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

Ли Александр Андреевич БД-241м

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

Направление подготовки/специальность

38.04.05 - Бизнес-информатика

Бизнес-аналитика и большие данные

(очная форма обучения)

Вариант 12

Москва

2024

**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:**

1.Создать динамические подключения к различным источникам данных.

2.Разработать процесс выявления и обработки дублирующихся записей.

3.Реализовать механизм объединения данных в единое хранилище.

Индивидуально задание

Фильтр: срок доставки > 5 дней , отчет по прибыли, анализ продаж

Общая структура трансформации

Изображение выглядит как текст, снимок экрана, программное обеспечение, Шрифт

Автоматически созданное описание

Рис.1 общая структура трансформации

Импортируем данные из samplestore, указываем путь и разделитель

Изображение выглядит как текст, снимок экрана, число, программное обеспечение

Автоматически созданное описание

Рис.2 указываем путь к данным

Проверяем полученные данные через preview data

Изображение выглядит как текст, снимок экрана, меню, число

Автоматически созданное описание

Рис.3 проверяем полученные данные

Добавляем select values, проверяем чтоб называние столбцов было без пробелов и их тип данных.

Изображение выглядит как текст, снимок экрана, программное обеспечение, число

Автоматически созданное описание

Рис.4 переименовываем столбцы

По заданию, считаем разницу во времени между доставкой и датой заказа.

Изображение выглядит как текст, снимок экрана, программное обеспечение, число

Автоматически созданное описание

Рис.5 рассчитываем дни на доставку

В pentaho при вычитании дат **Date A - Date B, результат возвращается в миллисекундах**

1 секунда = 1000 миллисекунд

1 минута = 60000 миллисекунд

1 час = 3 600 000 миллисекунд

1 день = 86 400 000 миллисекунд

Поэтому, чтобы перевести миллисекунды в дни, делим на 86 400 000

Для этого добавляем константу 86 400 000

Изображение выглядит как текст, линия, число, Шрифт

Автоматически созданное описание

Рис.6 добавляем константу

Делим время доставки на 86 400 000, чтобы рассчитать количество дней на доставку

Изображение выглядит как текст, снимок экрана, линия, число

Автоматически созданное описание

Рис.6 переводим миллисекунды в дни

Далее, добавляем фильтр, который оставляет только те значения, в которых количество дней на доставку больше 5

Изображение выглядит как текст, линия, Шрифт, число

Автоматически созданное описание

Рис.7 переводим миллисекунды в дни

Далее посчитаем общие продажи и процент маржи

Изображение выглядит как текст, снимок экрана, программное обеспечение, линия

Автоматически созданное описание

Рис.7 рассчитываем общие продажи и маржу в процентах

Посчитаем средние продажи

Изображение выглядит как текст, снимок экрана, число, Шрифт

Автоматически созданное описание

Рис.7 считаем средние продажи

И выберем данные для экспорта в целевую бд

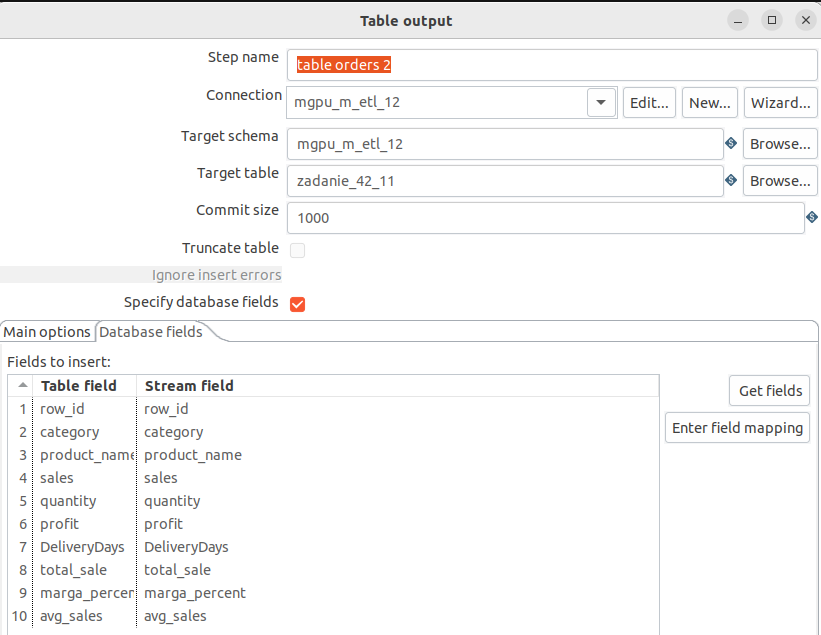


Рис.8 экспорт данных

Проверяем соединение с базой данных Изображение выглядит как текст, снимок экрана, программное обеспечение, число

Автоматически созданное описание

Рис.9 экспорт данных

Создаем таблицу в базе данных phpMyAdmin

CREATE TABLE zadanie\_42\_12(

row\_id INT AUTO\_INCREMENT PRIMARY KEY,

category VARCHAR(50),

product\_name VARCHAR(255),

sales int,

quantity int,

profit int,

DeliveryDays int,

total\_sale int,

marga\_percent int,

avg\_sales int

);

Отправляем данные в базу данных

Изображение выглядит как текст, снимок экрана, число, Параллельный

Автоматически созданное описание

Рис.10 добавляем данные в таблицу

Задание 2

Изображение выглядит как текст, снимок экрана, диаграмма, Шрифт

Автоматически созданное описание

Рис.11 структура трансформации

Select values, calculator, add constants, total\_deliverydays, filter rows такой же как и в прошлом задании

Добавляем фильтр, чтобы считались только значения, которые не вернули

Изображение выглядит как текст, снимок экрана, линия, Шрифт

Автоматически созданное описание

Рис.12 фильтр, оставляем то те заказы, которые не вернули

Считаем общие продажи по каждому региону

Изображение выглядит как текст, снимок экрана, число, Шрифт

Автоматически созданное описание

Рис.13 общие продажи по регионам

Делаем соединение с базой данных и указываем значения, которые хотим экспортировать

Изображение выглядит как текст, снимок экрана, программное обеспечение, число

Автоматически созданное описание

Рис.14 соединение с базой данных

Создаем базу данных в Phpmyadmin

**CREATE** **TABLE** zadanie\_42\_13(

id **INT** ***AUTO\_INCREMENT*** **PRIMARY** **KEY**,

total\_sales\_by\_region **int**,

region **varchar**(7));

проверяем полученные значения в целевой базе данных

Изображение выглядит как текст, снимок экрана, Шрифт, программное обеспечение

Автоматически созданное описание

Рис.15 проверка полученных значений

Заключение

В ходе проделанной лабораторной работы, была создана etl-трансформация, исходные данные, после трансформации были отправлены в целевую базу данных.