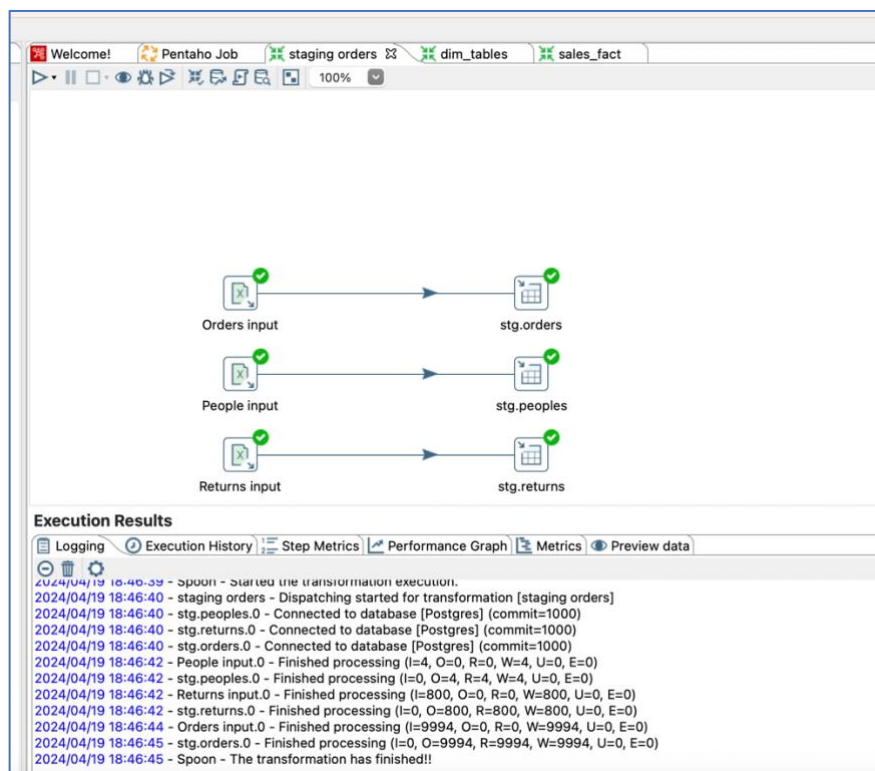


Отчёт по  
ПРАКТИЧЕСКОЙ РАБОТЕ № 4-1. ETL и ELT  
студента группы БД-231м  
Осокина Артёма Ибрагимовича

---

1. Скачать [отсюда](#) 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
2. [Скачать примеры Pentaho jobs](#) для **Staging** и **Dimension Tables**.
3. Создайте еще одну трансформацию, в которой создать **sales\_fact** таблицу.



Welcome! Pentaho Job staging orders dim\_tables sales\_fact

100%

stg.orders

Select values

Unique rows

Add sequence

dw.product\_dim

dw.customer\_dim

dw.shipping\_dim

### Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Preview data

2024/04/19 18:46:20 - Spoon - Started the transformation execution.

2024/04/19 18:46:27 - dim\_tables - Dispatching started for transformation [dim\_tables]

2024/04/19 18:46:27 - dw.customer\_dim.0 - Connected to database [Postgres] (commit=1000)

2024/04/19 18:46:27 - dw.product\_dim.0 - Connected to database [Postgres] (commit=1000)

2024/04/19 18:46:27 - dw.shipping\_dim.0 - Connected to database [Postgres] (commit=1000)

2024/04/19 18:46:27 - stg.orders.0 - Finished reading query, closing connection

2024/04/19 18:46:27 - Select values.0 - Finished processing (I=0, O=0, R=9994, W=9994, U=0, E=0)

2024/04/19 18:46:27 - stg.orders.0 - Finished processing (I=9994, O=0, R=0, W=9994, U=0, E=0)

2024/04/19 18:46:27 - Unique rows.0 - Finished processing (I=0, O=0, R=9994, W=9984, U=0, E=0)

2024/04/19 18:46:27 - Add sequence.0 - Finished processing (I=0, O=0, R=9984, W=9984, U=0, E=0)

2024/04/19 18:46:28 - dw.product\_dim.0 - Finished processing (I=0, O=9984, R=9984, W=9984, U=0, E=0)

2024/04/19 18:46:28 - Spoon - The transformation has finished!!

Welcome! Pentaho Job staging orders dim\_tables sales\_fact

100%

Start

Staging Tables

Dim Tables

Sales Fact

Success

Write to log 2

Write to log

Write to log 3

### Execution Results

Logging History Job metrics Metrics

2024/04/19 18:47:05 - merge join.0 - Finished processing (I=0, O=0, R=9994, W=9994, U=0, E=0)

2024/04/19 18:47:05 - Sort order\_id 12.0 - Finished processing (I=0, O=0, R=9994, W=9994, U=0, E=0)

2024/04/19 18:47:06 - Merge join 2.0 - Finished processing (I=0, O=0, R=10794, W=3226, U=0, E=0)

2024/04/19 18:47:06 - sales\_fact Select values.0 - Finished processing (I=0, O=0, R=3226, W=3226, U=0, E=0)

2024/04/19 18:47:06 - file output csv.0 - Finished processing (I=0, O=3227, R=3226, W=3226, U=0, E=0)

2024/04/19 18:47:06 - Pentaho Job - Starting entry [Success]

2024/04/19 18:47:06 - Pentaho Job - Finished job entry [Success] (result=[true])

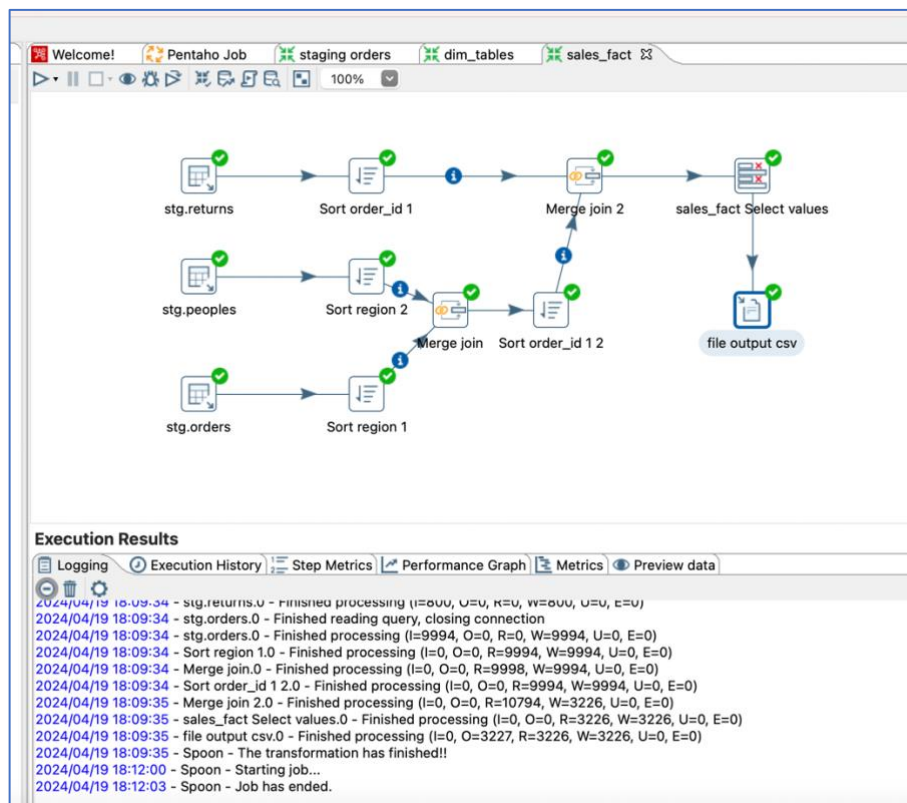
2024/04/19 18:47:06 - Pentaho Job - Finished job entry [Sales Fact] (result=[true])

2024/04/19 18:47:06 - Pentaho Job - Finished job entry [Dim Tables] (result=[true])







2024/04/19 18:47:06 - Pentaho Job - Finished job entry [Staging Tables] (result=[true])

2024/04/19 18:47:06 - Pentaho Job - Job execution finished

2024/04/19 18:47:06 - Spoon - Job has ended.



- 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.

Название	Описание
 <b>Database lookup</b>	Этот компонент используется для выполнения операции поиска в базе данных на основе определенных критериев. Он позволяет связывать данные из различных источников по ключевым полям.
 <b>XML output</b>	Компонент для чтения данных из XML-файлов. Он позволяет извлекать структурированные данные из XML-файлов и передавать их дальше для обработки.
 <b>Text file output</b>	Этот компонент используется для записи данных в текстовые файлы. Он обеспечивает возможность форматирования вывода и выбора разделителей для различных типов данных.
 <b>Sort rows</b>	Этот компонент позволяет сортировать строки данных по одному или нескольким полям. Он полезен при подготовке данных перед загрузкой в целевые системы.
 <b>Calculator</b>	Компонент, который выполняет различные арифметические и логические операции над данными. Он позволяет создавать новые вычисляемые поля на основе существующих данных.
 <b>CSV file input</b>	Этот компонент читает данные из файлов CSV (Comma-Separated Values). Он обрабатывает строки данных, разделенные запятыми, и передает их для дальнейшей обработки.

 <p>Select values</p>	<p>Компонент для выбора определенных полей из набора данных. Он позволяет фильтровать и выбирать только необходимые данные для дальнейшей обработки.</p>
 <p>Row normaliser</p>	<p>Этот компонент используется для преобразования строк данных в структурированный формат. Он преобразует строки с данными различных форматов в унифицированный формат.</p>
 <p>Microsoft Excel input</p>	<p>Компонент для чтения данных из файлов Microsoft Excel. Он позволяет извлекать данные из листов Excel и передавать их для дальнейшей обработки.</p>
 <p>Merge join</p>	<p>Этот компонент выполняет объединение данных из двух источников на основе определенного условия соединения. Он позволяет объединять данные из различных источников для дальнейшей обработки.</p>