

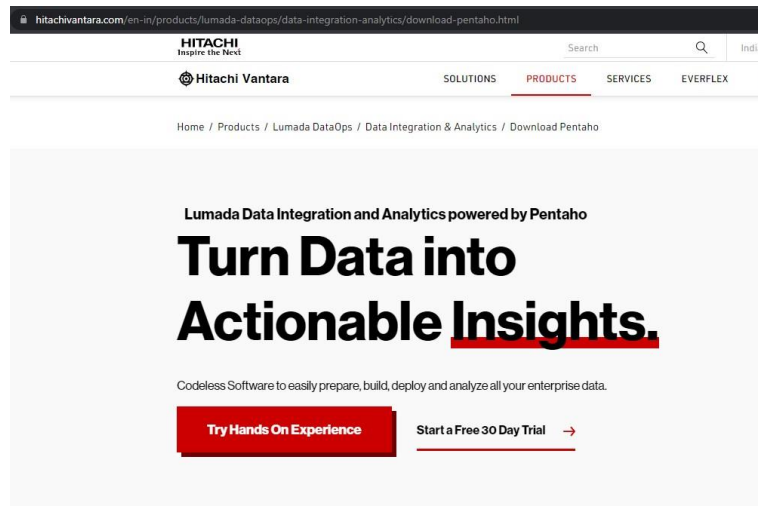
Nama : Samuel R Yosan

NIM : 201402067

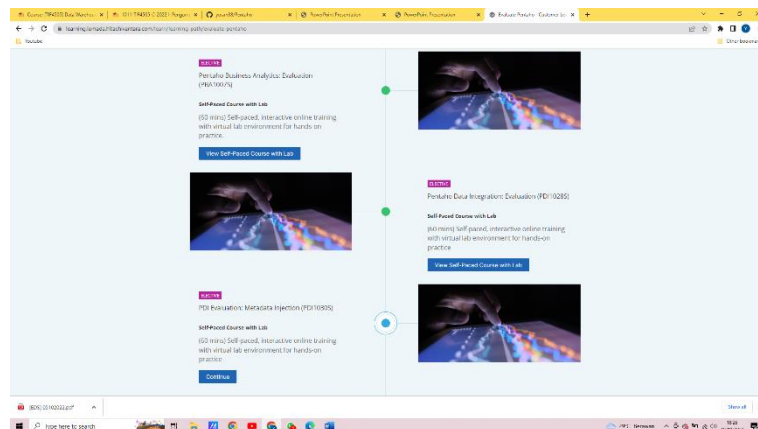
## HITACHI & Pentaho

Pada kesempatan saat ini saya akan menampilkan Pentaho Data Evaluation dimana memiliki hands-on yang dibuat juga oleh sang pemilik, Hitachi, dimana dapat dilihat dari web di bawah ini.

<https://www.hitachivantara.com/en-in/products/lumada-dataops/data-integration-analytics/download-pentaho.html>

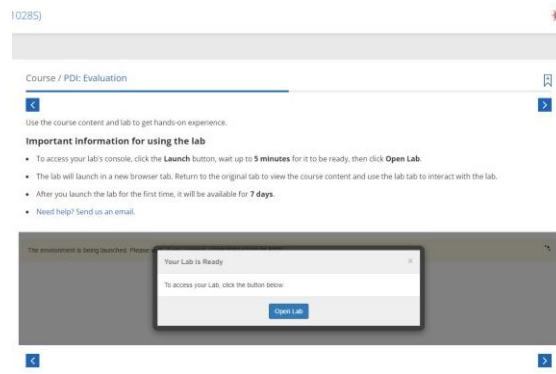


Hands-on Pentaho dapat dilihat di bawah ini.



Dari tampilan di atas, dapat dilihat materi hands-on dari Pentaho. Selanjutnya adalah tampilan materi yang disajikan oleh Pentaho.

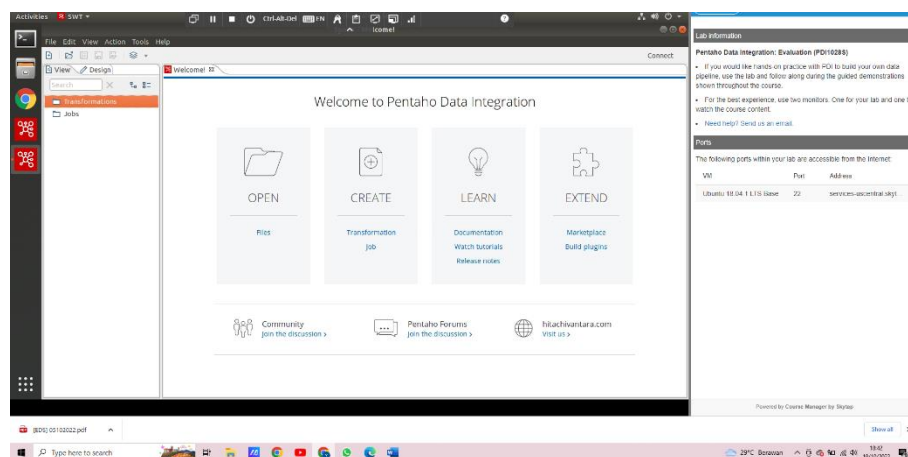




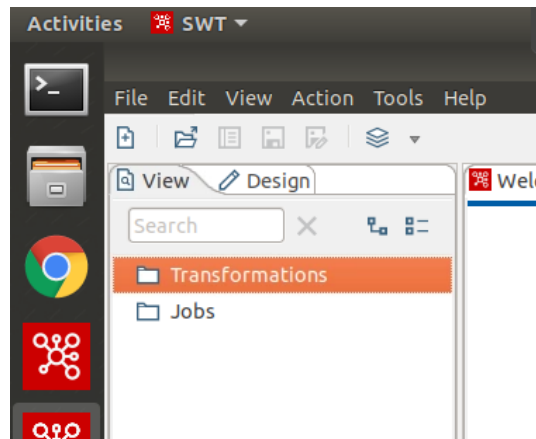
Saat kita sudah membuka materi hands-on, mari kita membuka lab yang disediakan



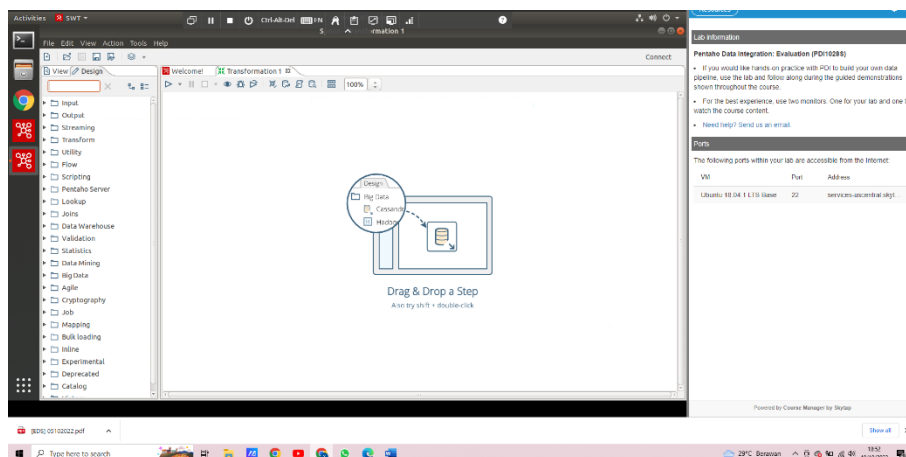
Pertama, kita akan membaca inputan data terlebih dahulu. Misalnya data customer. Untuk lebih lanjut, mari diawali dengan buka Spoon dari lab dimana kita membuka hands-on.



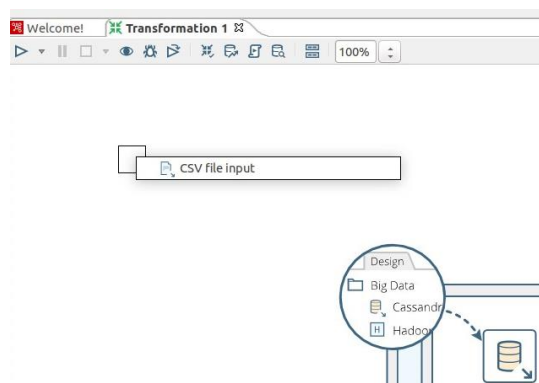
Pada saat ini, kita akan mencoba transformasi data pelanggan menjadi data yang bersih dan siap dipakai dalam database. Pada bagian kiri atas, pilih new kemudian transformation.



Dan muncul lembar baru untuk transformation data. Lembar kerja baru ini (Drag & Drop a Step pada gambar) dikenal juga sebagai canvas.



Kita mulai dengan step input customer data dengan drag CSV file input pada tab Design ke dalam canvas. Diikuti dengan double-click untuk konfigurasi.



Konfigurasi dapat berupa step name, filename dapat berupa sumber data, delimiter, enclosure, dan sebagainya sebagaimana yang sesuai untuk menginput data.

Step name: csvi-customers

Filename: /home/pentaho/course\_files/data\_prep/customers.csv

Delimiter: ,

Enclosure: "

NIO buffer size: 50000

Lazy conversion? ☒

Header row present? ☒

Add filename to result ☐

The row number field name (optional):

Running in parallel? ☐

New line possible in fields? ☐

Format: mixed

File encoding:

Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim type
1								

Buttons: Help, OK, Get Fields, Preview, Cancel

Dilanjutkan dengan click pada Get Fields untuk dapat nama kolom yang ada pada input data. Tetapkan 0 untuk Number of lines sample untuk mendapat nama kolom data secara keseluruhan.

Step name: csvi-customers

Filename: /home/pentaho/course\_files/data\_prep/customers.csv

Delimiter: ,

Enclosure: "

NIO buffer size: 50000

Lazy conversion? ☒

Header row present? ☒

Add filename to result ☐

The row number field name (optional):

Running in parallel? ☐

New line possible in fields? ☐

Format: mixed

File encoding:

Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim type
1								

Buttons: Help, OK, Get Fields, Preview, Cancel

Sample data dialog:

Number of lines to sample: 0

Show sample summary ☐

Buttons: OK, Cancel

Maka nama-nama kolom beserta propertinya akan muncul pada bagian bawah dialog konfigurasi step.

CSV file input

Step name:

Filename:

Delimiter:

Enclosure:

NIO buffer size:

Lazy conversion? ☒

Header row present? ☒

Add filename to result ☐

The row number field name (optional):

Running in parallel? ☐

New line possible in fields? ☐

Format:

File encoding:

	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim type
1	customer_number	Integer	#	15	0	\$	.	,	none
2	customer_name	String		28		\$	.	,	none
3	last_name	String		9		\$	.	,	none
4	first_name	String		9		\$	.	,	none
5	address	String		26		\$	.	,	none
6	city	String		13		\$	.	,	none
7	state	String		10		\$	.	,	none
8	postal_code	Integer	#	15	0	\$	.	,	none
9	country	String		3		\$	.	,	none
10	region	Integer	#	15	0	\$	.	,	none
11	sales_rep_employee_num	Integer	#	15	0	\$	.	,	none

Untuk dapat melihat contoh data tersedia dari file tersebut, klik tombol preview pada bagian bawah dialog dan masukkan preview size sesuai keinginan banyak data yang ditampilkan.

Preview size

Enter the number of rows to preview

Dan muncul kotak dialog baru yang memuat data preview dari file input yang ditentukan. Dari preview ini kita dapat mengamati bagian data yang belum bersih seperti masih terdapatnya null pada kolom sales\_rep\_employee\_num, penulisan state dan address yang tidak konsisten dalam penggunaan tanda titik, hingga tidak konsistennya penggunaan singkatan pada kolom address. Jika diclick di salah satu data kolom state, ternyata masih terdapat white-space yang tidak diperlukan. Untuk itu kita harus membersihkan data agar lebih efisien untuk digunakan.

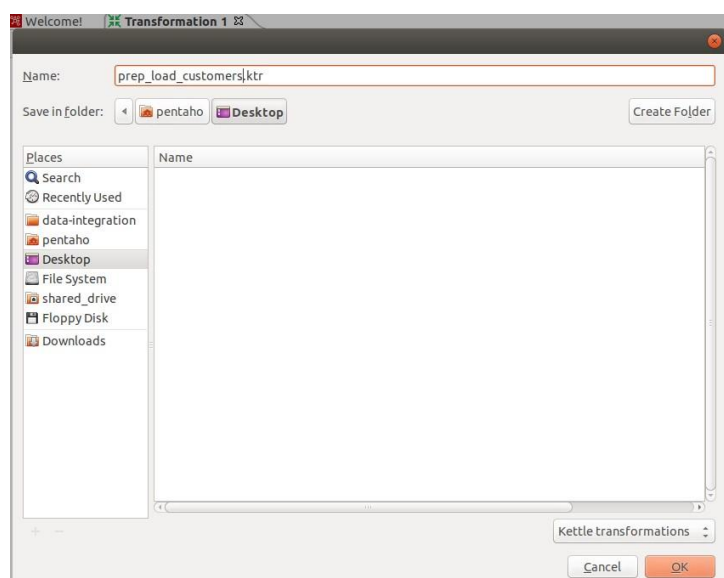
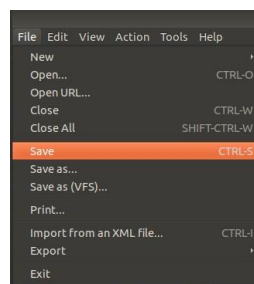
Examine preview data

Rows of step: csvi-customers (36 rows)

umber	customer_name	last_name	first_name	address	city	state	postal_code	country	region	sales_rep_employee_num
112	Signal Gift Stores	King	Sue	245 Edgefield St.	Las Vegas	NV.	83030	USA	14	1166
124	Mini Gifts Distributors Ltd.	Nelson	Valarie	982 Lilac Street	San Rafael	CA	97562	USA	15	<null>
129	Mini Wheels Co.	Murphy	Julie	125 Augusta St.	San Francisco	CA	94217	USA	15	<null>
131	Land of Toys Inc.	Yu	Kwai	9 South Summit Road	NYC	NY	10022	USA	10	1323
151	Muscle Machine Inc	Young	Jeff	9788 Union Avenue	NYC	NY	10022	USA	10	<null>
157	Diecast Classics Inc.	Yu	Kyung	907 Windsor St.	Allentown	PA	70267	USA	4	1216
161	Technics Stores Inc.	Hirano	Juri	7473 Bishop Ave.	Burlingame	CA.	94217	USA	16	1165
168	American Souvenirs Inc	Franco	Sue	541 East Halifax Street	New Haven	CT.	97823	USA	13	1286
173	Cambridge Collectables Co.	Tseng	Kyung	271 South Moon Drive	Cambridge	MA	51247	USA	5	1188
175	Gift Depot Inc.	King	Julie	9920 Pennsylvania Lane	Bridgewater	CT	97562	USA	3	1323
181	Vitachrome Inc.	Frick	Michael	48 Tower Lane	NYC	NY	10022	USA	10	<null>
198	Auto-Moto Classics Inc.	Taylor	Leslie	8132 East Riverside Street	Brickhaven	MA	58339	USA	2	1216
204	Online Mini Collectables	Barajas	Miguel	8093 Archer St	Brickhaven	MA	58339	USA	2	1188
205	Toys4GrownUps.com	Young	Julie	663 Blue Spring Lane	Pasadena	CA	90003	USA	16	1166
219	Boards & Toys Co.	Young	Leslie	377 Java Ave.	Glendale	CA	92561	USA	16	<null>
239	Collectable Mini Designs Co.	Thompson	Valarie	341 South Lumber St.	San Diego	CA	91217	USA	17	<null>
286	Marta's Replicas Co.	Hernandez	Marta	904 Meadowbrook Street	Cambridge	MA	51247	USA	5	1216
319	Mini Classics	Frick	Steve	429 Achorage St.	White Plains	NY.	24067	USA	11	1323
320	Mini Creations Ltd.	Tam	Wing C	357 John Drive	New Bedford	MA.	50553	USA	7	1188
321	Corporate Gift Ideas Co.	Brown	Julie	908 Gray St.	San Francisco	CA	94217	USA	15	1165
328	Tekni Collectables Inc.	Brown	William	266 Court Drive	Newark	NJ	94019	USA	9	1323
339	Classic Gift Ideas, Inc	Cervantes	Francisca	798 Pinnacle Ave.	Philadelphia	PA	71270	USA	12	1188
347	Men'R' US Retailers, Ltd.	Chandler	Michael	529 Carriage St.	Los Angeles	CA.	91003	USA	16	1166
362	Gifts4AllAges.com	Yoshido	Juri	222 Green Lake St.	Boston	MA	51003	USA	1	1216
363	Online Diecast Creations Co.	Young	Valarie	7987 Manchester St.	Nashua	NH	62005	USA	8	1216
379	Collectables For Less Inc.	Nelson	Allen	26 Corporation St.	Brickhaven	MA	58339	USA	2	<null>
424	Classic Legends Inc.	Hernandez	Maria	8847 West Oceanview St.	NYC	NY	10022	USA	10	1286
447	Gift Ideas Corp.	Lewis	Dan	655 Wentworth Street	Glendale	CT	97561	USA	6	1323
450	The Sharp Gifts Warehouse	Frick	Sue	817 Pin Oak St.	San Jose	CA	94217	USA	15	1165
455	Super Scale Inc.	Murphy	Leslie	834 Bank Lane	New Haven	CT.	97823	USA	13	1286
456	Microscale Inc.	Kuo	Kee	43 Fairfield Court	NYC	NY	10022	USA	10	1286

Close Show Log

Setelah menyelesaikan konfigurasi pada file input, click OK. Sekarang, kita akan menyimpan perubahan pada canvas transformation terlebih dahulu. Pada tab file pojok kiri atas, pilih save dan simpan pada lokasi yang diinginkan.



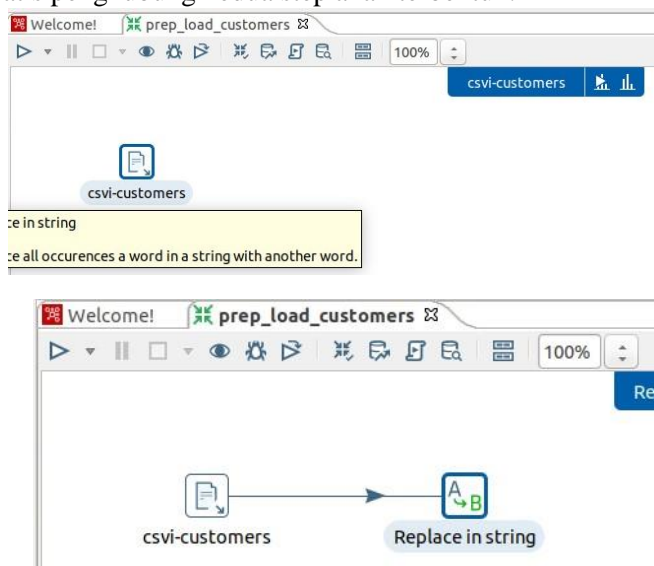


# Data Preparation Scenario

HITACHI  
Inspire the Next

## Standardize Address and State Data (Replace in String Step)

Selanjutnya kita akan melakukan standarisasi format untuk beberapa kolom, dalam penerapan kali ini kolom address dan state. Pertama, mari mencari step Replace in string pada tab design. Untuk menghubungkan kedua step, click step csvi-customers dan double-click pada step Replace in string, secara otomatis penghubung kedua step akan terbentuk.



Kemudian dilanjutkan konfigurasi step Replace in string lewat double-click padanya. Standarisasi diperlukan adalah menghapus tanda titik pada kolom address dan state, dan mengubah singkatan St dan Ave menjadi Street dan Avenue. Sekarang, click OK.

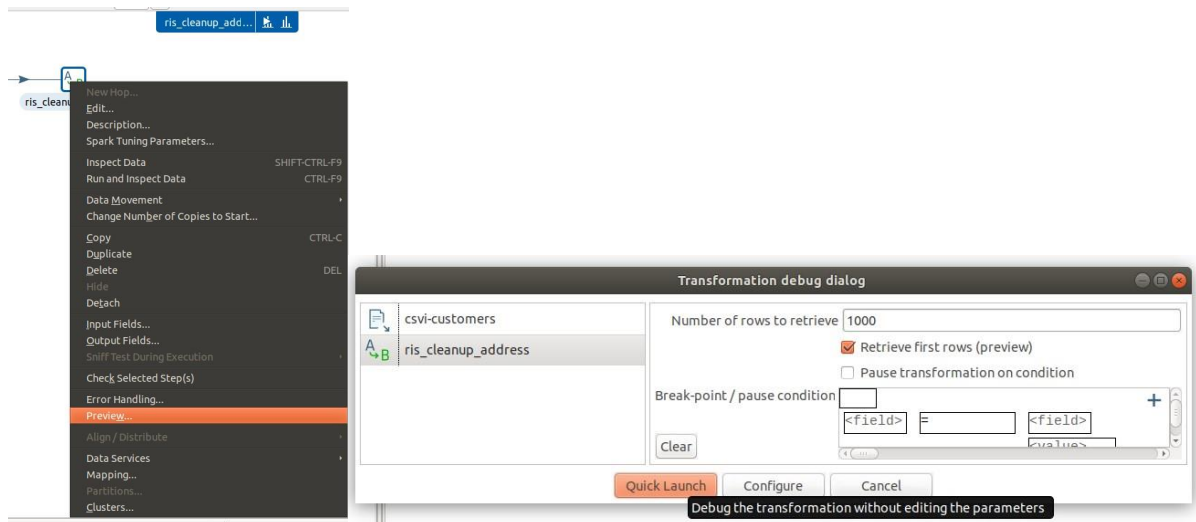
Replace in string

Step name:

	In stream field	Out stream field	use RegEx	Search	Replace with	Set empty string?	Replace with field	Whole Word	Case sensitive
1	address			.	Street			N	
2	address			St	Street			Y	
3	address			Ave	Avenue			Y	
4	state			.					

Buttons: Help, OK, Get fields, Cancel

Right-click step tersebut, pilih preview agar menampilkan data hasil standarisasi. Sebelum preview selesai, kita dapat konfigurasi pada transformation debug dialog. Bila konfigurasi selesai, click Quick Launch.



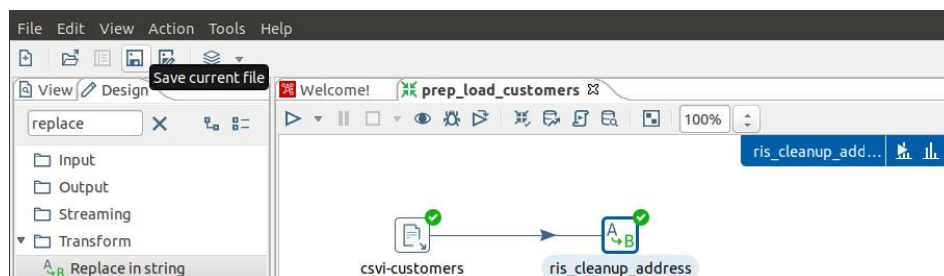
Dengan demikian, kita dapat melihat hasil standarisasi data customer di kotak dialog baru. Dan kita bisa lakukan save progress.

Examine preview data

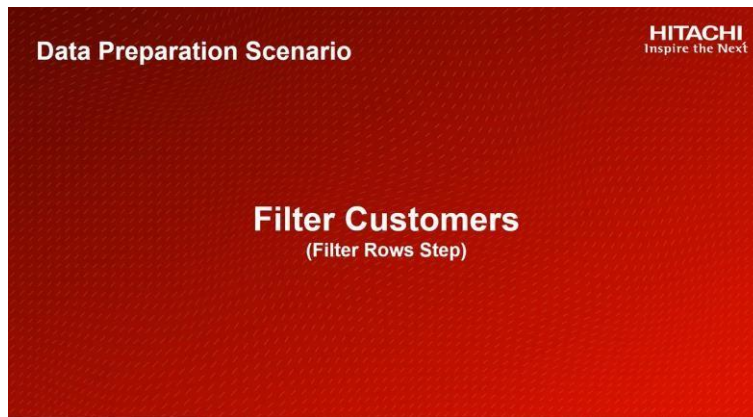
Rows of step: ris\_cleanup\_address (36 rows)

	customer_number	customer_name	last_name	first_name	address	city	state	postal_code	country	region	sales
1	112	Signal Gift Stores	King	Sue	245 Edgefield Street	Las Vegas	NV	83030	USA	14	
2	124	Mini Gifts Distributors Ltd.	Nelson	Valarie	982 Lilac Street	San Rafael	CA	97562	USA	15	
3	129	Mini Wheels Co.	Murphy	Julie	125 Augusta Street	San Francisco	CA	94217	USA	15	
4	131	Land of Toys Inc.	Yu	Kwai	9 South Summit Road	NYC	NY	10022	USA	10	
5	151	Muscle Machine Inc.	Young	Jeff	9788 Union Avenue	NYC	NY	10022	USA	10	
6	157	Diecast Classics Inc.	Yu	Kyung	907 Windsor Street	Allentown	PA	70267	USA	4	
7	161	Technics Stores Inc.	Hirano	Juri	7473 Bishop Avenue	Burlingame	CA	94217	USA	16	
8	168	American Souvenirs Inc.	Franco	Sue	541 East Halifax Street	New Haven	CT	97823	USA	13	
9	173	Cambridge Collectables Co.	Tseng	Kyung	271 South Moon Drive	Cambridge	MA	51247	USA	5	
10	175	Gift Depot Inc.	King	Julie	9920 Pennsylvania Lane	Bridgewater	CT	97562	USA	3	
11	181	Vitachrome Inc.	Frick	Michael	48 Tower Lane	NYC	NY	10022	USA	10	
12	198	Auto-Moto Classics Inc.	Taylor	Leslie	8132 East Riverside Street	Brickhaven	MA	58339	USA	2	
13	204	Online Mini Collectables	Barajas	Miguel	8093 Archer Street	Brickhaven	MA	58339	USA	2	
14	205	Toys4GrownUps.com	Young	Julie	663 Blue Spring Lane	Pasadena	CA	90003	USA	16	
15	219	Boards & Toys Co.	Young	Leslie	377 Java Avenue	Glendale	CA	92561	USA	16	
16	239	Collectable Mini Designs Co.	Thompson	Valarie	341 South Lumber Street	San Diego	CA	91217	USA	17	
17	286	Marta's Replicas Co.	Hernandez	Marta	904 Meadowbrook Street	Cambridge	MA	51247	USA	5	
18	319	Mini Classics	Frick	Steve	429 Achorage Street	White Plains	NY	24067	USA	11	
19	320	Mini Creations Ltd.	Tam	Wing C	357 John Drive	New Bedford	MA	50553	USA	7	
20	321	Corporate Gift Ideas Co.	Brown	Julie	908 Gray Street	San Francisco	CA	94217	USA	15	
21	328	Tekni Collectables Inc.	Brown	William	266 Court Drive	Newark	NJ	94019	USA	9	
22	339	Classic Gift Ideas, Inc	Cervantes	Francisca	798 Pinnacle Avenue	Philadelphia	PA	71270	USA	12	
23	347	Men 'R' US Retailers, Ltd.	Chandler	Michael	529 Carriage Street	Los Angeles	CA	91003	USA	16	
24	362	Gifts4AllAges.com	Yoshido	Juri	222 Green Lake Street	Boston	MA	51003	USA	1	
25	363	Online Diecast Creations Co.	Young	Valarie	7987 Manchester Street	Nashua	NH	62005	USA	8	
26	379	Collectables For Less Inc.	Nelson	Allen	26 Corporation Street	Brickhaven	MA	58339	USA	2	
27	424	Classic Legends Inc.	Hernandez	Maria	8847 West Oceanview Street	NYC	NY	10022	USA	10	
28	447	Gift Ideas Corp.	Lewis	Dan	655 Wentworth Street	Glendale	CT	97561	USA	6	
29	450	The Sharp Gifts Warehouse	Frick	Sue	817 Pin Oak Street	San Jose	CA	94217	USA	15	
30	455	Super Scale Inc.	Murphy	Leslie	834 Bank Lane	New Haven	CT	97823	USA	13	
31	456	Microscale Inc.	Kuo	Kee	43 Fairfield Court	NYC	NY	10022	USA	10	

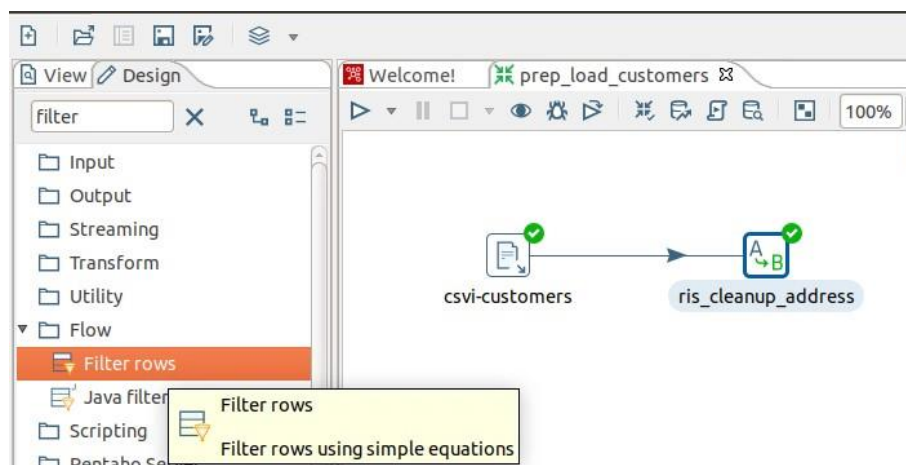
Close



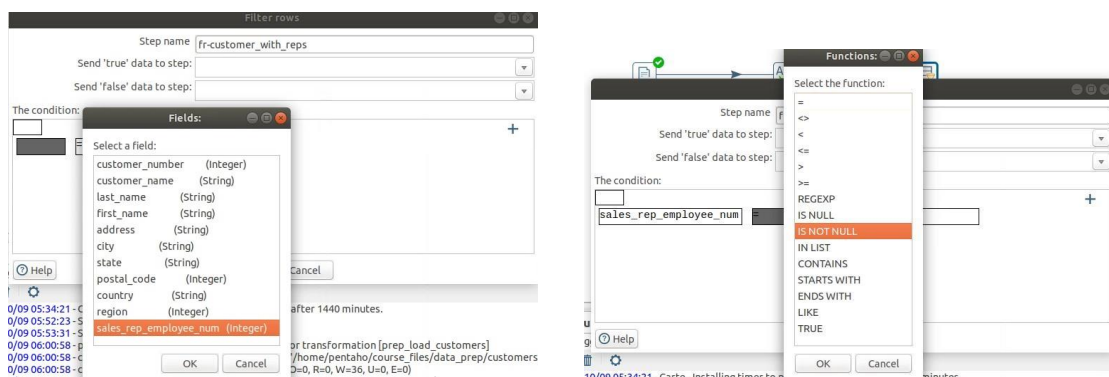




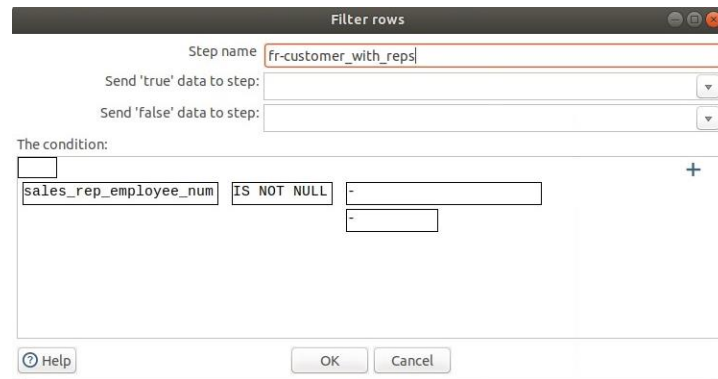
Kita lanjutkan dengan filter data terstandarisasi. Sebelumnya, cari step Filter rows pada tab Design terlebih dahulu dan double-click. Pastikan step ini terhubung dengan step ris\_cleanup\_address.



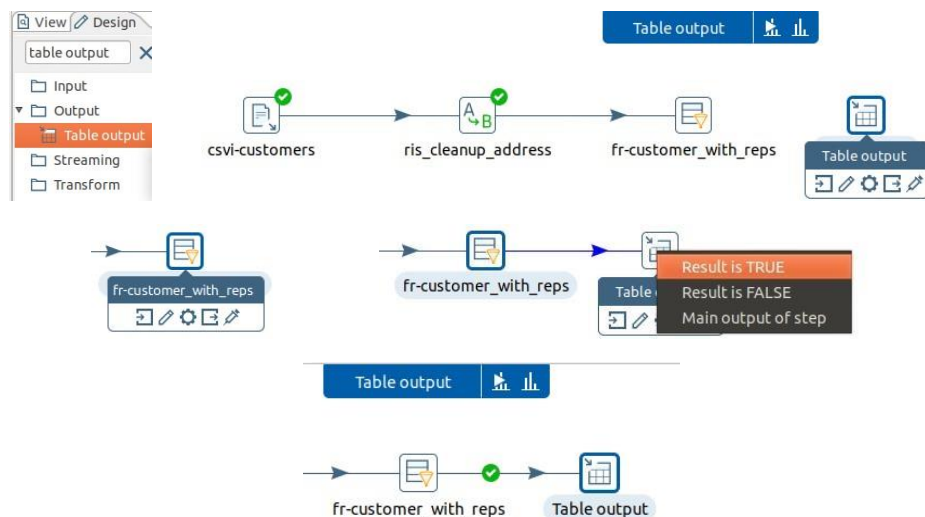
Double-click step untuk memulai konfigurasi. Bagian send 'true' dan 'false' dapat dikosongkan karena kita akan menambahkan penghubung step secara manual nantinya. Di kotak condition, pilih kolom sales\_rep\_employee\_num karena kolom ini adalah yang akan kita filtrasi. Di sebelah kanan, ada kotak yang bertujuan untuk menentukan function yang akan diterapkan pada nama kolom, pilih IS NOT NULL.



Konfigurasi ini bertujuan untuk filter kolom sales\_rep\_employee\_num yang bernilai null. Melalui step ini, data dapat dialihkan ke step selanjutnya dan data nilai null pada kolom dapat diproses lebih lanjut pada step lainnya. Konfigurasi dapat dilihat pada gambar berikut.



Pada tab Design, di step Table output dan drag berada di dalam canvas. Hover step filter dan pilih ikon kedua dari kanan agar garis terhubung ke step Table output dan pilih Result is TRUE.



Lakukan double-click pada output untuk memulai konfigurasi. Di dialog konfigurasi ini, kita dapat melihat SQL yang dibuat dengan click SQL pada bagian bawah.

Click tombol Execute dan table customers\_fixed pada penerapan contoh kali ini akan terbentuk pada database sampledata beserta struktur-struktur tablenya.

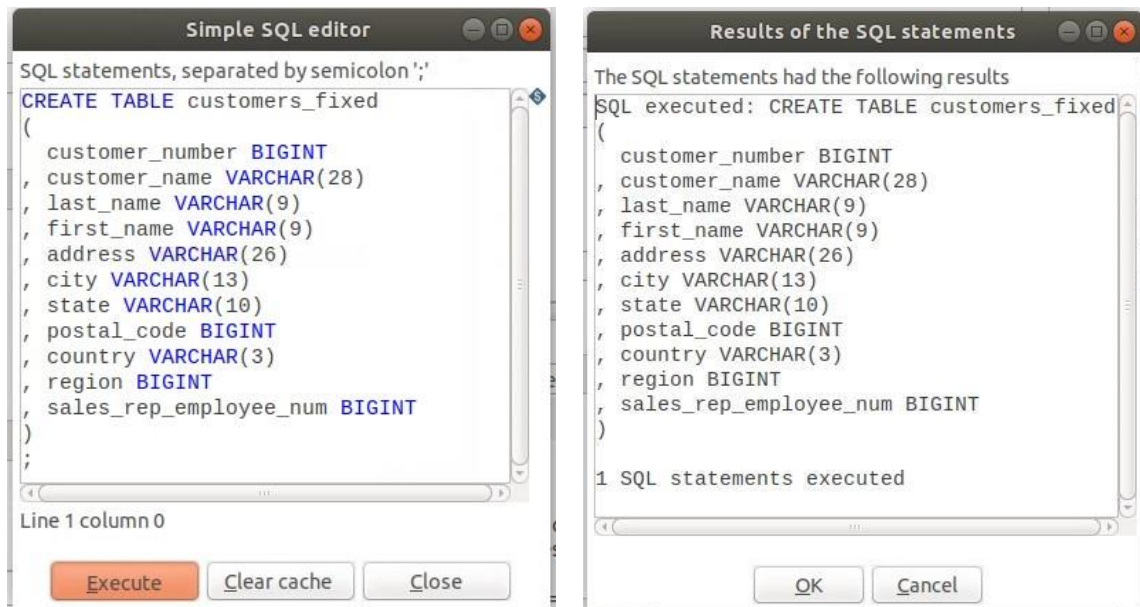
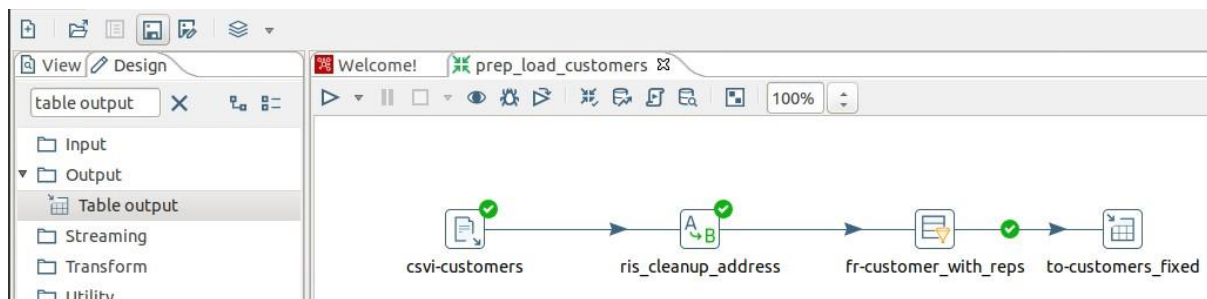


Table dan struktur terbentuk akan dapat dilihat pada penerapan selanjutnya. Selesai konfigurasi, tidak lupa untuk OK dan close. Dan simpan progress Anda.

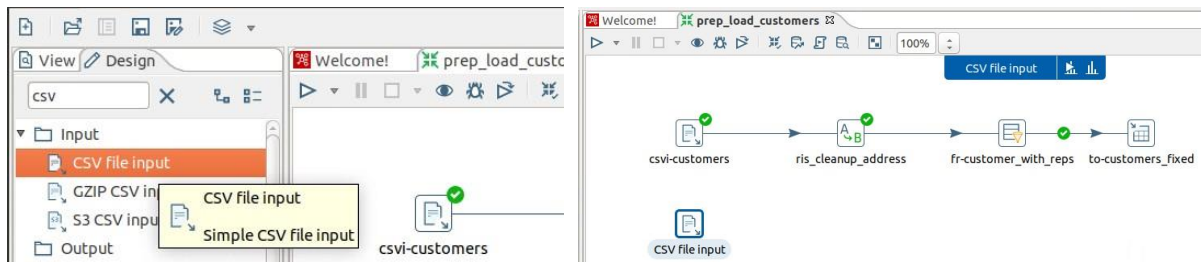


# Data Preparation Scenario

HITACHI  
Inspire the Next

## Reading Employee Data (CSV File Input Step)

Selanjutnya kita menginput file employee data sebagai dasar penanganan untuk bagian false pada filtered data bernilai null. Cari step CSV file input dan drag ke dalam canvas.



Lakukan konfigurasi dengan double-click pada dan juga Get Fields dan Preview pada data employee ini. Pada saat ini, file input data employee sudah bersih untuk kolom tertentu yang diperlukan saja.

**CSV file input**

Step name: csvi-employees

Filename: /home/pentaho/course\_files/data\_prep/employees.csv Browse...

Delimiter: , Insert TAB

Enclosure: "

NIO buffer size: 50000

Lazy conversion? ☒

Header row present? ☒

Add filename to result: ☐

The row number field name (optional):

Running in parallel? ☐

New line possible in fields? ☐

Format: mixed

File encoding:

Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim type
1 employee_number	Integer	#	15	0	\$	.	,	none
2 last_name	String		9		\$	.	,	none
3 first_name	String		8		\$	.	,	none
4 email	String		31		\$	.	,	none
5 region	Integer	#	15	0	\$	.	,	none
6 job_title	String		27		\$	.	,	none

Help
OK
Get Fields
Preview
Cancel

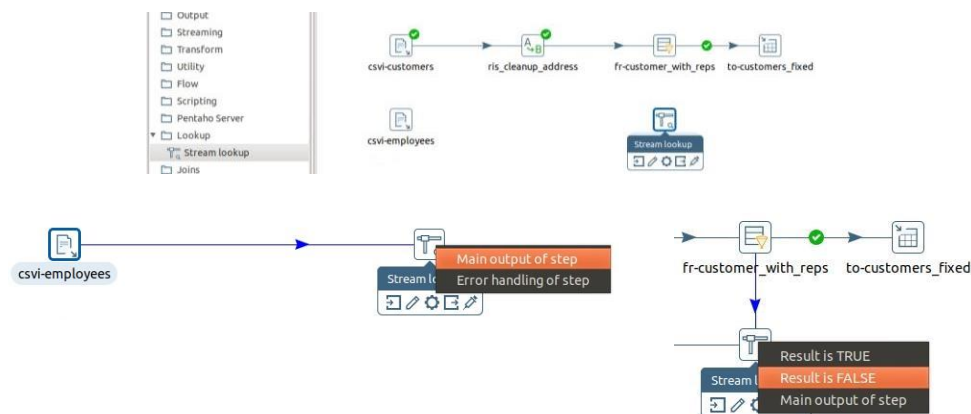
**Examine preview data**

Rows of step: csvi-employees (23 rows)

	employee_number	last_name	first_name	email	region	job_title
1	1002	Murphy	Diane	dmurphy@classicmodelcars.com	<null>	President
2	1056	Patterson	Mary	mpatterso@classicmodelcars.com	<null>	VP Sales
3	1076	Firrelli	Jeff	jfirrelli@classicmodelcars.com	<null>	VP Marketing
4	1088	Patterson	William	wpatterson@classicmodelcars.com	<null>	Sales Manager (JAPAN, APAC)
5	1102	Bondur	Gerard	athompson@classicmodelcars.com	<null>	Sale Manager (EMEA)
6	1143	Bow	Anthony	bhoward@classicmodelcars.com	<null>	Sales Manager (NA)
7	1165	Jennings	Leslie	ljennings@classicmodelcars.com	2	Sales Rep
8	1166	Thompson	Leslie	lthompson@classicmodelcars.com	1	Sales Rep
9	1188	Firrelli	Julie	ifirrelli@classicmodelcars.com	4	Sales Rep

## Lookup Employee Number (Stream Lookup Step)

Selanjutnya, kita ke Stream lookup pada tab Design dan drag ke dalam canvas. Dan hubungkan csvi-employees dengan step sekarang, dengan Shift dan menarik garis dari step csvi menuju step ini, set sebagai Main output of step. Dan hubungkan filtrasi sebelumnya (fr-customer\_with\_reps) dengan step ini dengan cara yang sama dan pilihan Result adalah FALSE.



Double-click step dan lakukan konfigurasi. Pada step ini, data mengacu pada kolom region sebagai lookup value untuk data employee dan customer. Juga tetapkan employee\_number dari data employee sebagai sales\_rep\_employee\_num\_found yang nantinya menggantikan nilai null agar data customer terintegrasi. Jangan lupa click OK dan save progress setelah selesai konfigurasi.

Stream lookup			
Step name		slkp-get_employess_num	
Lookup step		csvi-employees	
The key(s) to look up the value(s):			
Field	LookupField		
1	region	region	
Specify the fields to retrieve :			
Field	New name	Default	Type
1	employee_number	sales_rep_employee_num_found	
<input checked="" type="checkbox"/> Preserve memory (costs CPU)			
<input type="radio"/> Key and value are exactly one integer field			
<input type="radio"/> Use sorted list (i.s.o. hashtable)			
<input type="button" value="Help"/> <input type="button" value="OK"/> <input type="button" value="Cancel"/> <input type="button" value="Get Fields"/> <input type="button" value="Get lookup fields"/>			

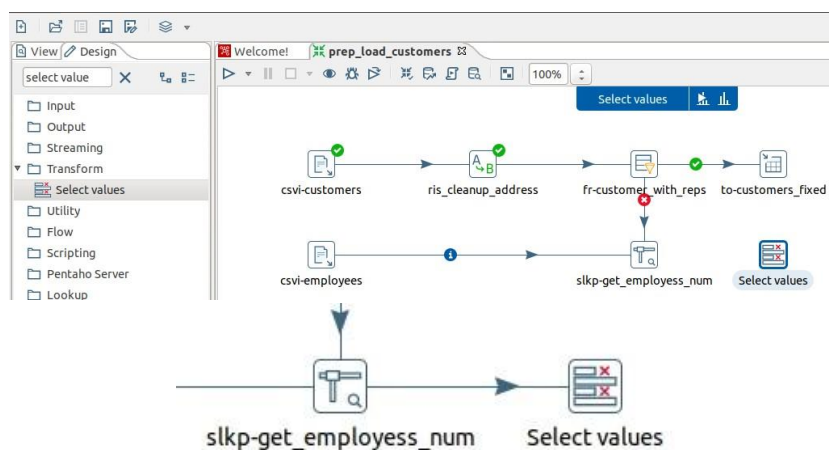


# Data Preparation Scenario

HITACHI  
Inspire the Next

## Format Field Layout (Select Values Step)

Selanjutnya, cari Select Values pada tab Design dan drag pada canvas, serta hubungkan stream lookup sebelumnya dengan step ini



Double-click untuk memulai konfigurasi. Silahkan click tombol Get fields to select daftar nama kolom (fields) pada tampilan konfigurasi tidak ada.

Select values

Step name: sv-format\_field\_layout

Select & Alter
Remove
Meta-data

Fields:

▼	Fieldname	Rename to	Length	Precision
1	customer_number			
2	customer_name			

Get fields to select  
Edit Mapping

Cari nama kolom sales\_rep\_employee\_num dan delete selected lines. Kolom akan dihapus karena berisi nilai null untuk data yang terfilter. Serta isi kolom sales\_rep\_employee\_num\_found pada kolom rename di sebelah kanannya dengan sales\_rep\_employee\_num.

5	address				
6	city				
7	state				
8	postal_code				
9	country				
10	region				
11	sales_rep_employee_num				
12	sales_rep_employee_num_found				

Cut selected lines
CTRL-X

Delete selected lines
DEL

Keep only selected lines
CTRL-K

Copy field value to all rows

Undo : not available
CTRL-Z

Redo : not available
CTRL-Y



Select values

Step name: sv-format\_field\_layout

Select & Alter Remove Meta-data

Fields:

	Fieldname	Rename to
1	customer_number	
2	customer_name	
3	last_name	
4	first_name	
5	address	
6	city	
7	state	
8	postal_code	
9	country	
10	region	
11	sales_rep_employee_num_found	sales rep employee num

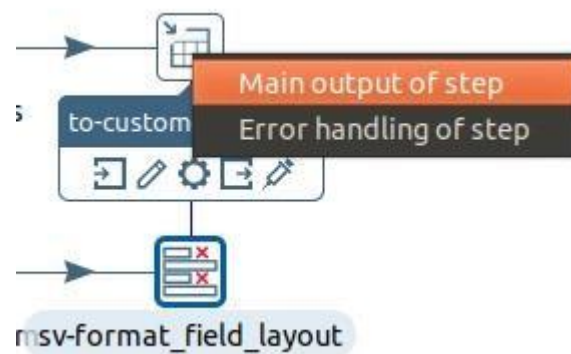
Get fields to select

Edit Mapping

Include unspecified fields, ordered t ☐

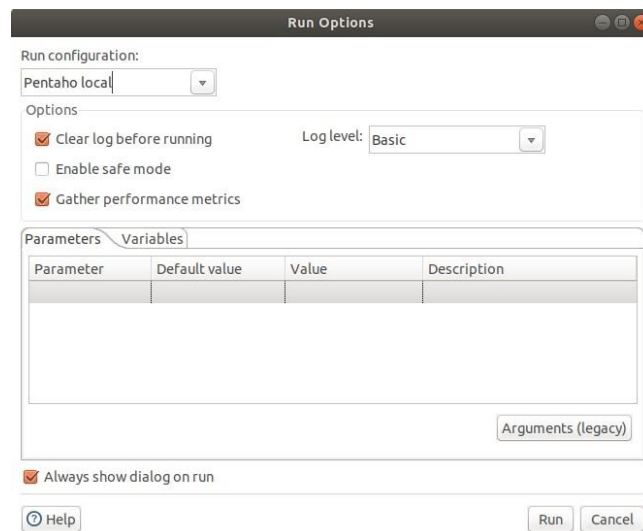
Help OK Cancel

Click OK. Hubungkan dengan step table output sebelumnya (to-customer\_fixed) dan click Main output of step of step. Dan save progress Anda.



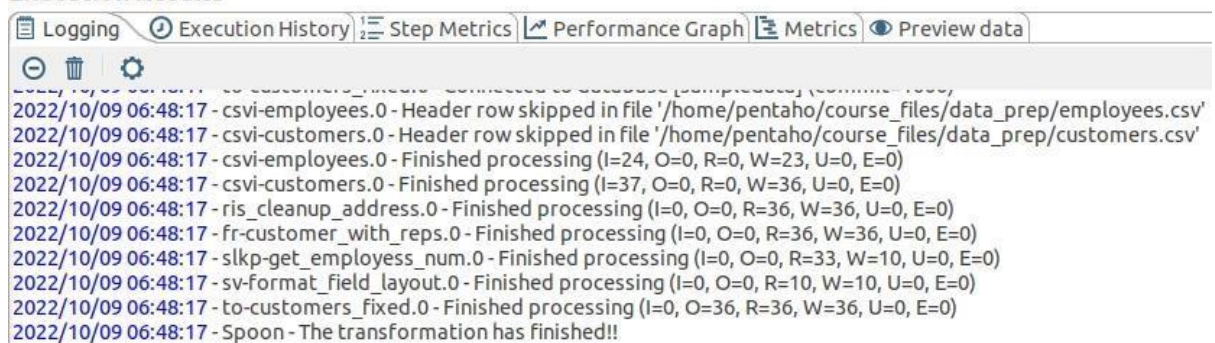


Untuk menjalankan hasil transformasi, click icon play pada bagian kiri atas canvas transformation. Akan muncul kotak dialog Run Options, click Run setelah selesai konfigurasi.



Proses dan hasil transformasi akan muncul pada bagian bawah aplikasi.

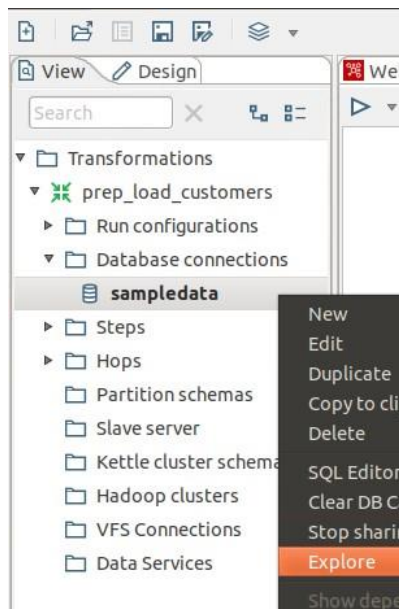
### Execution Results



Tab Metrics dapat menunjukkan pemrosesan data pada setiap step, sebagai contoh pada to-customers\_fixed memiliki property output bernilai 24, yang artinya terdapat 24 data yang masuk ke table customer\_fixed pada database sample data.

Stepname	Copypnr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)
1 csvi-customers	0	0	36	37	0	0	0	0	Finished	0.0s	5,286
2 ris_cleanup_address	0	36	36	0	0	0	0	0	Finished	0.0s	2,118
3 csvi-employees	0	0	23	24	0	0	0	0	Finished	0.0s	4,000
4 fr-customer_with_reps	0	36	36	0	0	0	0	0	Finished	0.0s	1,800
5 slkp-get_employess_num	0	33	10	0	0	0	0	0	Finished	0.2s	205
6 sv-format_field_layout	0	10	10	0	0	0	0	0	Finished	0.2s	61
7 to-customers_fixed	0	36	36	0	36	0	0	0	Finished	0.2s	199

Untuk melihat hasil transformation pada table, click View dan expand folder Transformations, prep\_load\_customers (nama transformation), Database connections. Click sampledata, right-click serta pilih explore.



Setelah muncul kotak dialog Database Explorer, right-click CUSTOMER\_FIXED pada Tables dan pilih Preview first 100. Kemudian akan tampil data hasil transformation yang sudah masuk ke dalam table tersebut sebanyak 100 data di awal.

