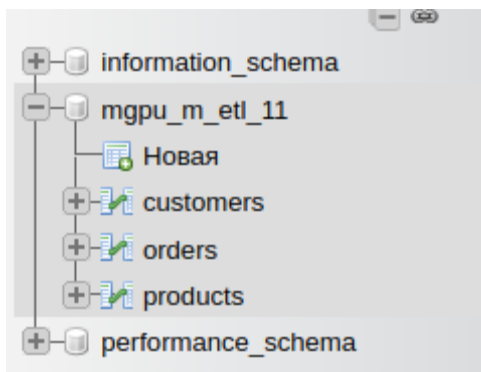
The screenshot displays the Pentaho Data Integration (Spoon) interface for a job named "lab\_02\_1\_csv\_orders". The main workspace shows a sequence of transformation steps: "CSV file input", "Select values", "Memory group by", "Filter rows", "Value mapper", and "table orders". A yellow box labeled "Поток Orders" is positioned above the "Filter rows" step. The left sidebar shows the project structure under "Transformations". The bottom pane, titled "Execution Results", contains a log of the job's execution, showing successful completion of each step with timestamps and metrics.

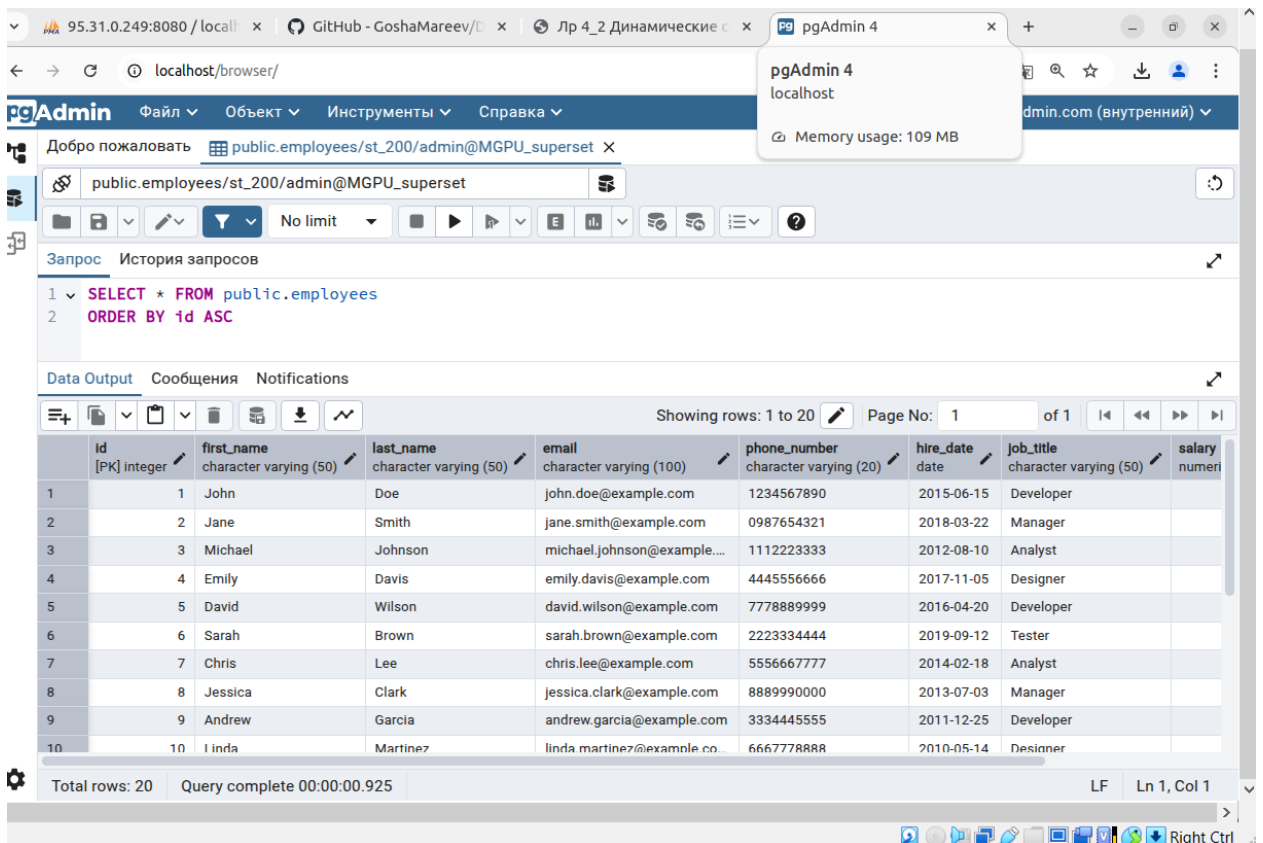
**Execution Results Log:**

- 2025/03/22 22:29:53 - CSV file input.0 - Header row skipped in file 'file:///home/dev/Downloads/lab\_etl/data\_for\_labs/lab\_csvexl\_to\_mysql/dat
- 2025/03/22 22:29:54 - CSV file input.0 - Finished processing (I=9995, O=0, R=0, W=9994, U=0, E=0)
- 2025/03/22 22:29:55 - Select values.0 - Finished processing (I=0, O=0, R=9994, W=19988, U=0, E=0)
- 2025/03/22 22:29:55 - Dummy (do nothing).0 - Finished processing (I=0, O=0, R=9994, W=9994, U=0, E=0)
- 2025/03/22 22:29:56 - Memory group by.0 - Finished processing (I=0, O=0, R=9994, W=9994, U=0, E=0)
- 2025/03/22 22:29:56 - Filter rows.0 - Finished processing (I=0, O=0, R=9994, W=9994, U=0, E=0)
- 2025/03/22 22:29:56 - Value mapper.0 - Finished processing (I=0, O=0, R=9994, W=9994, U=0, E=0)



Делаем это с остальными таблицами





## Заключение

В рамках лабораторной работы были получены практические навыки создания ETL-процесса для загрузки данных из CSV-файла в базу данных MySQL с использованием Pentaho Data Integration. Были созданы динамические подключения к различным источникам данных, разработан механизм обработки дублирующихся записей и объединения данных в единое хранилище. Также была настроена обработка ошибок при выполнении трансформации.

### Основные выводы:

- Динамические подключения позволяют гибко работать с различными источниками данных.
- Обработка дубликатов и ошибок является важным этапом для обеспечения целостности данных.
- Использование ETL-инструментов значительно упрощает процесс интеграции и анализа данных.