

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

Башкатова Анна Денисовна

ОТЧЕТ

по дисциплине «Инструменты для хранения и обработки больших данных»

**Тема: «Домашняя работа 2 “ETL Компоненты и начало работы с ETL на
примере Pentaho Data Integration” Часть 2»**

Направление подготовки (специальность) 38.03.05 – бизнес-информатика

Направленность (профиль) образовательной программы «Аналитика данных
и эффективное управление»

Курс обучения: 3

Форма обучения: очная

Руководитель: Босенко Т. М.

Москва
2023

Создание подключения к БД.

Database Connection

General
Advanced
Options
Pooling
Clustering

Connection name:
MySQL|Башкатова

Connection type:
MySQL
Native Mondrian
Neoview
Netezza
Oracle
Oracle RDB
Palo MOLAP Server
Pentaho Data Services
PostgreSQL
Redshift
Remedy Action Request System
SAP ERP System

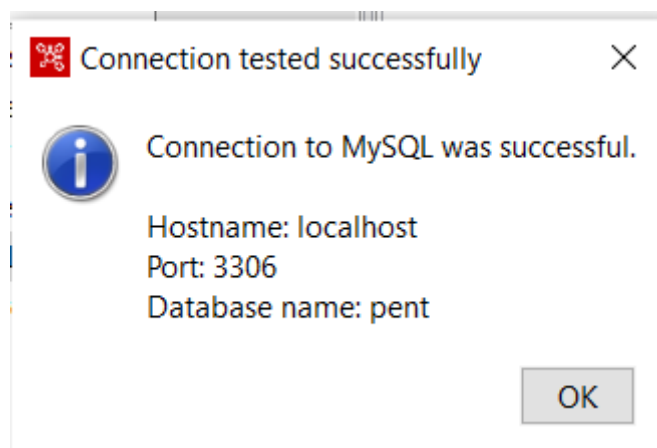
Access:
Native (JDBC)
ODBC
JNDI

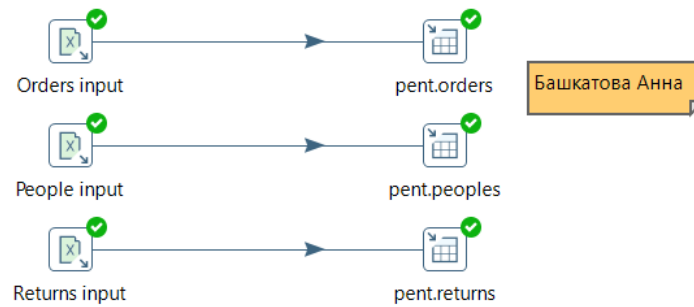
Settings
Host Name:
localhost
Database Name:
pent
Port Number:
3306
Username:
root
Password:

☒ Use Result Streaming Cursor

Test Feature List Explore OK Cancel

Подключение успешно.





Execution Results

[Logging](#)
[Execution History](#)
[Step Metrics](#)
[Performance Graph](#)
[Metrics](#)
[Preview data](#)

2023/05/04 21:48:26 - Spoon - Running transformation using the Kettle execution engine

2023/05/04 21:48:26 - Spoon - Transformation opened.

2023/05/04 21:48:26 - Spoon - Launching transformation [staging orders]...

2023/05/04 21:48:26 - Spoon - Started the transformation execution.

2023/05/04 21:48:26 - staging orders - Dispatching started for transformation [staging orders]

2023/05/04 21:48:26 - pent.peoples.0 - Connected to database [MySQL] (commit=1000)

2023/05/04 21:48:26 - pent.returns.0 - Connected to database [MySQL] (commit=1000)

2023/05/04 21:48:26 - pent.orders.0 - Connected to database [MySQL] (commit=1000)

2023/05/04 21:48:27 - Returns input.0 - Finished processing (I=800, O=0, R=0, W=800, U=0, E=0)

2023/05/04 21:48:27 - People input.0 - Finished processing (I=4, O=0, R=0, W=4, U=0, E=0)

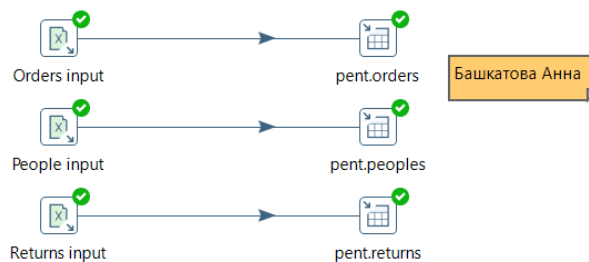
2023/05/04 21:48:27 - pent.peoples.0 - Finished processing (I=0, O=4, R=4, W=4, U=0, E=0)

2023/05/04 21:48:27 - Orders input.0 - Finished processing (I=9994, O=0, R=0, W=9994, U=0, E=0)

2023/05/04 21:48:28 - pent.returns.0 - Finished processing (I=0, O=800, R=800, W=800, U=0, E=0)

2023/05/04 21:48:29 - pent.orders.0 - Finished processing (I=0, O=9994, R=9994, W=9994, U=0, E=0)

2023/05/04 21:48:29 - Spoon - The transformation has finished!!



Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Preview data

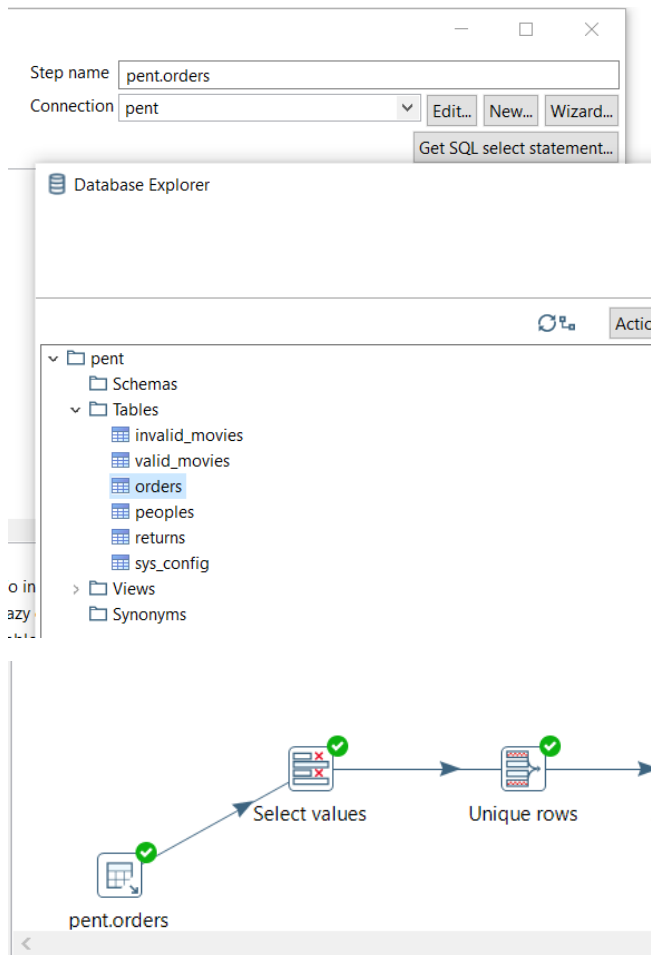
#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	input/output
1	Orders input	0	0	9994	9994	0	0	0	0	Finished	1.7s	5 750	-
2	Returns input	0	0	800	800	0	0	0	0	Finished	0.8s	1 000	-
3	People input	0	0	4	4	0	0	0	0	Finished	1.2s	3	-
4	pent.orders	0	9994	9994	0	9994	0	0	0	Finished	3.0s	3 278	-
5	pent.peoples	0	4	4	0	4	0	0	0	Finished	1.4s	3	-
6	pent.returns	0	800	800	0	800	0	0	0	Finished	1.8s	439	-

Вид в БД

Фильтры

Содержит слово:

Таблица	Действие	Строки	Тип	Сравнение	Размер	Фрагментировано
orders		9 994	InnoDB	utf8_general_ci	3.5 МБ	-
peoples		4	InnoDB	utf8_general_ci	16.0 КиБ	-
returns		800	InnoDB	utf8_general_ci	64.0 КиБ	-
3 таблицы	Всего	10 798	InnoDB	utf8_general_ci	3.6 МБ	0 Байт



Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Preview data

2023/05/04 22:37:24 - Spoon - Running transformation using the Kettle execution engine

2023/05/04 22:37:25 - Spoon - Transformation opened.

2023/05/04 22:37:25 - Spoon - Launching transformation [dim_tables]...

2023/05/04 22:37:25 - Spoon - Started the transformation execution.

2023/05/04 22:37:25 - dim_tables - Dispatching started for transformation [dim_tables]

2023/05/04 22:37:25 - pent.product_pent.0 - Connected to database [MySQL] (commit=1000)

2023/05/04 22:37:25 - pent.orders.0 - Finished reading query, closing connection

2023/05/04 22:37:25 - pent.orders.0 - Finished processing (I=9994, O=0, R=0, W=9994, U=0, E=0)

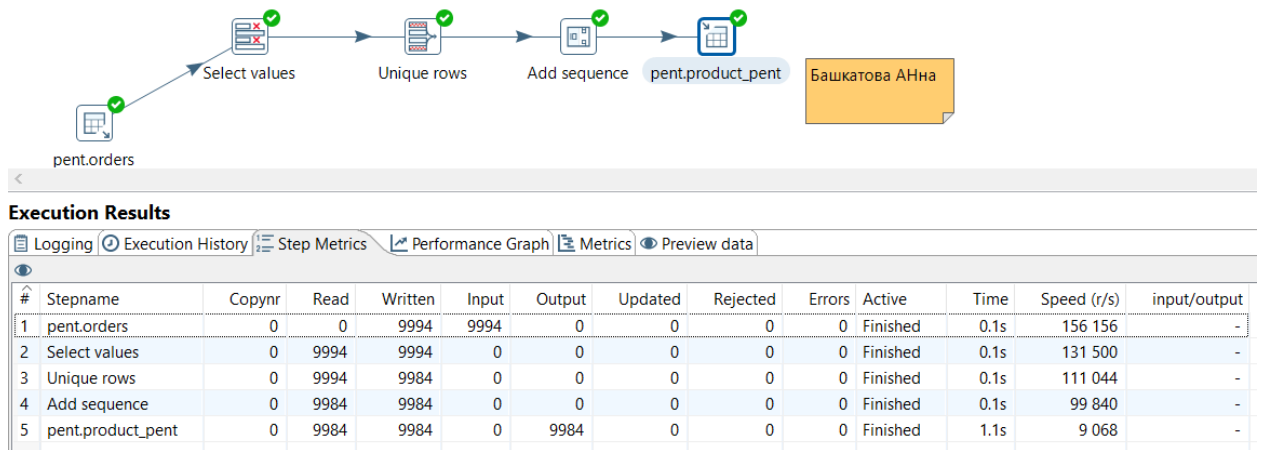
2023/05/04 22:37:25 - Select values.0 - Finished processing (I=0, O=0, R=9994, W=9994, U=0, E=0)

2023/05/04 22:37:25 - Unique rows.0 - Finished processing (I=0, O=0, R=9994, W=9984, U=0, E=0)

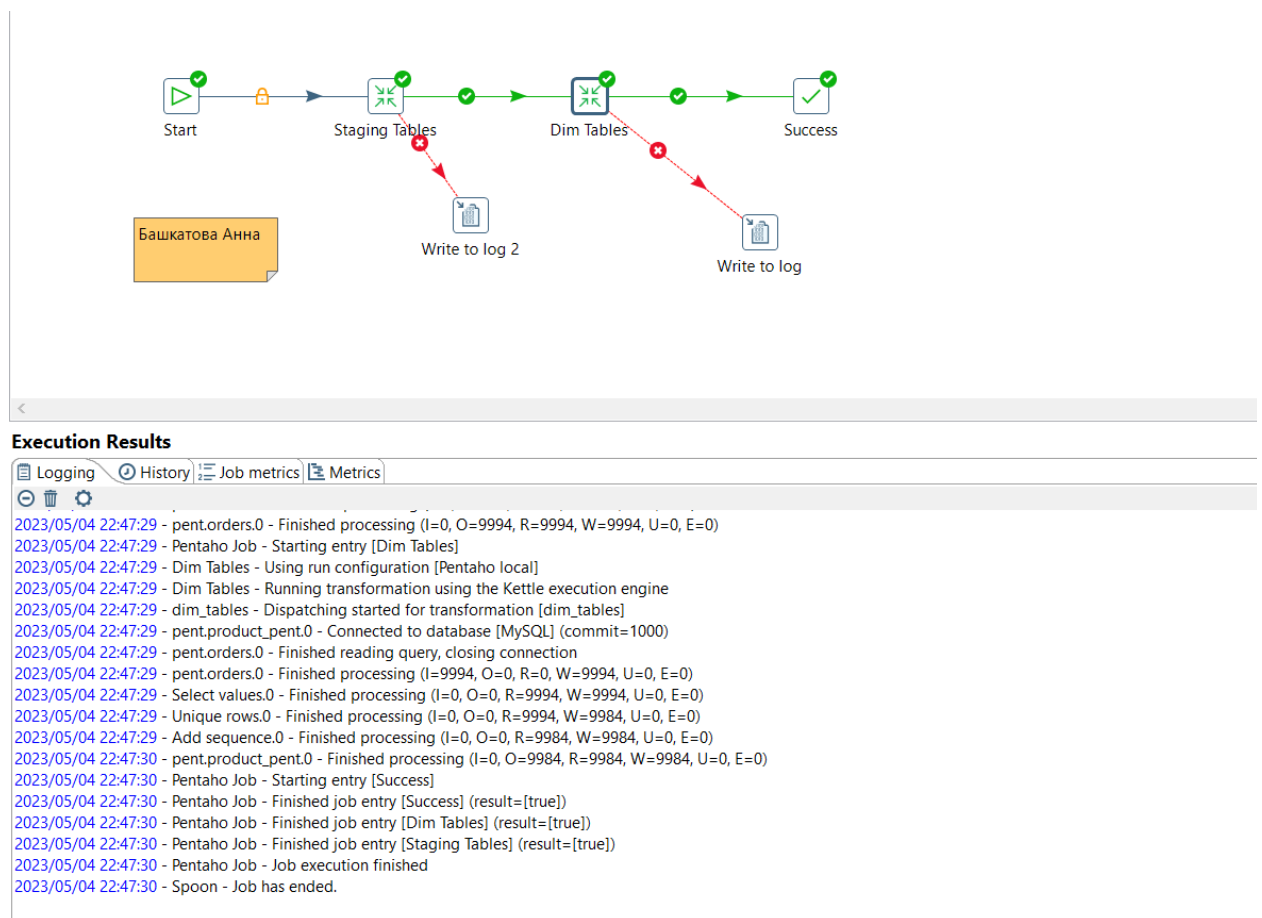
2023/05/04 22:37:25 - Add sequence.0 - Finished processing (I=0, O=0, R=9984, W=9984, U=0, E=0)

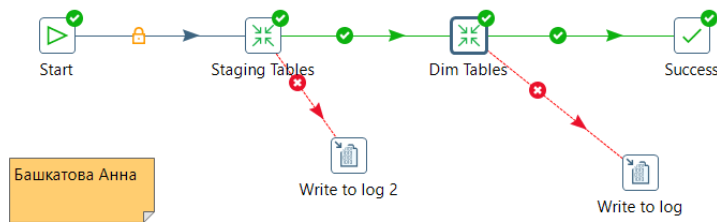
2023/05/04 22:37:26 - pent.product_pent.0 - Finished processing (I=0, O=9984, R=9984, W=9984, U=0, E=0)

2023/05/04 22:37:26 - Spoon - The transformation has finished!!



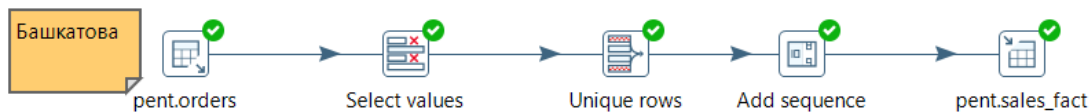
Работа jobs





Execution Results

Job / Job Entry	Comment	Result	Reason	Filename	Nr	Log date
Job: Pentaho Job	Start of job execution		start			2023/05/04 22:...
Start	Start of job execution		start			2023/05/04 22:...
Start	Job execution finished	Success			0	2023/05/04 22:...
Staging Tables	Start of job execution		Followed unconditional link	C:\Users\Bashk\Downloads...		2023/05/04 22:...
Staging Tables	Job execution finished	Success		C:\Users\Bashk\Downloads...	1	2023/05/04 22:...
Dim Tables	Start of job execution		Followed link after success	C:\Users\Bashk\Downloads...		2023/05/04 22:...
Dim Tables	Job execution finished	Success		C:\Users\Bashk\Downloads...	2	2023/05/04 22:...
Success	Start of job execution		Followed link after success			2023/05/04 22:...
Success	Job execution finished	Success			2	2023/05/04 22:...
Job: Pentaho Job	Job execution finished	Success	finished		2	2023/05/04 22:...



Execution Results

Logging	Execution History	Step Metrics	Performance Graph	Metrics	Preview data
2023/05/04 23:00:56	Spoon - Running transformation using the Kettle execution engine				
2023/05/04 23:00:56	Spoon - Transformation opened.				
2023/05/04 23:00:56	Spoon - Launching transformation [sales_fact]...				
2023/05/04 23:00:56	Spoon - Started the transformation execution.				
2023/05/04 23:00:56	sales_fact - Dispatching started for transformation [sales_fact]				
2023/05/04 23:00:56	pent.sales_fact.0 - Connected to database [MySQL] (commit=1000)				
2023/05/04 23:00:56	pent.orders.0 - Finished reading query, closing connection				
2023/05/04 23:00:56	pent.orders.0 - Finished processing (I=9994, O=0, R=0, W=9994, U=0, E=0)				
2023/05/04 23:00:56	Select values.0 - Finished processing (I=0, O=0, R=9994, W=9994, U=0, E=0)				
2023/05/04 23:00:56	Unique rows.0 - Finished processing (I=0, O=0, R=9994, W=9991, U=0, E=0)				
2023/05/04 23:00:56	Add sequence.0 - Finished processing (I=0, O=0, R=9991, W=9991, U=0, E=0)				
2023/05/04 23:00:57	pent.sales_fact.0 - Finished processing (I=0, O=9991, R=9991, W=9991, U=0, E=0)				
2023/05/04 23:00:57	Spoon - The transformation has finished!!				

Создание трансформации, в которой создаётся sales_fact таблица:

Башкатова

pent.orders

Select values

Unique rows

Add sequence

pent.sales_fact

Execution Results

LoggingExecution HistoryStep MetricsPerformance GraphMetricsPreview data

#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	input/output
1	pent.orders	0	0	9994	9994	0	0	0	0	Finished	0.1s	124 925	-
2	Select values	0	9994	9994	0	0	0	0	0	Finished	0.1s	109 824	-
3	Unique rows	0	9994	9991	0	0	0	0	0	Finished	0.1s	81 918	-
4	Add sequence	0	9991	9991	0	0	0	0	0	Finished	0.2s	65 301	-
5	pent.sales_fact	0	9991	9991	0	9991	0	0	0	Finished	1.1s	9 225	-