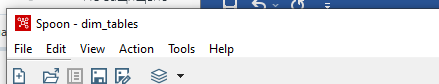
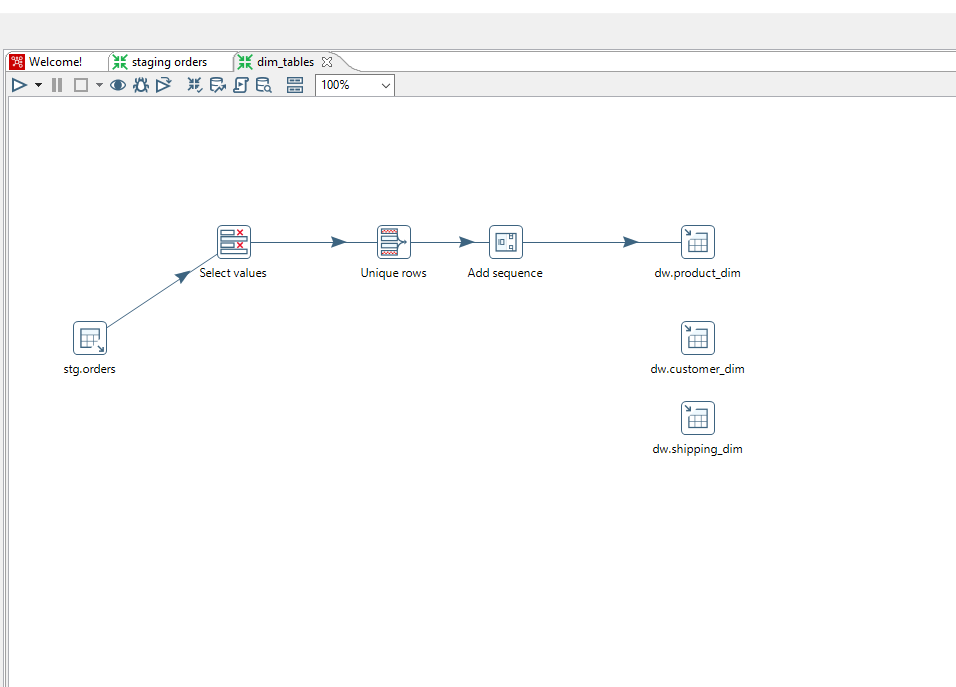
Кайгородцев Михаил БД-231м

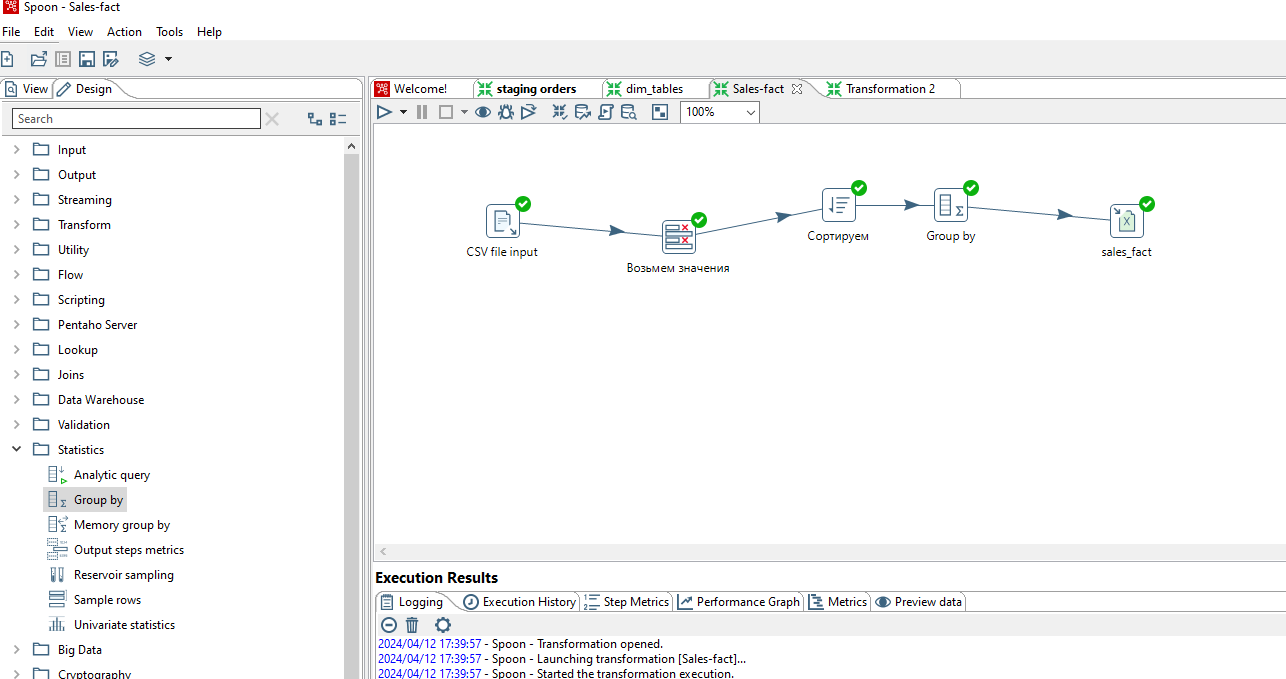
Скачать [отсюда](https://sourceforge.net/projects/pentaho/) и запустить Pentaho DI. Pentaho DI требует установку Java 8. Попробуйте скачать архив и распаковать его. Необходимо запустить spoon.sh для Linux/Mac и spoon.bat для Windows. Видео по установке Pentaho DI на примере Windows 10



[Скачать примеры Pentaho jobs](https://github.com/BosenkoTM/Data-Engineering-Platforms/tree/master/modules/Module04/Lab%204.4#%D0%BF%D0%B5%D1%80%D0%B5%D0%B4-%D0%BF%D0%B5%D1%80%D0%B2%D1%8B%D0%BC-%D0%B7%D0%B0%D0%BF%D1%83%D1%81%D0%BA%D0%BE%D0%BC) для Staging и Dimension Tables.

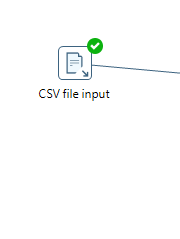


Создайте еще одну трансформацию, в которой создать sales\_fact таблицу.



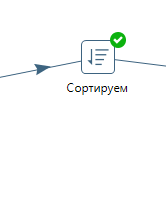
Выявить 8-10 подсистем в ETL Pentaho DI и написать небольшой отчет, в котором приложить print screen компонента (ETL подсистемы) и написать про его свойства. Результат сохраните в Git.

1. CSV file input



Выгружает данные из файлов формата csv

1. Sort rows



Помогает отсортировать данные по выбранной колонке.

1. Калькулятор



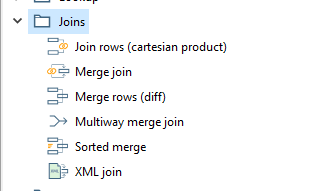
С помощью него можно получить новые данные, используя исходные данные и различные выражения.

1. Logs



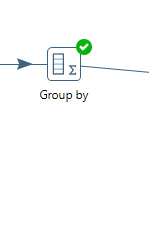
Получение логов о проделанной работе

1. Joins



Дает возможность соединить таблицы.

1. Group by



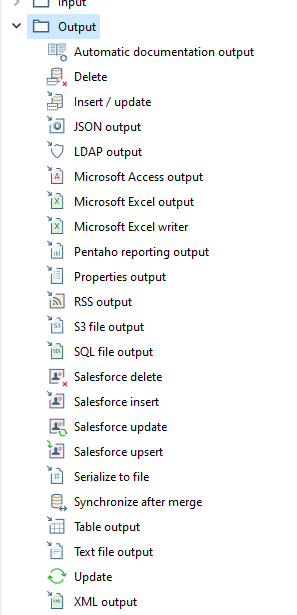
Помогает сгруппировать таблицы

1. Split fields to rows



Переводит колонки данных в строки

1. Output



Группа компонентов, позволяющая обработанные данные, вывести различными способами