

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

ДИСЦИПЛИНА:

Инструменты для хранения и обработки больших данных

Домашнее задание 2

Тема:

Pentaho.

Выполнила: Соколова М. С., группа: АДЭУ-201

Преподаватель: Босенко Т. М.

Москва

2023

Database Connection

General

Advanced

Options

Pooling

Clustering

Connection name:

MySQL

Connection type:

MS Access

MS SQL Server

MS SQL Server (Native)

MariaDB

MaxDB (SAP DB)

MonetDB

MySQL

Native Mondrian

Neoview

Netezza

Oracle

Oracle RDB

Access:

Native (JDBC)

ODBC

JNDI

Settings

Host Name:

localhost

Database Name:

Sokolova_pentaho

Port Number:

3306

Username:

root

Password:

☒ Use Result Streaming Cursor

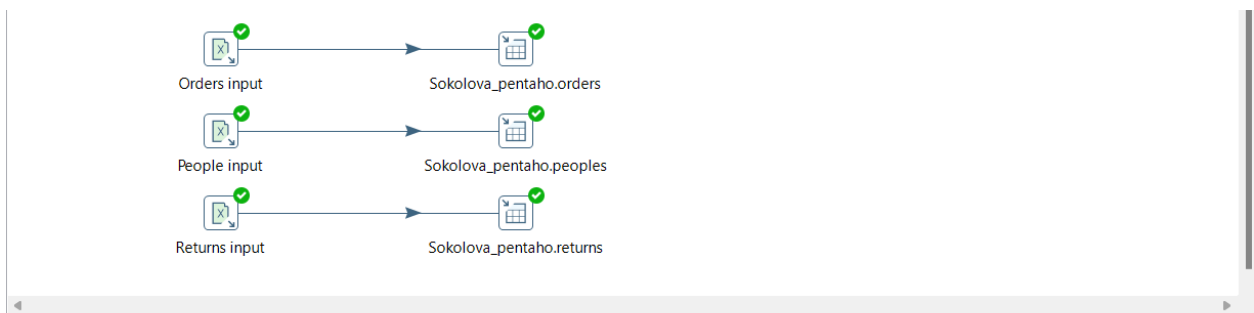
Test

Feature List

Explore

OK

Cancel



Execution Results

<div> <div>Logging</div> <div>Execution History</div> <div>Step Metrics</div> <div>Performance Graph</div> <div>Metrics</div> <div>Preview data</div> </div>													
#	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	2.0s	5 042	-
2	Returns input	0	0	800	800	0	0	0	0	Finished	1.3s	622	-
3	People input	0	0	4	4	0	0	0	0	Finished	1.3s	3	-
4	Sokolova_pentaho.orders	0	9994	9994	0	9994	0	0	0	Finished	5.2s	1 908	-
5	Sokolova_pentaho.peoples	0	4	4	0	4	0	0	0	Finished	1.5s	2	-
6	Sokolova_pentaho.returns	0	800	800	0	800	0	0	0	Finished	2.2s	372	-

Имя схемы:

Sokolova_pentaho

SQL путь:

Кодировка по умолчанию:

utf8mb4

Размер:

0

Представление по умолчанию:

utf8mb4_general_ci

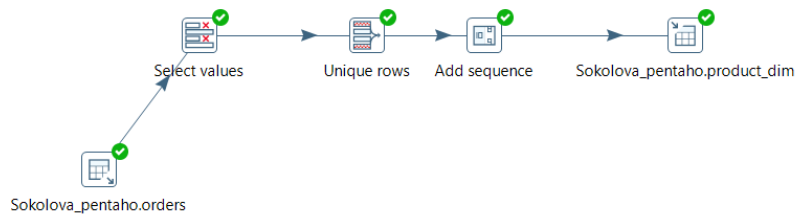
Таблицы

Представления

Индексы

Процедуры

Название	Движок	Авто-увеличение	Размер	Описание
orders	InnoDB	0	3,5M	
peoples	InnoDB	0	16K	
returns	InnoDB	0	64K	



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/c

1

Sokolova_pentaho.orders

0

0

9994

9994

0

0

0

0

Finished

0.3s

34 822

2

Select values

0

9994

9994

0

0

0

0

0

Finished

0.3s

34 462

3

Unique rows

0

9994

9984

0

0

0

0

0

Finished

0.3s

32 875

4

Add sequence

0

9984

9984

0

0

0

0

0

Finished

0.3s

29 714

5

Sokolova_pentaho.product_dim

0

9984

9984

0

9984

0

0

0

Finished

2.7s

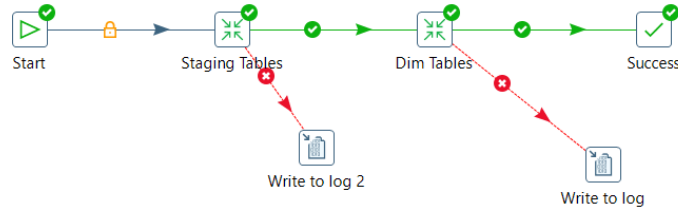
3 760

Имя схемы: SQL путь:

Кодировка по умолчанию: Размер:

Представление по умолчанию:

	Название	Движок	Авто-увеличение	Размер	Описание
Таблицы	orders	InnoDB	0	3,5M	
Представления	peoples	InnoDB	0	16K	
Индексы	product_dim	InnoDB	0	16K	
Процедуры	returns	InnoDB	0	64K	



Execution Results

<div> <div>Logging</div> <div>History</div> <div>Job metrics</div> <div>Metrics</div> </div>						
Job / Job Entry	Comment	Result	Reason	Filename	Nr	Log date
Pentaho Job						
Job: Pentaho Job	Start of job execution		start			2023/05/04 18:...
Start	Start of job execution		start			2023/05/04 18:...
Start	Job execution finished	Success			0	2023/05/04 18:...
Staging Tables	Start of job execution		Followed unconditional link	C:/Users/marin/OneDrive/P...		2023/05/04 18:...
Staging Tables	Job execution finished	Success		C:/Users/marin/OneDrive/P...	1	2023/05/04 18:...
Dim Tables	Start of job execution		Followed link after success	C:/Users/marin/OneDrive/P...		2023/05/04 18:...
Dim Tables	Job execution finished	Success		C:/Users/marin/OneDrive/P...	2	2023/05/04 18:...
Success	Start of job execution		Followed link after success			2023/05/04 18:...
Success	Job execution finished	Success			2	2023/05/04 18:...
Job: Pentaho Job	Job execution finished	Success	finished		2	2023/05/04 18:...

<div> <div>Sokolova_pentaho.orders</div> <div>Select values</div> <div>Group by</div> <div>Sokolova_pentaho.sales_fact</div> </div>													
Execution Results													
<div> <div>Logging</div> <div>Execution History</div> <div>Step Metrics</div> <div>Performance Graph</div> <div>Metrics</div> <div>Preview data</div> </div>													
#	Stepname	Copypnr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	input/output
1	Sokolova_pentaho.orders	0	0	9994	9994	0	0	0	0	0	Finished	0.2s	52 325
2	Select values	0	9994	9994	0	0	0	0	0	0	Finished	0.2s	49 970
3	Group by	0	9994	1	0	0	0	0	0	0	Finished	0.2s	42 893
4	Sokolova_pentaho.sales_fact	0	1	1	0	1	0	0	0	0	Finished	0.3s	4

Group by

Step name

Group by

Include all rows?

Temporary files directory

%%java.io.tmpdir%%

Browse...

TMP-file prefix

grp

Add line number, restart in each group

Line number field name

Always give back a result row

The fields that make up the group:

#

Group field

1

Get Fields

Aggregates :

#

Name

Subject

Type

Value

1

Sales

Sales

Sum

2

Get lookup fields

Table output

Step name Sokolova_pentaho.sales_fact

Connection MySQL Edit... New... Wizard...

Target schema Sokolova_pentaho Browse...

Target table sales_fact Browse...

Commit size 1000

Truncate table ☒

Ignore insert errors ☐

Specify database fields ☒

Main options Database fields

Partition data over tables ☐

Partitioning field

Partition data per month ☒

Partition data per day ☐

Use batch update for inserts ☒

Is the name of the table defined in a field? ☐

Field that contains name of table:

Store the tablename field ☐

Help OK Cancel SQL

Simple SQL editor

SQL statements, separated by semicolon ';' :

```
CREATE TABLE Sokolova_pentaho.sales_fact
(
  Sales DOUBLE
)
;
```

```
SELECT *
FROM Sokolova_pentaho.sales_fact
```

sales_fact 1 ×

SELECT * FROM Sokolova_pentaho.sale Введите SQL выражение чтобы отфильтровать результаты

	123 Sales
1	2 297 201