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

Мошенина Елена Дмитриевна БД-241м

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

**Вариант 15**

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,

кандидат технических наук

Москва

2025

**Цель работы**: получить практические навыки создания ETL-процесса для загрузки данных из CSV-файла в базу данных MySQL с использованием Pentaho [Data](http://95.131.149.21/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.

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

Заходим по ссылке: <http://95.131.149.21:8080/phpmyadmin>

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

Контент, сгенерированный ИИ, может содержать ошибки.

CREATE TABLE orders (

row\_id INT PRIMARY KEY,

order\_date DATE,

ship\_date DATE,

ship\_mode VARCHAR(50),

sales DECIMAL(10,2),

quantity INT,

discount DECIMAL(4,2),

profit DECIMAL(10,2),

returned TINYINT(1) DEFAULT 0 -- 1 = Yes, 0 = No

);

CREATE TABLE customers (

id INT AUTO\_INCREMENT PRIMARY KEY,

customer\_id VARCHAR(20) NOT NULL,

customer\_name VARCHAR(100),

segment VARCHAR(50),

country VARCHAR(100),

city VARCHAR(100),

state VARCHAR(100),

postal\_code VARCHAR(20),

region VARCHAR(50),

INDEX idx\_customer\_id (customer\_id),

INDEX idx\_region (region)

);

CREATE TABLE products (

id INT AUTO\_INCREMENT PRIMARY KEY,

product\_id VARCHAR(20) NOT NULL,

category VARCHAR(50),

sub\_category VARCHAR(50),

product\_name VARCHAR(255),

person VARCHAR(100),

INDEX idx\_product\_id (product\_id),

INDEX idx\_category (category),

INDEX idx\_subcategory (sub\_category)

);

ALTER TABLE orders ADD INDEX idx\_order\_date (order\_date);

ALTER TABLE orders ADD INDEX idx\_ship\_date (ship\_date);

ALTER TABLE customers ADD INDEX idx\_region (region);

ALTER TABLE products ADD INDEX idx\_category (category);

ALTER DATABASE mgpu\_ico\_etl\_prepod CHARACTER SET utf8mb4 COLLATE

utf8mb4\_unicode\_ci;

ALTER TABLE orders CONVERT TO CHARACTER SET utf8mb4 COLLATE

utf8mb4\_unicode\_ci;

ALTER TABLE customers CONVERT TO CHARACTER SET utf8mb4 COLLATE

utf8mb4\_unicode\_ci;

ALTER TABLE products CONVERT TO CHARACTER SET utf8mb4 COLLATE

utf8mb4\_unicode\_ci;

Изображение выглядит как программное обеспечение, текст, Мультимедийное программное обеспечение, Значок на компьютере

Контент, сгенерированный ИИ, может содержать ошибки.

Настраиваем все соединения:

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

Контент, сгенерированный ИИ, может содержать ошибки.

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

Контент, сгенерированный ИИ, может содержать ошибки.

Для каждой таблицы проделываем Run:

Изображение выглядит как текст, программное обеспечение, Значок на компьютере, Мультимедийное программное обеспечение

Контент, сгенерированный ИИ, может содержать ошибки.

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

Контент, сгенерированный ИИ, может содержать ошибки.

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

Контент, сгенерированный ИИ, может содержать ошибки.

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

Контент, сгенерированный ИИ, может содержать ошибки.

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

Контент, сгенерированный ИИ, может содержать ошибки.

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

Контент, сгенерированный ИИ, может содержать ошибки.

Структура Job:

Start

├── Set Variables (FILE\_PATH)

├── Check File Exists

├── HTTP Download

├── Transform: Load Orders (lab\_02\_1)

├── Transform: Load Customers (lab\_02\_2)

├── Transform: Load Products (lab\_02\_3)

├── Transform: Analytics 1 (lab\_02\_4)

└── Transform: Analytics 2 (lab\_02\_5)

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

Контент, сгенерированный ИИ, может содержать ошибки.

**Варианты индивидуальных заданий 4.2.**

15 Фильтр по подкатегории: только Art

**Изображение выглядит как снимок экрана, текст, программное обеспечение, компьютер

Контент, сгенерированный ИИ, может содержать ошибки.**

**SELECT \* FROM products WHERE sub\_category = 'Art'**

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

Контент, сгенерированный ИИ, может содержать ошибки.

**Вывод.**

В ходе работы были достигнуты поставленные цели и задачи. Практические навыки создания ETL-процесса для загрузки данных из CSV-файла в базу данных MySQL с использованием Pentaho Data Integration были успешно получены.

Были созданы динамические подключения к различным источникам данных, что позволило гибко настраивать процесс извлечения информации. Разработан процесс выявления и обработки дублирующихся записей, что обеспечило качество и целостность данных в целевом хранилище. Реализован механизм объединения данных, что способствовало консолидации информации из разных источников в едином месте. Настроена обработка ошибок при выполнении трансформации, что повысило надёжность и устойчивость ETL-процесса.

Полученные навыки и опыт могут быть полезны при работе с большими объёмами данных и в задачах, требующих автоматизации процессов извлечения, преобразования и загрузки данных.