МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ

УЧРЕЖДЕНИЕ ОБРАЗОВАНИЯ

БелорусскиЙ государственный университет

информатики и радиоэлектроники

Инженерно-экономический факультет

Кафедра экономической информатики

Дисциплина: Системы и технологии интеллектуальной обработки данных

|  |
| --- |
|  |
|  |
|  |

Лабораторная работа №1

# «Консолидация данных и ETL»

|  |  |  |
| --- | --- | --- |
| Выполнил: |  | Д. Г. Слобин  группа 772303 |
| Проверил: |  | А. А. Кунцевич |

Минск 2020

**Теоретическая часть:**

1. Что такое csv файл?

Это текстовый файл, в котором содержится информация.

Каждая строка - это отдельная строка таблицы, а столбцы отделены один от другого специальными символами - разделителями (например, запятой).

В последнее время разделителем может быть не только запятая, но и другие символы (пробел, точка с запятой, табуляция, другое).

1. XSLX

XLSX - это формат электронных таблиц, который создает Microsoft Office Excel начиная с версии 2007.

**Практическая часть:**

1. Создать CSV файл. Используя Talend, написать по генерации из этого файла другого CSV файла.
2. Создать Excel файл. Используя Talend, написать сценарий по генерации из этого файла другого Excel файла.
3. Создать Excel файл. Используя элемент tMap Talend(а), написать сценарий по генерации другого Excel файла с заданной структурой полей. Например, некоторые поля одного файла объединены в одно поле другого файла. Таких объединений должно быть 2.

На рисунках изображено выполнение лабораторной работы №1.

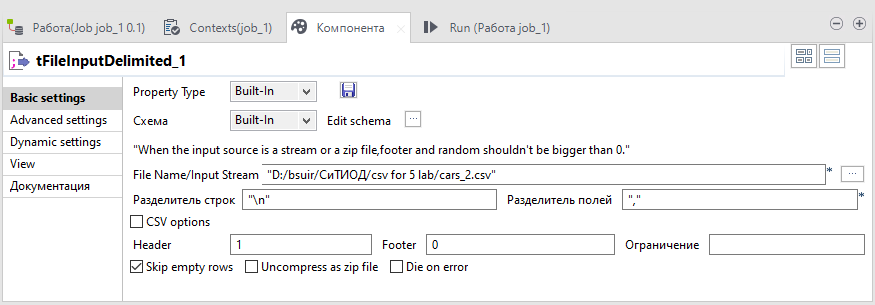


Рисунок 1 – Настройки компонента tFileInputDelimited\_1

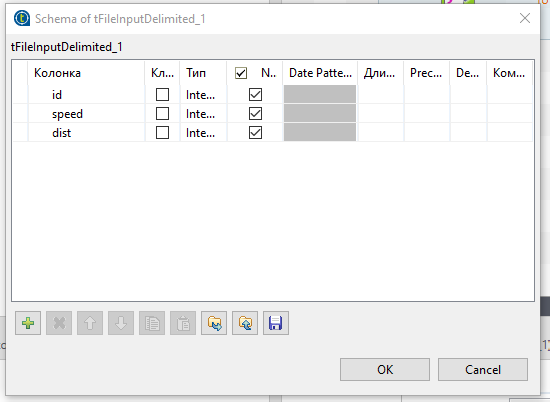


Рисунок 2 – Содержание компонента tFileInputDelimited\_1

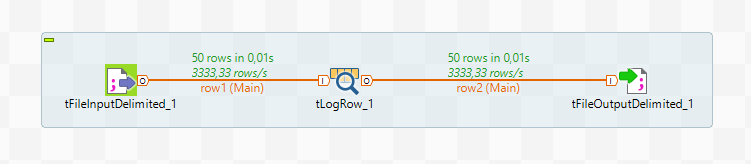


Рисунок 3 – Схема элементов CSV файла

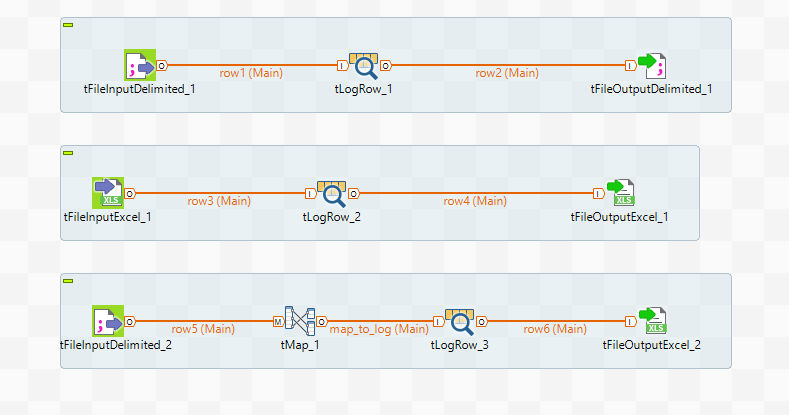


Рисунок 4 – Схема элементов всех файлов

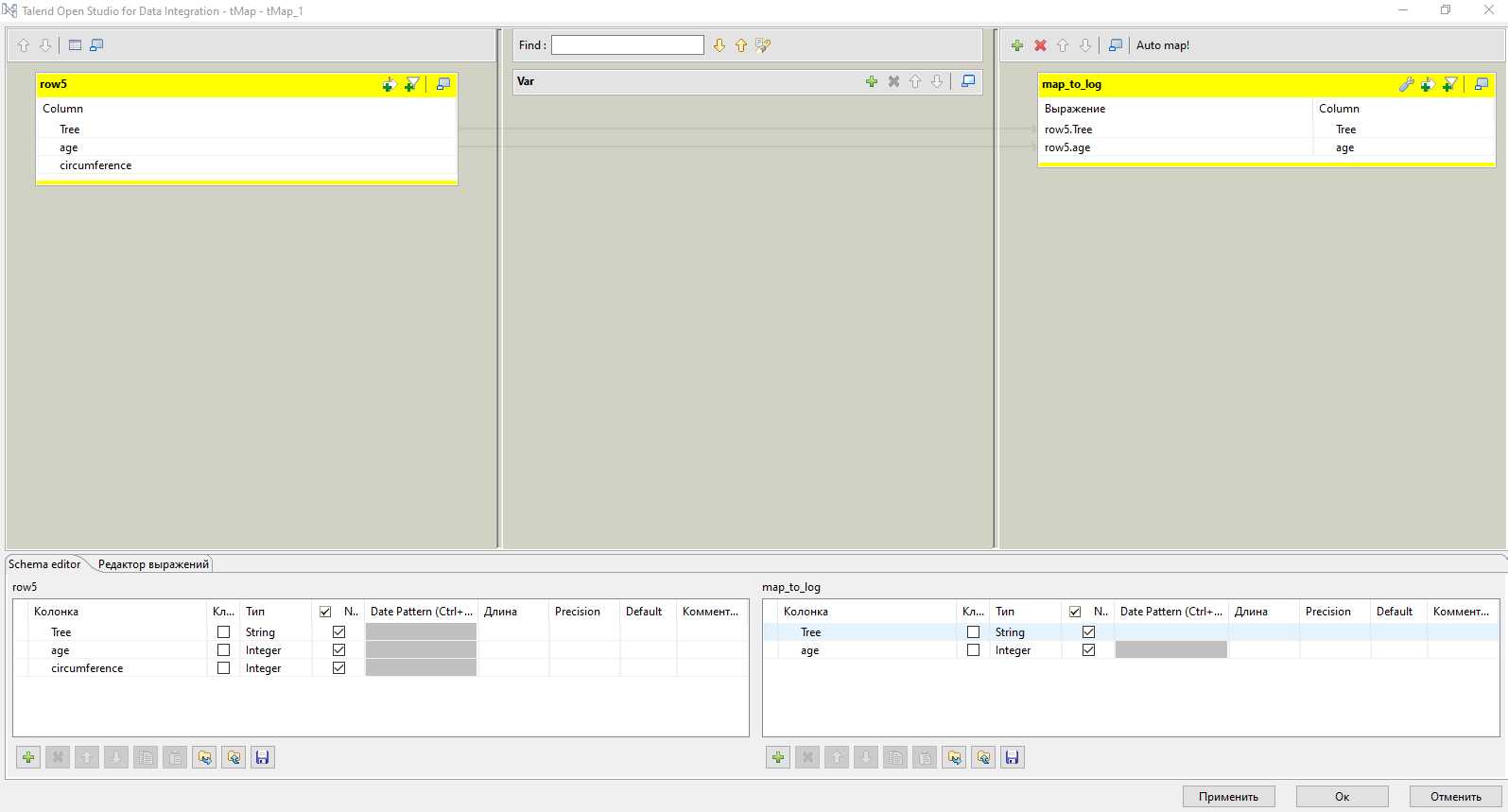


Рисунок 5 – Свойства элемента tMap

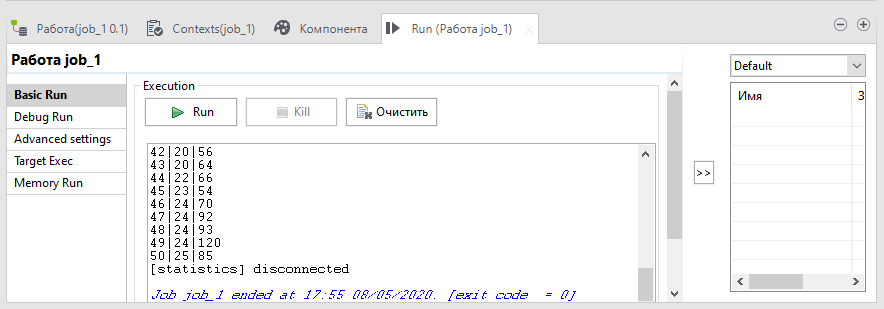


Рисунок 6 – Результат выполнения работы

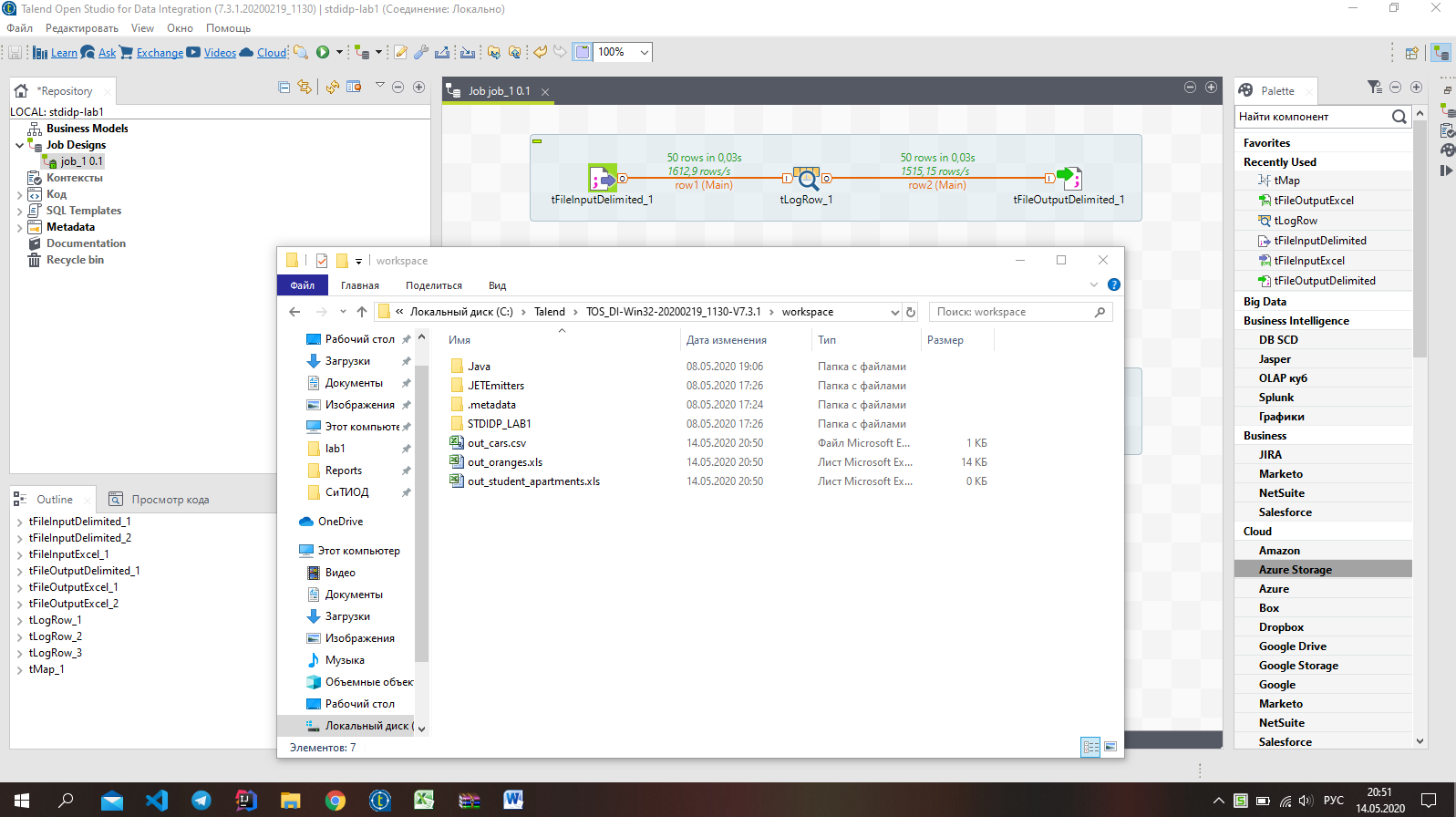


Рисунок 7 – Результат выполнения работы