Департамент образования города Москвы

Государственное автономное образовательное учреждение

высшего образования города Москвы

«Московский городской педагогический университет»

Институт цифрового образования

Департамент информатики, управления и технологий

ОТЧЕТ

по дисциплине «Проектный практикум по разработке ETL-решений»

Направление подготовки 38.03.05 – бизнес-информатика

Профиль подготовки «Аналитика данных и эффективное управление»

(очная форма обучения)

Лабораторная работа 1.1 Установка и настройка ETL-инструмента. Создание конвейеров данных

Выполнил:

st-89

Руководитель:

Москва  
2025

**Цель работы:** изучение основных принципов работы с ETL-инструментами на примере Pentaho Data Integration (PDI), настройка конвейера обработки данных, фильтрация и замена значений в Excel-файле, а также выгрузка обработанных данных в базу данных MySQL/PostgreSQL.

**Задача:** Анализ электронной коммерции: очистка данных о заказах, сегментация клиентов **Вариант 2** - <https://www.kaggle.com/datasets/carrie1/ecommerce-data>

**Ход работы:**

1. Для начала скачаем файл csv Kaggle. Далее уставим MySQL на виртуальную машину (Рис. 1)

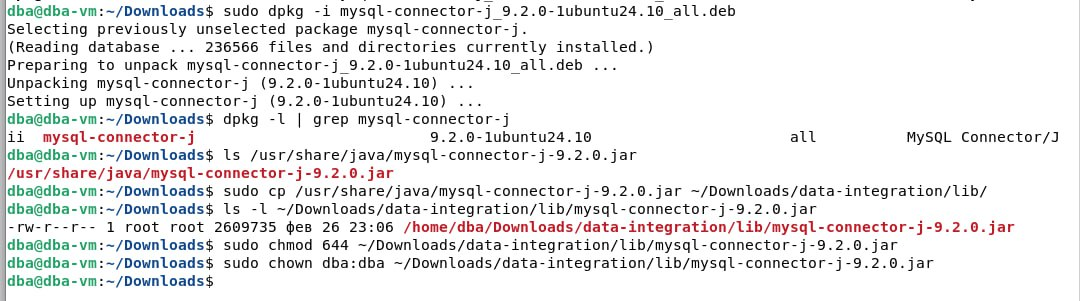


Рис. 1 – Установка MySQL

Далее загрузим файл CSV с помощью узла **File input** и разделим ETL процесс на два «потока» (Рис. 2):

* Группировка по CustomerID и с помощью узла **Filter Rows** убираем NULL значения для дальнейшей загрузки в базу. Итог – Вывод CustomerID, Country, price\_sum и count\_tovar (Рис. 3)
* Группировка по StockCode и с помощью узла **Filter Rows** убираем NULL значения для дальнейшей загрузки в базу. Итог – Вывод StockCode и count\_tovar (Рис. 4)

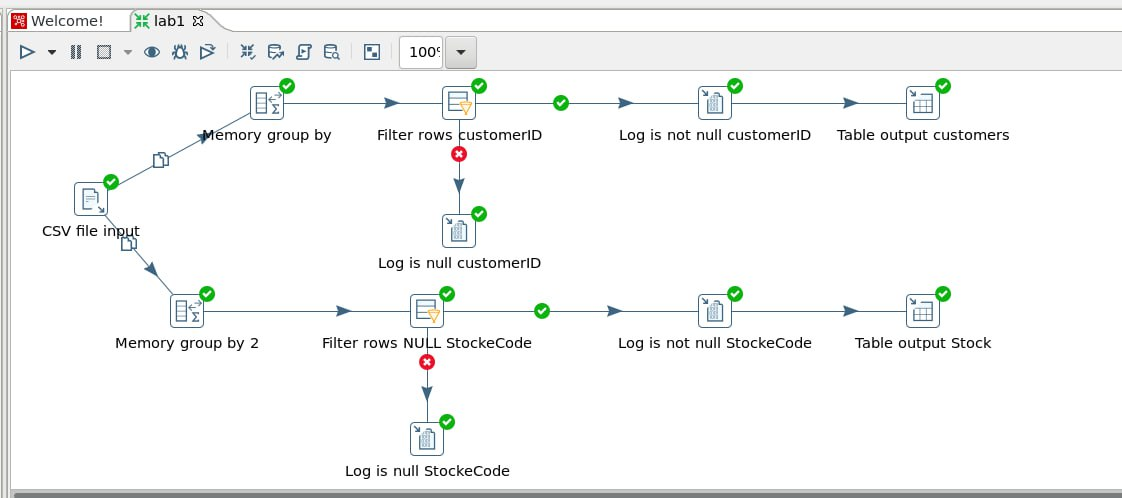


Рис. 2 – ETL процесс обработки данных заказов

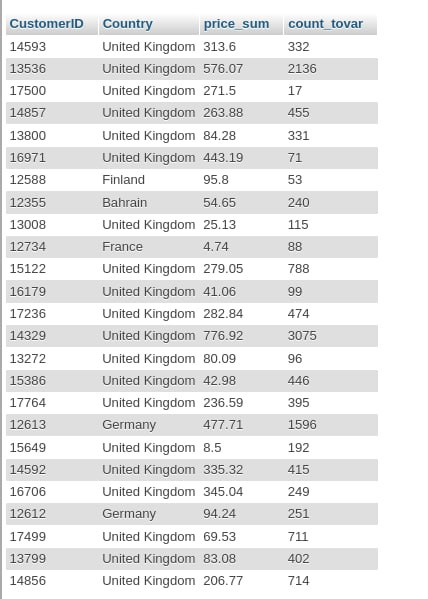


Рис. 3 – Группировка по CustomerID

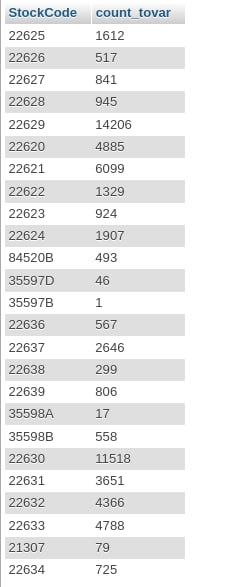


Рис. 4 – Группировка по stockCode

**Вывод:** Была проведена предобработка данных и дальнейшая загрузка в phpMyAdmin с помощью ETL процесса.

**ПРИЛОЖЕНИЕ**

CREATE TABLE infoCustomers (

CustomerID VARCHAR(20) NOT NULL,

Country VARCHAR(50) NOT NULL,

price\_sum VARCHAR(20) NOT NULL,

count\_tovar VARCHAR(20) NOT NULL

);

CREATE TABLE stock (

StockCode VARCHAR(20) NOT NULL,

count\_tovar VARCHAR(20) NOT NULL

);