Building Sales Data mart Using **Pentaho** **Part 2 (continued)**
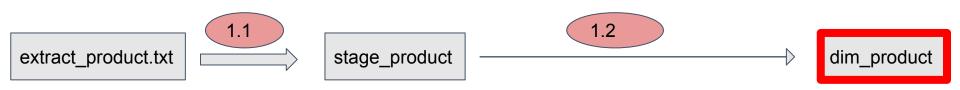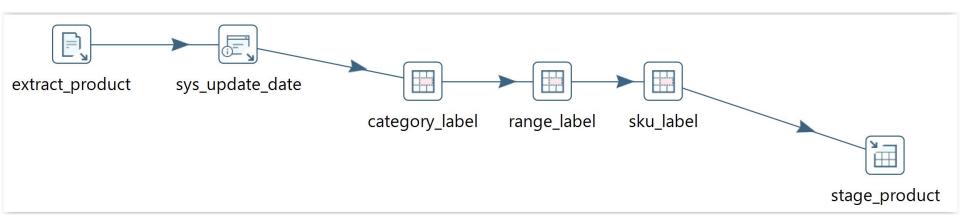
By Naheed Anjum Arafat

Task 1 - Product Dimension Table

Task 1.1
`Product file → stage_product`
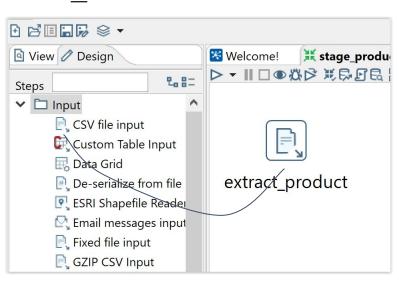
Objective

Task 1.1 - Create Transformation

- New transformation in spoon
- **Save as** `stage_product.ktr` **in folder** `dim_product`

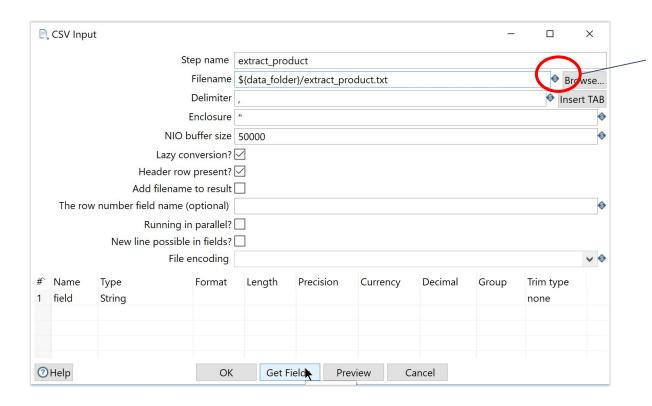# Task 1.1 - Create Transformation
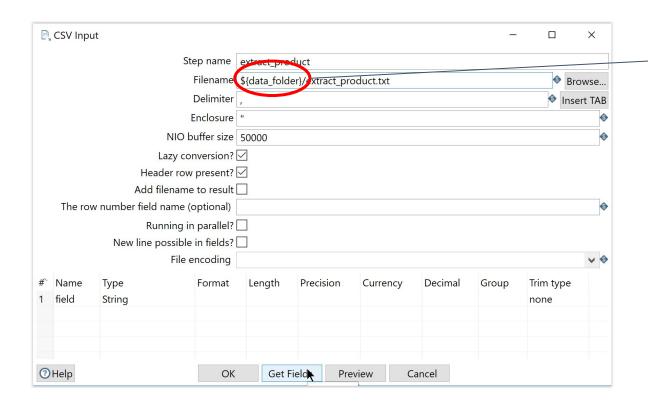
- Edit the CSV Input step
- Set the file path to
  `${data_folder}/extract_product.txt`
- Get fields as ALL STRING

CSV Input Step

CSV Input Step

CSV Input Step
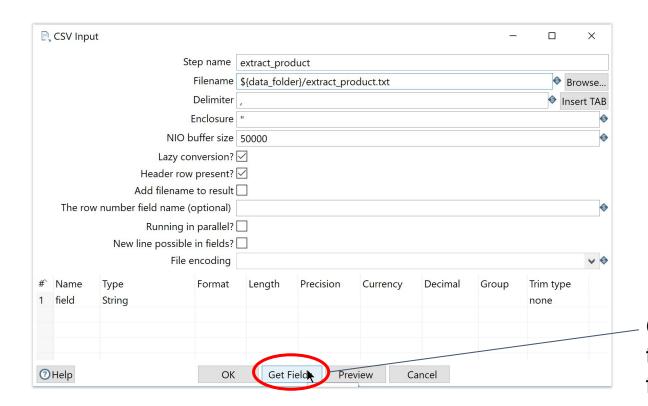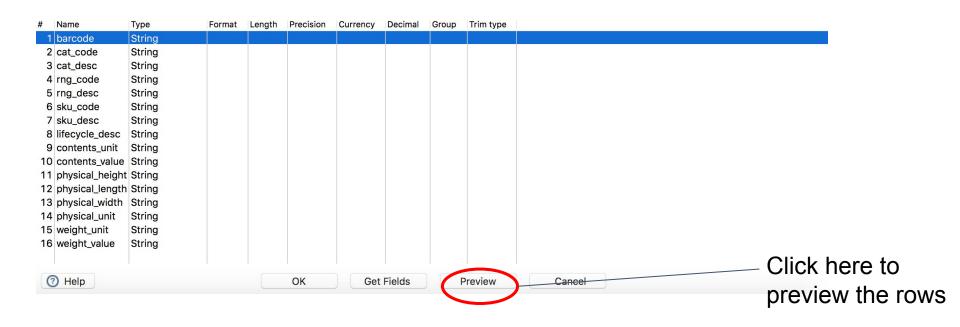
Click here to get the fields from text file

CSV Input Step

#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim type
1	barcode	String							
2	cat_code	String							
3	cat_desc	String							
4	rng_code	String							
5	rng_desc	String							
6	sku_code	String							
7	sku_desc	String							
8	lifecycle_desc	String							
9	contents_unit	String							
10	contents_value	String							
11	physical_height	String							
12	physical_length	String							
13	physical_width	String							
14	physical_unit	String							
15	weight_unit	String							
16	weight_value	String							

? Help   OK   Get Fields   Preview   Cancel

Click here to preview the rows

Preview rows

Rows of step: CSV file input (1000 rows)

#	barcode	cat_code	cat_desc	rng_code	rng_desc	sku_code	sku_desc
1	1010000008	10	FACIAL SKIN CARE/1	100	VITAMIN C	0000000000...	VIT.C SKIN BOOST 30ML
2	101000008	10	FACIAL SKIN CARE/1	100	VITAMIN C	0000000000...	VIT.C SKIN BOOST 30ML
3	2000000000...	10	FACIAL SKIN CARE/1	100	VITAMIN C	0000000000...	VIT.C SKIN BOOST 30ML
4	5028197245...	10	FACIAL SKIN CARE/1	100	VITAMIN C	0000000000...	VIT.C SKIN BOOST 30ML
5	5028197290...	10	FACIAL SKIN CARE/1	100	VITAMIN C	0000000000...	VIT.C SKIN BOOST 30ML
6	126000017	12	WOMENS FRAGRANCE/3	600	WHITE MUSK	0000000000...	NEW WHITE MUSK LUXURY SOAP 100GR
7	2000000004...	12	WOMENS FRAGRANCE/3	600	WHITE MUSK	0000000000...	NEW WHITE MUSK LUXURY SOAP 100GR
8	5028197253...	12	WOMENS FRAGRANCE/3	600	WHITE MUSK	0000000000...	NEW WHITE MUSK LUXURY SOAP 100GR
9	111060000	24	BATH AND SHOWER/26	261	ALOE VERA	0000000000...	ALOE VERA ANTI-PERSPIRANT DEODORANT 50ML
10	2000000000...	24	BATH AND SHOWER/26	261	ALOE VERA	0000000000...	ALOE VERA ANTI-PERSPIRANT DEODORANT 50ML
11	5028197258...	24	BATH AND SHOWER/26	261	ALOE VERA	0000000000...	ALOE VERA ANTI-PERSPIRANT DEODORANT 50ML
12	101000016	10	FACIAL SKIN CARE/1	100	VITAMIN C	0000000000...	VIT.C INTENS NIGHT TREAT 30ML
13	2000000000...	10	FACIAL SKIN CARE/1	100	VITAMIN C	0000000000...	VIT.C INTENS NIGHT TREAT 30ML
14	5028197246...	10	FACIAL SKIN CARE/1	100	VITAMIN C	0000000000...	VIT.C INTENS NIGHT TREAT 30ML
15	111060018	23	BODY CARE/25	261	ALOE VERA	0000000000...	NEW ALOE ANTI-PERSPIRANT DEODORANT 50ML
16	2000000000...	23	BODY CARE/25	261	ALOE VERA	0000000000...	NEW ALOE ANTI-PERSPIRANT DEODORANT 50ML
17	5028197257...	23	BODY CARE/25	261	ALOE VERA	0000000000...	NEW ALOE ANTI-PERSPIRANT DEODORANT 50ML
18	5028197431...	23	BODY CARE/25	261	ALOE VERA	0000000000...	NEW ALOE ANTI-PERSPIRANT DEODORANT 50ML
19	126000025	12	WOMENS FRAGRANCE/3	600	WHITE MUSK	0000000000...	NEW WHITE MUSK BATH & SHOWER GEL 200ML
20	2000000000...	12	WOMENS FRAGRANCE/3	600	WHITE MUSK	0000000000...	NEW WHITE MUSK BATH & SHOWER GEL 200ML
21	5028197373...	12	WOMENS FRAGRANCE/3	600	WHITE MUSK	0000000000...	NEW WHITE MUSK BATH & SHOWER GEL 200ML
22	101000024	10	FACIAL SKIN CARE/1	100	VITAMIN C	0000000000...	VIT.C SPF15 DAILY MOIST 50ML
23	2000000000...	10	FACIAL SKIN CARE/1	100	VITAMIN C	0000000000...	VIT.C SPF15 DAILY MOIST 50ML
24	5028197137...	10	FACIAL SKIN CARE/1	100	VITAMIN C	0000000000...	VIT.C SPF15 DAILY MOIST 50ML
25	5028197246...	10	FACIAL SKIN CARE/1	100	VITAMIN C	0000000000...	VIT.C SPF15 DAILY MOIST 50ML
26	0	23	BODY CARE/25	261	ALOE VERA	0000000000...	ALOE CARING BODY BALM 200ML
27	111060026	23	BODY CARE/25	261	ALOE VERA	0000000000...	ALOE CARING BODY BALM 200ML
28	2000000000...	23	BODY CARE/25	261	ALOE VERA	0000000000...	ALOE CARING BODY BALM 200ML
29	5028197257...	23	BODY CARE/25	261	ALOE VERA	0000000000...	ALOE CARING BODY BALM 200ML
30	126000033	12	WOMENS FRAGRANCE/3	600	WHITE MUSK	0000000000...	NEW WHITE MUSK BODY LOTION 200ML
31	2000000000...	12	WOMENS FRAGRANCE/3	600	WHITE MUSK	0000000000...	NEW WHITE MUSK BODY LOTION 200ML
32	5028197253...	12	WOMENS FRAGRANCE/3	600	WHITE MUSK	0000000000...	NEW WHITE MUSK BODY LOTION 200ML
33	111060034	23	BODY CARE/25	261	ALOE VERA	0000000000...	ALOE COOLING BODY GEL 250ML
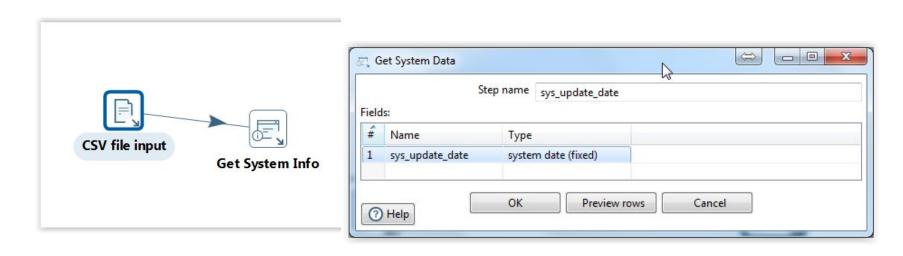
Close   Show Log

When you do not want all the fields

The row number field name (optional)

Running in parallel?

New line poss

Insert before this row	
Insert after this row	
Move up	⌘↑
Move down	⌘↓
Optimal Column size incl. header	F3
Optimal Column size excl. header	F4
Clear all	
Select all rows	⌘A
Clear selection	↻
Filtered selection	⌘F
Copy selected lines to clipboard	⌘C
Paste clipboard to table	⌘V
Cut selected lines	⌘X
Delete selected lines	⌧
Keep only selected lines	⌘K
Copy field value to all rows	
Undo : not available	⌘Z
Redo : not available	⌘Y

#	Name	Type	Format			Trim type
1	barcode	String				
2	cat_code	String				
3	cat_desc	String				
4	rng_code	String				
5	rng_desc	String				
6	sku_code	String				
7	sku_desc	String				
8	lifecycle_desc	String				
9	contents_unit	String				
10	contents_value	String				
11	physical_height	String				
12	physical_length	String				
13	physical_width	String				
14	physical_unit	String				
15	weight_unit	String				
16	weight_value	String				

Select the attributes you don't want. Right click and delete the line.

? Help

eview   Cancel

Good practice: Add an update date
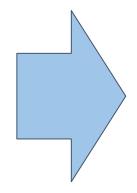
User friendly attribute name

cat_code	cat_desc	rng_code	rng_desc	sku_code	sku_desc
10	FACIAL SKIN CARE/1	100	VITAMIN C	0000000000010	VIT.C SKIN BOOST 30ML
23	BODY CARE/25	261	ALOE VERA	0000000000013	NEW ALOE ANTI-PERSPIRANT DEODORANT 50ML
23	BODY CARE/25	261	ALOE VERA	0000000000016	ALOE CARING BODY BALM 200ML
23	BODY CARE/25	261	ALOE VERA	0000000000018	ALOE COOLING BODY GEL 250ML
25	MAKE UP/27	018	ACCESSORIES	0000000000025	DIVA BRUSH RETRACTABLE LIP

User friendly name
Ease of use in
Tableau

category_label	range_label	sku_label
10 - FACIAL SKIN CARE/1	100 - VITAMIN C	0000000000010 - VIT.C SKIN BOOST 30ML
23 - BODY CARE/25	261 - ALOE VERA	0000000000013 - NEW ALOE ANTI-PERSPIRANT DEODORANT 50ML
23 - BODY CARE/25	261 - ALOE VERA	0000000000016 - ALOE CARING BODY BALM 200ML
23 - BODY CARE/25	261 - ALOE VERA	0000000000018 - ALOE COOLING BODY GEL 250ML
25 - MAKE UP/27	018 - ACCESSORIES	0000000000025 - DIVA BRUSH RETRACTABLE LIP

Transform fields

#	Fieldname	Type
1	barcode	String
2	cat_code	String
3	cat_desc	String
4	rng_code	String
5	rng_desc	String
6	sku_code	String
7	sku_desc	String
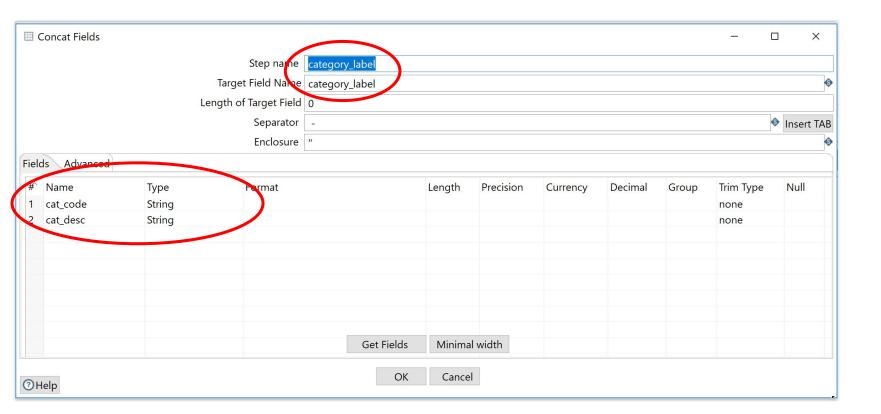
#	Fieldname	Type
1	barcode	String
2	category_label	String
3	range_label	String
4	sku_label	String

Concat fields

- cat_code + " - " + cat_desc → category_label
- rng_code + " - " + rng_desc → range_label
- sku_code + " - " + sku_desc → sku_label

Concat fields - category_label

Concat fields

Click on Advanced tab

Remove the fields

extract_product → sys_update_date → category_label

Concat fields - range_label

Concat fields

Length of Target Field 0

Separator	- · Insert TAB
Enclosure	"

Fields **Advanced**

Click on Advanced tab

Remove selected fields? ☑

Force the enclosure around ☐

Disable the enclosure fix? ☐

Header ☐

Footer ☐

Encoding ▼

Right pad fields ☐

Fast data dump (no formatting) ☐

Split every ... rows 0

Add Ending line after last row

Remove the fields

⑦ Help   OK   Cancel

extract_product → sys_update_date → category_label → range_label

Concat fields - sku_label

Concat fields

Click on Advanced tab

Remove the fields

Length of Target Field | 0
Separator | - | Insert TAB
Enclosure | "

Fields | Advanced

Remove selected fields? ☑
Force the enclosure around ☐
Disable the enclosure fix? ☐
Header ☐
Footer ☐
Encoding [▼]
Right pad fields ☐
Fast data dump (no formatting) ☐
Split every ... rows | 0
Add Ending line after last row [   ]

⊘ Help

OK   Cancel

extract_product → sys_update_date → category_label → range_label → sku_label

Table output

- Edit Table output step
- change name of step to `stage_product`
- change table name to `stage_product`

Table output

Step name	stage_product
Connection	postgres [Edit...] [New...] [Wizard...]
Target schema	[_____] [Browse...]
Target table	stage_product [_____] [Browse...]
	stage_product
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	☑
Return auto-generated key	☐
Name of auto-generated key field	[_____]

[? Help]   [OK]   [Cancel]   [SQL]

Set Truncate table to YES

Finally, Click on SQL

SQL statement to execute - automatic

```
CREATE TABLE stage_product
(
  barcode TEXT
, sys_update_date TIMESTAMP
, category_label TEXT
, range_label TEXT
, sku_label TEXT
)
;
```

Simple SQL editor

SQL statements, separated by semicolon ';'

Line 9 column 1

Execute   Clear cache   Close

Click Execute
to create the table

The SQL statements had the following results

```
SQL executed: CREATE TABLE stage_product
(
  barcode TEXT
, sys_update_date TIMESTAMP
, category_label TEXT
, range_label TEXT
, sku_label TEXT
)

1 SQL statements executed
```

SQL was executed
Table created
No errors

Click OK
to close the dialog
box

Simple SQL editor

SQL statements, separated by semicolon ';'

```
CREATE TABLE stage_product
(
  barcode TEXT
, sys_update_date TIMESTAMP
, category_label TEXT
, range_label TEXT
, sku_label TEXT
)
;
```

Line 9 column 1

Execute   Clear cache   Close

Click Close
to close dialog box

Table output

Step name: stage_product

Connection: postgres   Edit...   New...   Wizard...

Target schema:

Target table: stage_product   Browse...

stage_product

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: ☑

Return auto-generated key: ☐

Name of auto-generated key field:

? Help   OK   Cancel   SQL
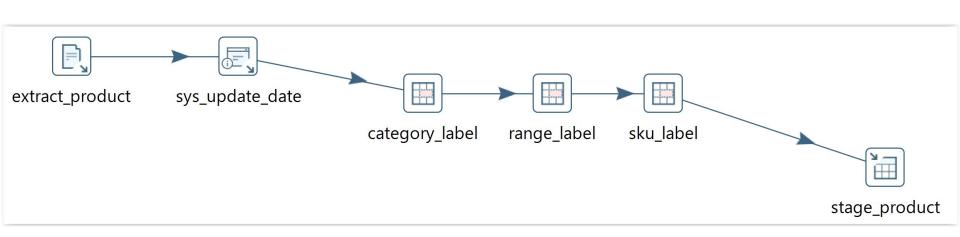
Click OK
to close the dialog
box

Task 1.1 - Run Transformation

- Run Transformation
- Watch step metrics
- Watch for errors in logging

Show results pane

Click on this icon to show Results pane

Show the execution results pane

extract_product → sys_update_date → category_label → range_label → sku_label → stage_product
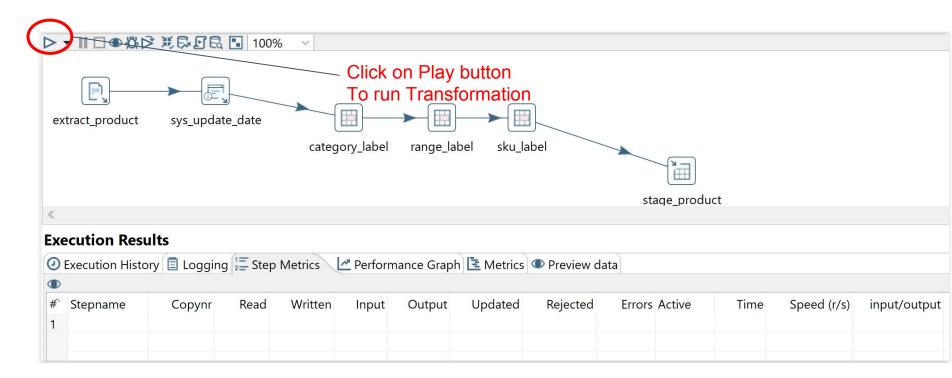
100%

Show results pane

Click on Step metrics tab

Show results pane

Click on Play button
To run Transformation

extract_product → sys_update_date → category_label → range_label → sku_label → stage_product

Execution Results

Execution History | Logging | Step Metrics | Performance Graph | Metrics | Preview data

#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	input/output
1													
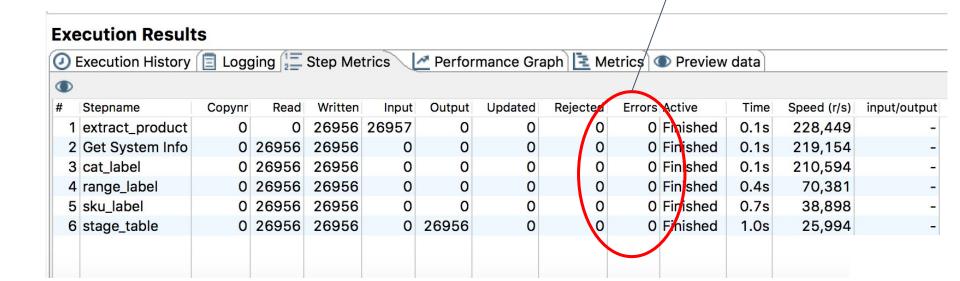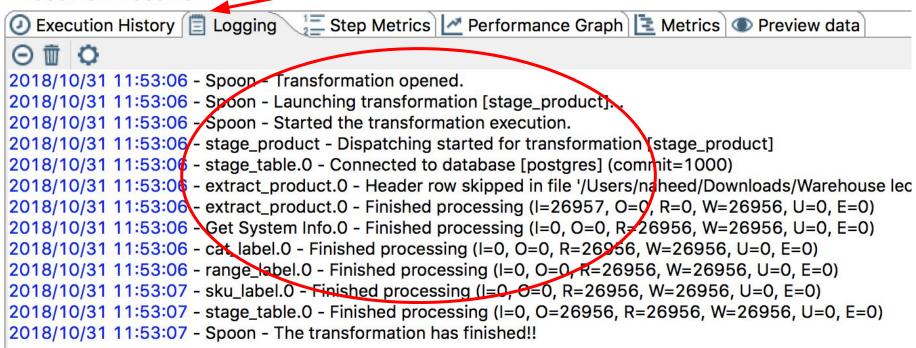
Run transformation

Run is completed
No errors

Execution Results

Execution History | Logging | Step Metrics | Performance Graph | Metrics | Preview data

#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	input/output
1	extract_product	0	0	26956	26957	0	0	0	0	Finished	0.1s	228,449	-
2	Get System Info	0	26956	26956	0	0	0	0	0	Finished	0.1s	219,154	-
3	cat_label	0	26956	26956	0	0	0	0	0	Finished	0.1s	210,594	-
4	range_label	0	26956	26956	0	0	0	0	0	Finished	0.4s	70,381	-
5	sku_label	0	26956	26956	0	0	0	0	0	Finished	0.7s	38,898	-
6	stage_table	0	26956	26956	0	26956	0	0	0	Finished	1.0s	25,994	-

Check the logging Tab

Execution Results

Execution History | 📋 Logging | Step Metrics | Performance Graph | Metrics | Preview data

⊖ 🗑 ⚙

2018/10/31 11:53:06 - Spoon - Transformation opened.
2018/10/31 11:53:06 - Spoon - Launching transformation [stage_product]...
2018/10/31 11:53:06 - Spoon - Started the transformation execution.
2018/10/31 11:53:06 - stage_product - Dispatching started for transformation [stage_product]
2018/10/31 11:53:06 - stage_table.0 - Connected to database [postgres] (commit=1000)
2018/10/31 11:53:06 - extract_product.0 - Header row skipped in file '/Users/naheed/Downloads/Warehouse led
2018/10/31 11:53:06 - extract_product.0 - Finished processing (I=26957, O=0, R=0, W=26956, U=0, E=0)
2018/10/31 11:53:06 - Get System Info.0 - Finished processing (I=0, O=0, R=26956, W=26956, U=0, E=0)
2018/10/31 11:53:06 - cat_label.0 - Finished processing (I=0, O=0, R=26956, W=26956, U=0, E=0)
2018/10/31 11:53:06 - range_label.0 - Finished processing (I=0, O=0, R=26956, W=26956, U=0, E=0)
2018/10/31 11:53:07 - sku_label.0 - Finished processing (I=0, O=0, R=26956, W=26956, U=0, E=0)
2018/10/31 11:53:07 - stage_table.0 - Finished processing (I=0, O=26956, R=26956, W=26956, U=0, E=0)
2018/10/31 11:53:07 - Spoon - The transformation has finished!!
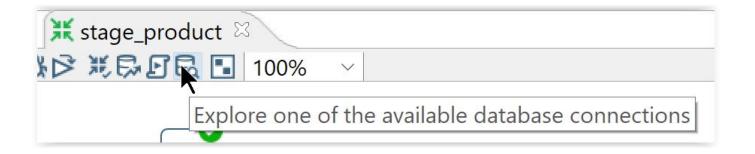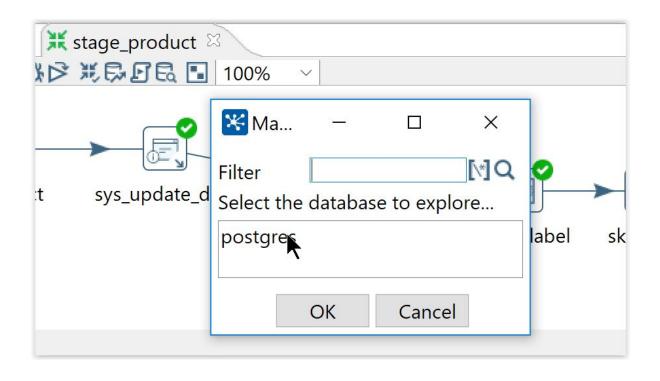
Check your work

- Run twice
- Browse the database
- Row count stage_product
- Check whether this count is the same as that in the source text file
- If not, check whether truncate table is set
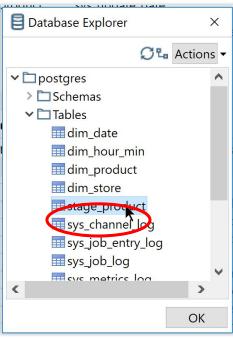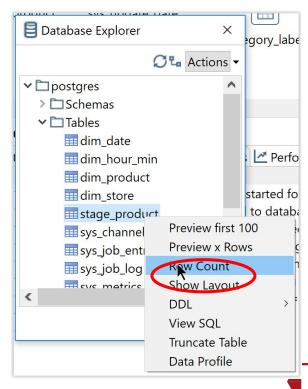
Browse database

Browse database

Browse database

Click on ">"

Navigate to `tables`
Right-click on `stage_product`

Click on Row Count

End of Task 1.1