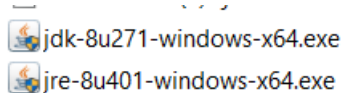


Кузьменко Сергей БД 231-м

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

Выполнено

https://gist.github.com/wavezhang/ba8425f24a968ec9b2a8619d7c2d86a6?permalink_comment_id=3541602 (ссылка для скачивания Jdk)



<https://www.oracle.com/java/technologies/downloads/#java8-windows> (скачивание jre)

<https://www.hitachivantara.com/pentaho/pentaho-plus-platform/data-integration-analytics/pentaho-community-edition.html> (ссылка для скачивания Пентаго)

pdi-ce-9.4.0.0-343.zip

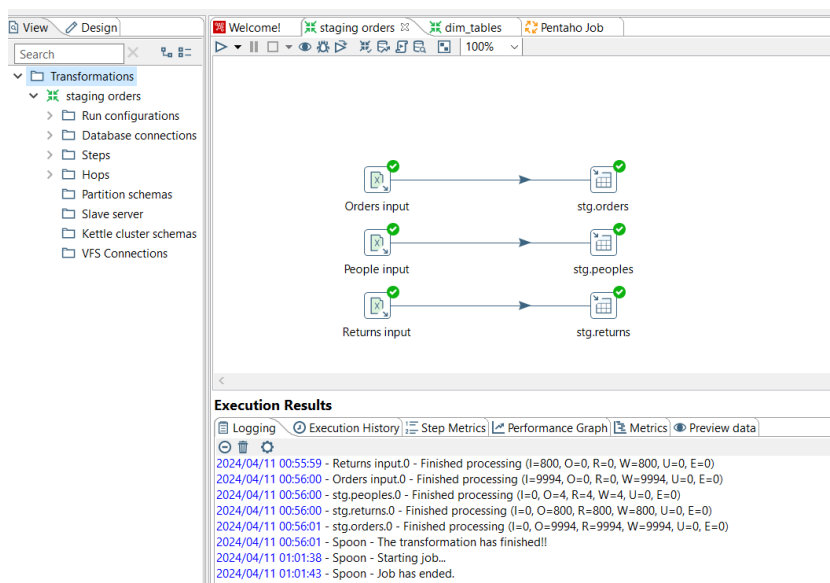
Pentaho Data Integration (Base Install)

Download →

Важно установить переменные как в инструкции

JAVA_HOME	C:\Program Files\Java\jdk1.8.0_271\
NUMBER_OF_PROCESSORS	8
OS	Windows_NT
Path	C:\Program Files (x86)\Common Files\Oracle\Java\javapath;C...
PATHEXT	.COM;.EXE;.BAT;.CMD;.VBS;.VBE;.JS;.JSE;.WSF;.WSH;.MSC
PENTHO_JAVA	C:\Program Files\Java\jre1.8.0_271\bin\java.exe
PENTHO_JAVA_HOME	C:\Program Files\Java\jre1.8.0_271\

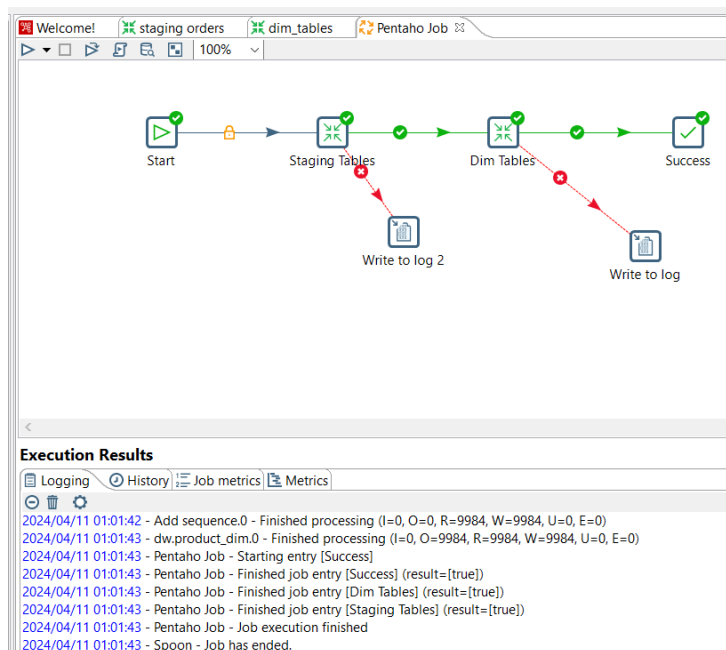
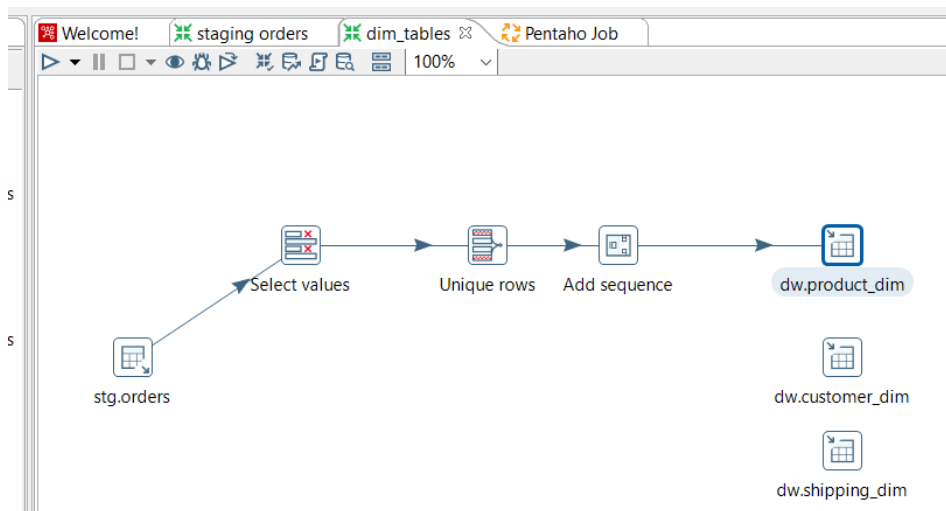
2. Скачать примеры Pentaho jobs для Staging и Dimension Tables. Выполнено



The screenshot shows the Pentaho DI (Spoon) interface. On the left, the 'Transformations' tree is expanded, showing 'staging orders' and its sub-items: 'Run configurations', 'Database connections', 'Steps', 'Hops', 'Partition schemas', 'Slave server', 'Kettle cluster schemas', and 'VFS Connections'. The main workspace displays a job design with three input nodes ('Orders input', 'People input', 'Returns input') connected to three output nodes ('stg.orders', 'stg.peoples', 'stg.returns'). Below the workspace, the 'Execution Results' tab is active, showing a log of execution steps and their completion times.

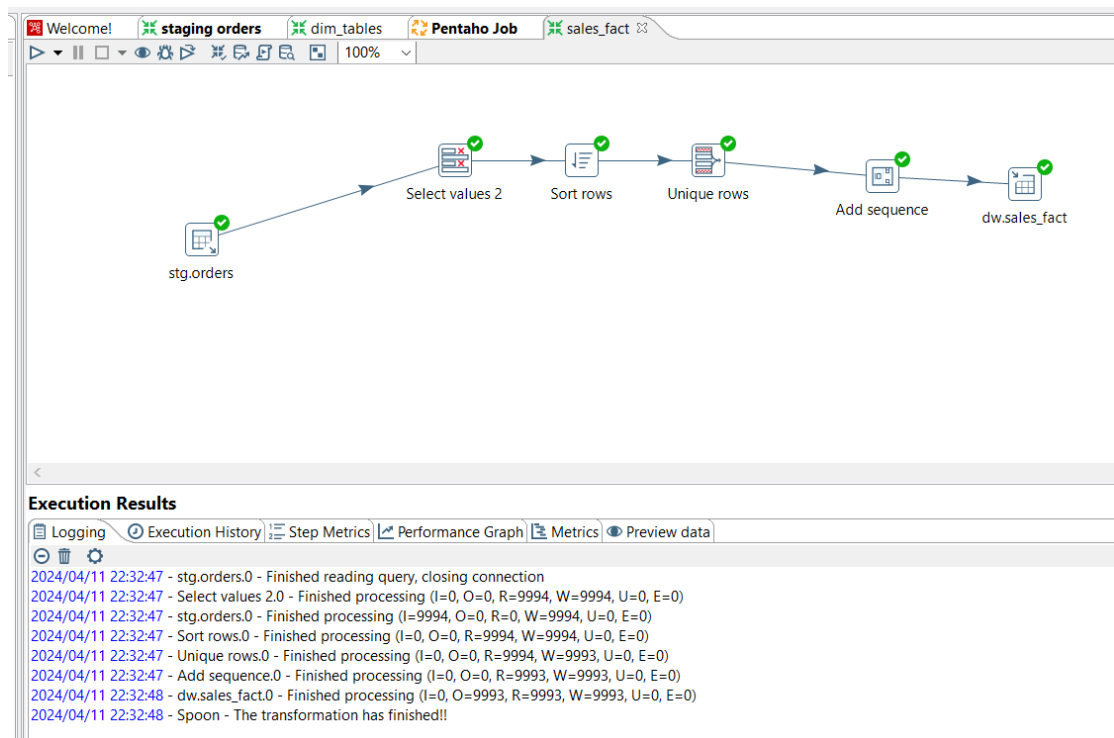
Execution Results

- 2024/04/11 00:55:59 - Returns input.0 - Finished processing (I=800, O=0, R=0, W=800, U=0, E=0)
- 2024/04/11 00:56:00 - Orders input.0 - Finished processing (I=9994, O=0, R=0, W=9994, U=0, E=0)
- 2024/04/11 00:56:00 - stg.peoples.0 - Finished processing (I=0, O=4, R=4, W=4, U=0, E=0)
- 2024/04/11 00:56:00 - stg.returns.0 - Finished processing (I=0, O=800, R=800, W=800, U=0, E=0)
- 2024/04/11 00:56:01 - stg.orders.0 - Finished processing (I=0, O=9994, R=9994, W=9994, U=0, E=0)
- 2024/04/11 00:56:01 - Spoon - The transformation has finished!!
- 2024/04/11 01:01:38 - Spoon - Starting job...
- 2024/04/11 01:01:43 - Spoon - Job has ended.

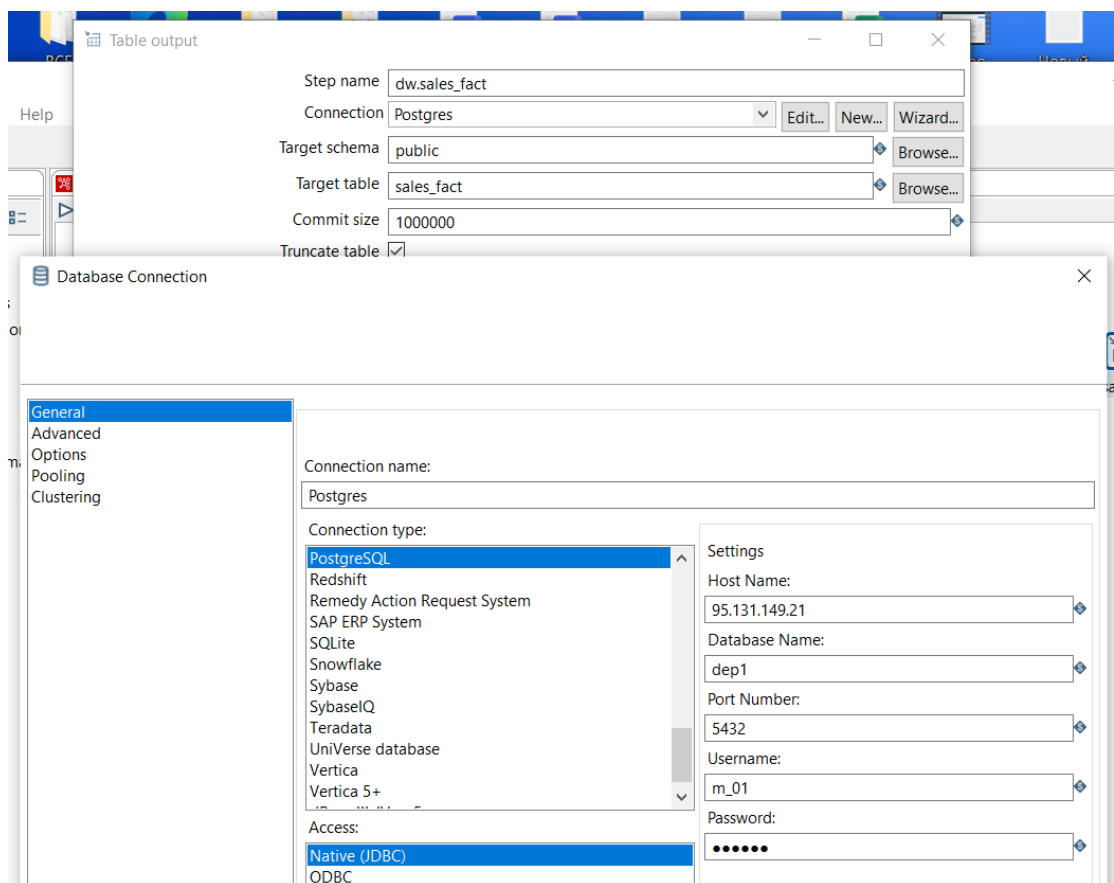


3. Создайте еще одну трансформацию, в которой создать **sales_fact** таблицу.

ВЫПОЛНЕНО



Используется БД Postgres, доступ к которой был выдан в начале курса.




Вид таблицы Sales_fact:

order_id	customer_id	customer_name	ship_date	ship_mode	product_id	region	country	city	state	quantity	sales	profit	discount
1016-101000	DP-13000	Darren Powers	16-01-07 00:00:00.000	Standard Class	OFF-PA-10000174	Central	United States	Houston	Texas	2	16,448	5,5512	
1016-106054	JO-15145	Jack O'Brian	16-01-07 00:00:00.000	First Class	OFF-BR-10002399	South	United States	Athens	Georgia	3	12,78	5,2398	
1016-112326	PO-19195	Phillina Ober	16-01-08 00:00:00.000	Standard Class	OFF-BI-10004094	Central	United States	Naperville	Illinois	2	3,54	-5,487	
1016-112326	PO-19195	Phillina Ober	16-01-08 00:00:00.000	Standard Class	OFF-LA-10003223	Central	United States	Naperville	Illinois	3	11,784	4,2717	
1016-112326	PO-19195	Phillina Ober	16-01-08 00:00:00.000	Standard Class	OFF-ST-10002743	Central	United States	Naperville	Illinois	3	272,736	-64,7748	
1016-130813	LS-17230	Lycoris Saunders	16-01-08 00:00:00.000	Second Class	OFF-PA-10002005	West	United States	Los Angeles	California	3	19,44	9,3312	
1016-167199	ME-17320	Maria Etezadi	16-01-10 00:00:00.000	Standard Class	FUR-CH-10004063	South	United States	Henderson	Kentucky	9	2,573.82	746,4078	
1016-167199	ME-17320	Maria Etezadi	16-01-10 00:00:00.000	Standard Class	OFF-AR-10001662	South	United States	Henderson	Kentucky	2	5,48	1,4796	
1016-167199	ME-17320	Maria Etezadi	16-01-10 00:00:00.000	Standard Class	OFF-BI-10004632	South	United States	Henderson	Kentucky	2	609,98	274,491	
1016-167199	ME-17320	Maria Etezadi	16-01-10 00:00:00.000	Standard Class	OFF-FA-10001883	South	United States	Henderson	Kentucky	4	31,12	0,3112	
1016-167199	ME-17320	Maria Etezadi	16-01-10 00:00:00.000	Standard Class	OFF-PA-10000955	South	United States	Henderson	Kentucky	1	6,54	3,0084	

Main options		Database fields	
Fields to insert:			
#	Table field	Stream field	
1	order_id	order_id	
2	customer_id	customer_id	
3	customer_n...	customer_na...	
4	ship_date	ship_date	
5	ship_mode	ship_mode	
6	product_id	product_id	
7	region	region	
8	country	country	
9	city	city	
1..	state	state	
1..	quantity	quantity	
1..	sales	sales	
1..	profit	profit	
1..	discount	discount	
1..	sales_PK	sales_PK	

- 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.

Выполнено (см. В другом файле)

 ETL_подсистемы(компоненты)_Кузьме...