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 4.3. Интеграция данных из нескольких источников.**

**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](#_Toc193677859)

[Основная часть 2](#_Toc193677860)

[Заключение 4](#_Toc193677861)

## Введение

**Цель работы**: получить практические навыки создания ETL-процесса для загрузки данных из CSV-файла в базу данных MySQL с использованием Pentaho [Data](http://95.31.0.249/moodle/mod/folder/view.php?id=1500) 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

Изображение выглядит как текст, программное обеспечение, снимок экрана, веб-страница

Автоматически созданное описание

Изображение выглядит как текст, программное обеспечение, снимок экрана, Значок на компьютере

Автоматически созданное описание

Делаем это с остальными таблицами

Изображение выглядит как текст, снимок экрана, Шрифт, число

Автоматически созданное описание

Изображение выглядит как текст, программное обеспечение, Значок на компьютере, веб-страница

Автоматически созданное описание

Изображение выглядит как текст, снимок экрана, программное обеспечение, веб-страница

Автоматически созданное описание

# Заключение

Целью работы было получение навыков интеграции, обработки и согласования данных из различных источников. Были изучены методы чтения данных из CSV, Excel и баз данных (MySQL, PostgreSQL), а также техники их очистки и согласования. Обработанные данные были сохранены в единое хранилище для дальнейшего анализа.

* Интеграция данных из нескольких источников требует тщательной подготовки и согласования форматов данных.
* Использование ETL-процессов позволяет автоматизировать обработку и объединение данных.
* Качество данных напрямую влияет на точность аналитических выводов, поэтому важна предварительная очистка и проверка данных.