

Государственное автономное образовательное  
учреждение высшего образования «Московский городской педагогический  
университет»

Институт цифрового образования  
Департамент информатики, управления и технологий

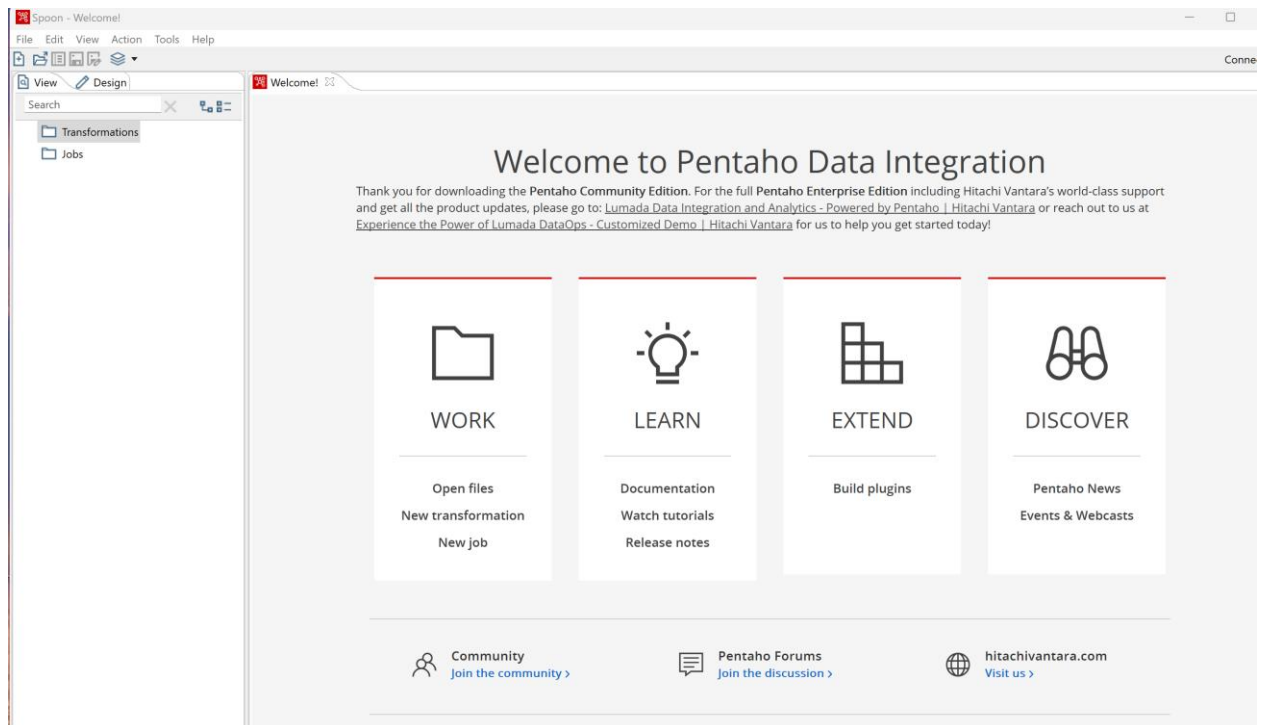
Отчет по лабораторной работе № 4.1

Выполнил: студент БД-231  
Кряжева К.С.

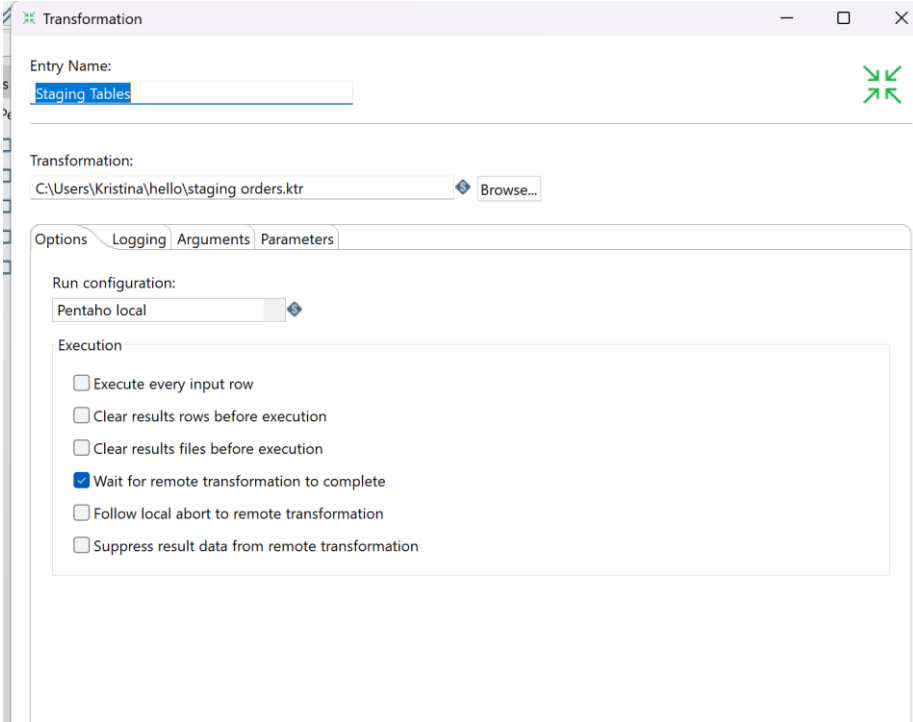
2024 г

## 1. Установка Pentaho

ИМЯ	ТИП
kitchen.sh	Исходный файл SH
LICENSE.txt	Текстовый документ
Pan.bat	Пакетный файл Windows
pan.sh	Исходный файл SH
PentahoDataIntegration_OSS_Licenses.ht...	Yandex Browser HTML Docu
purge-utility.bat	Пакетный файл Windows
purge-utility.sh	Исходный файл SH
README.txt	Текстовый документ
runSamples.bat	Пакетный файл Windows
runSamples.sh	Исходный файл SH
set-pentaho-env.bat	Пакетный файл Windows
set-pentaho-env.sh	Исходный файл SH
Spoon.bat	Пакетный файл Windows
spoon.command	Файл "COMMAND"
spoon.ico	Файл "ICO"
spoon.png	Файл "PNG"
spoon.sh	Исходный файл SH
SpoonConsole.bat	Пакетный файл Windows
SpoonDebug.bat	Пакетный файл Windows
SpoonDebug.sh	Исходный файл SH
yarn.sh	Исходный файл SH



## 2. Примеры Staging и Dimensions Table



Transformation

Entry Name:

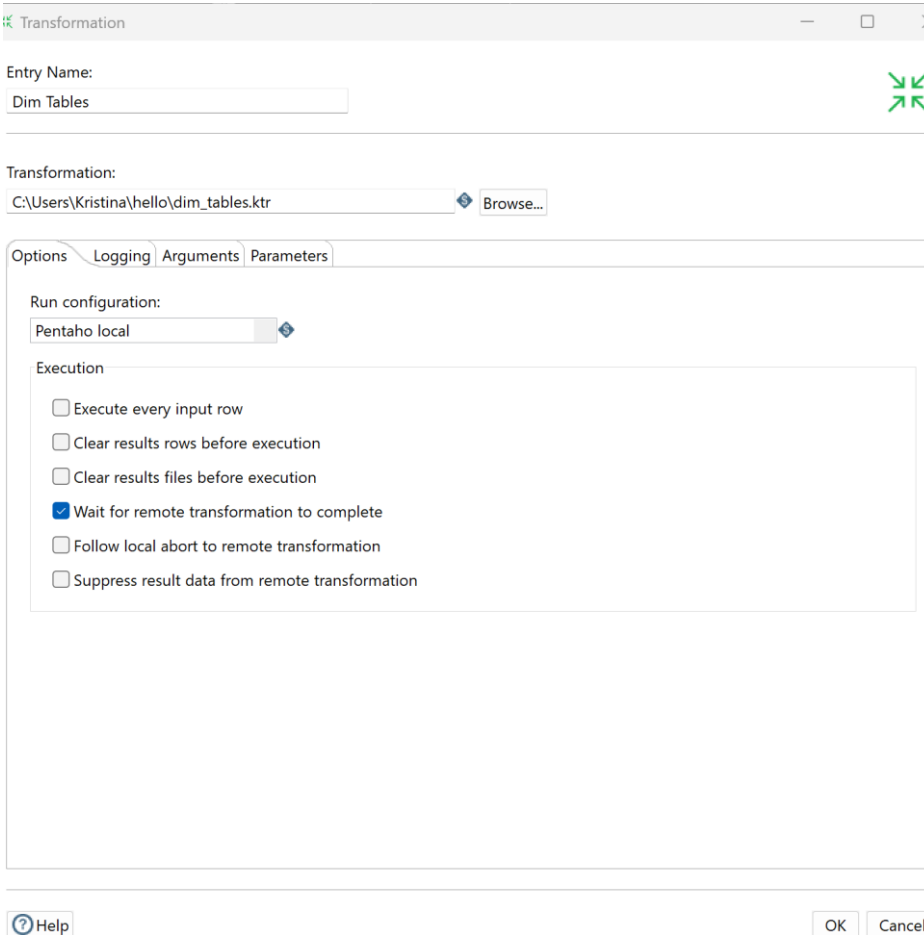
Transformation:  [Browse...](#)

Options | Logging | Arguments | Parameters

Run configuration:

Execution

- ☐ Execute every input row
- ☐ Clear results rows before execution
- ☐ Clear results files before execution
- ☒ Wait for remote transformation to complete
- ☐ Follow local abort to remote transformation
- ☐ Suppress result data from remote transformation



Transformation

Entry Name:

Transformation:  [Browse...](#)

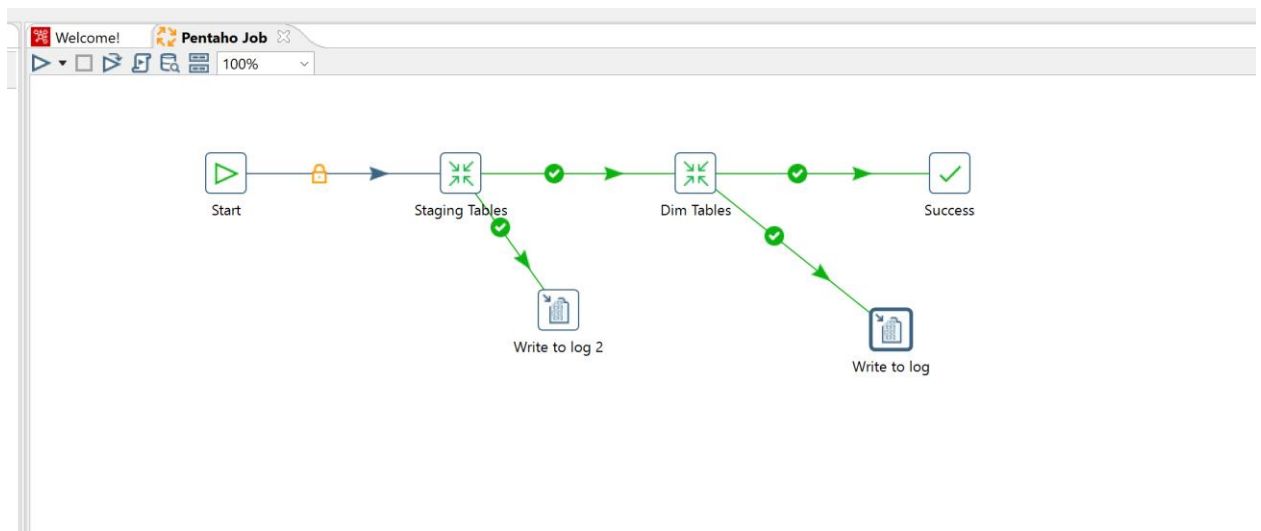
Options | Logging | Arguments | Parameters

Run configuration:

Execution

- ☐ Execute every input row
- ☐ Clear results rows before execution
- ☐ Clear results files before execution
- ☒ Wait for remote transformation to complete
- ☐ Follow local abort to remote transformation
- ☐ Suppress result data from remote transformation

[? Help](#) [OK](#) [Cancel](#)



### 3. 8-10 подсистем в ETL Pentaho DI

- 1) Table input: этот компонент позволяет считывать данные из таблицы с возможностью определения SQL-запроса.



Table input

- 2) Text file input: этот компонент считывает данные из текстового файла.



Text file input

- 3) Filter rows: компонент используется для фильтрации строк по условию.



Filter rows

Filter rows

Step name:

Send 'true' data to step:

Send 'false' data to step:

The condition:

- 4) Sort rows: компонент сортирует строки по определенным столбцам.



Sort rows

Sort rows

Step name:

Sort directory:

TMP-file prefix:

Sort size (rows in memory):

Free memory threshold (in %):

Compress TMP Files? ☒

Only pass unique rows? (verifies keys only) ☐

Fields :

#	Fieldname	Ascending	Case sensitive compare?	Sort based on current locale?	Collator Strength	Presorted?
1						

- 5) Merge Join: компонент используется для объединения двух наборов данных на основе равенства столбцов.



Merge join

Merge join

Step name:

First Step:

Second Step:

Join Type:

Keys for 1st step:

#	Key field
1	

Keys for 2nd step:

#	Key field
1	

- 6) Memory Group By: компонент позволяет группировать данные в памяти. Можно указать столбцы для группировки и агрегатные функции для каждой группы.



Memory group by

Memory group by

Step name: Memory group by

Always give back a result row: ☐

The fields that make up the group:

#	Group field
1	

Get Fields

Aggregates :

#	Name	Subject	Type	Value
1				

Get lookup fields

- 7) Detect Empty Stream: компонент используется для выявления пустых потоков данных.



Detect empty stream

Det...

Step name: Detect empty stream

Help OK Cancel

- 8) Unique Rows: удаляет дубликаты строк из набора данных.



Unique rows

Unique rows

Step name: Unique rows

Settings

Add counter to output? ☐ Counter field

Redirect duplicate row ☐ Error description

Fields to compare on (no entries means: compare complete row)

#	Fieldname	Ignore case
1		

- 9) Add Constants: добавляет константные значения в результирующий набор данных.



Add constants

Add constants

Step name: Add constants

Fields :

#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Value	Set empty string?
1										

Help OK Cancel