

Nama : Nanda Amelia

Nim : 191402015

Matkul : Data Warehouse & Business Intelligence

HITACHI & Pentaho

Pada Langkah pertama ini kita mulai dari halaman utama HITACHI yang berisikan penawaran pengenalan mengenai produknya Pentaho dapat diakses di link berikut <https://www.hitachivantara.com/en-in/products/lumada-dataops/data-integration-analytics/download-pentaho.html> .

The screenshot shows the homepage of the Hitachi Lumada Data Integration and Analytics powered by Pentaho. The top navigation bar includes links for Search, India, E-waste Management, Community, Get Support, and Partner. Below the navigation is a secondary menu with links for SOLUTIONS, PRODUCTS, SERVICES, INSIGHTS, NEWS, PARTNERS, COMPANY, and CONTACT US. A breadcrumb trail indicates the current page: Home / Products / Lumada DataOps / Data Integration & Analytics / Download Pentaho. The main content features a large headline "Turn Data into Actionable Insights." with a subtext "Codeless Software to easily prepare, build, deploy and analyze all your enterprise data." Below the headline are two buttons: "Try Hands On Experience" and "Start a Free 30 Day Trial". To the right of the text is a photograph of two workers standing on the nacelle of a wind turbine, looking out over a green landscape. At the bottom of the page, there is a cookie consent banner.

setelah itu kita akan mencoba hands-on dari Pentaho Data Integration dan akan dialihkan kehalaman <https://learning.lumada.hitachivantara.com/learning-paths/evaluate-pentaho>

The screenshot shows the "MILESTONE" section of the Lumada Learning Paths website under the "Pentaho Evaluation" category. It lists three evaluation courses:

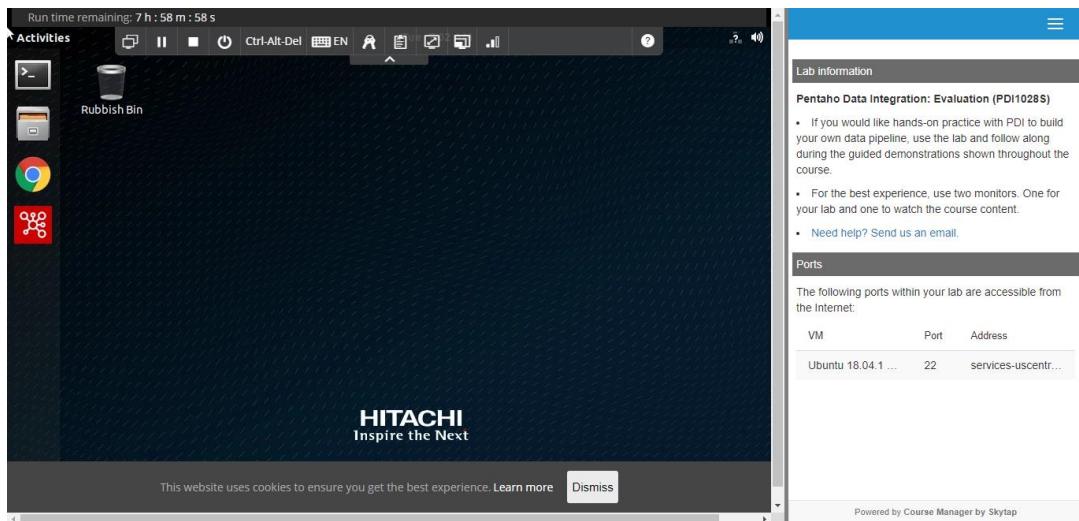
- Pentaho Business Analytics: Evaluation (PBA1007S)** (ELECTIVE)
Self-Paced Course with Lab
(60 mins) Self-paced, interactive online training with virtual lab environment for hands-on practice.
[View Details ↗](#)
- Pentaho Data Integration: Evaluation (PDI1028S)** (ELECTIVE)
Self-Paced Course with Lab
(60 mins) Self-paced, interactive online training with virtual lab environment for hands-on practice.
[View Details ↗](#)
- PDI Evaluation: Metadata Injection (PDI1030S)** (ELECTIVE)
Self-Paced Course with Lab
(60 mins) Self-paced, interactive online training with virtual lab environment for hands-on practice.
[View Details ↗](#)

Pada halaman ini kita dapat melihat daftar pembelajaran yang tersedia dari pentaho dan disini kita akan mencoba Pentaho Data Integration: Evaluation. Langkah pertama daftar terlebih dahulu pada pembelajaran tersebut disini juga tersedia Lab pembelajaran yang dapat kita akses melalui website untuk penerapan pembelajaran yang ada tanpa melakukan instalasi

The screenshot shows the Lumada Learning interface. On the left, there's a sidebar with a 'LESSONS' section. Under '01. COURSE', 'PDI: Evaluation' is expanded, showing 'LAB_PDI1028S' which is selected. Other items include 'Introduction', 'PDI: Evaluation', and 'Completion and Survey'. Below this is a 'NOTES' section with 'SUPPORT' and 'SIGN OUT' buttons. The main content area is titled 'Course / PDI: Evaluation'. It contains a brief description: 'Use the course content and lab to get hands-on experience.' A section titled 'Important information for using the lab' provides instructions: 'To access your lab's console, click the Launch button, wait up to 5 minutes for it to be ready, then click Open Lab.' It also states that the lab will launch in a new browser tab and be available for 7 days. A 'Launch' button is visible. At the bottom, a status message says 'Status: Ready to provision' with a yellow dot. The URL 'https://learning.lumada.hitachivantara.com/learn/sign_out' is shown at the bottom.

This screenshot shows the same Lumada Learning interface after the 'Launch' button was clicked. A modal window titled 'Your Lab is Ready' appears in the center. It contains the text 'The environment is being provisioned... Please wait.' and 'To access your Lab, click the button below.' with a blue 'Open Lab' button. The background content remains the same as in the previous screenshot, including the sidebar and course details.

Selanjutnya setelah kita melakukan permintaan akses pada lab ini, kita dapat mengakses virtual machine pada browser. Tampilan yang akan digunakan saat pembelajaran dapat kita lihat pada gambar berikut.



Pada bagian PDI: Evaluation untuk dapat memulai pembelajaran dan browser berisikan materi pembelajaran akan tampil.

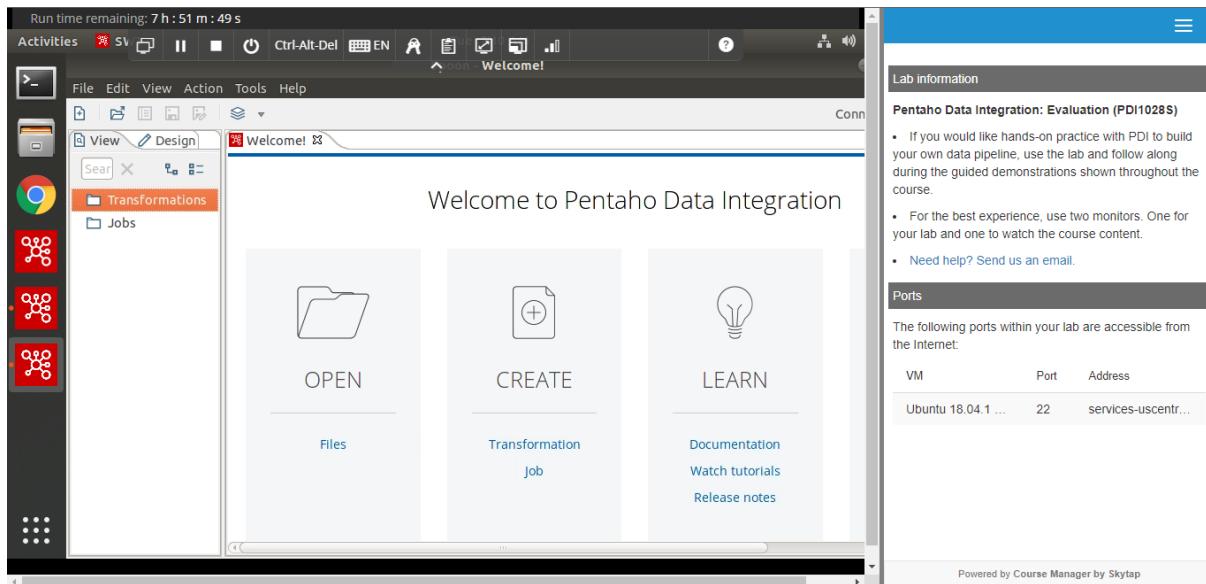
A screenshot of an eLearning platform interface. On the left is a sidebar with a 'LESSONS' section containing '01. COURSE' and a 'NOTES' section. The main content area is titled 'Course / PDI: Evaluation' and contains a message: 'Please complete this eLearning before continuing to the next page. (< will change to > after completion.)'. Below this is a large empty box labeled 'PDI: Evaluation' with a blue button 'Open Content In New Window'.

A screenshot of a course landing page. At the top right is a user profile for 'Joshua Morrison'. The main title is 'PDI Evaluation: Getting Started'. Below the title are two buttons: 'START COURSE' and 'DETAILS'. The background is red with a pattern of white diagonal lines.

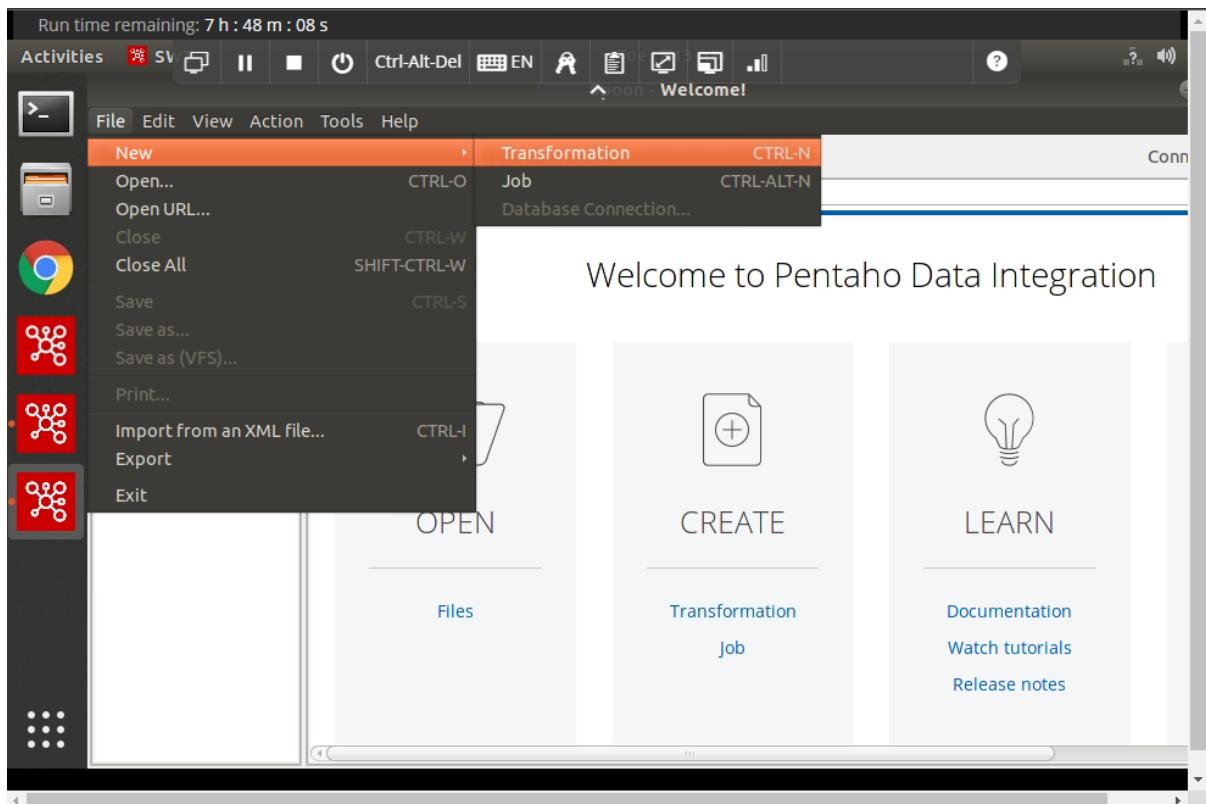
Data preparation is all about bringing data together, getting it all standardized and compliant with your company's policies, and then putting the data where it needs to go to make it most useful. In this course you're

Pentaho Data Integration: Evaluation

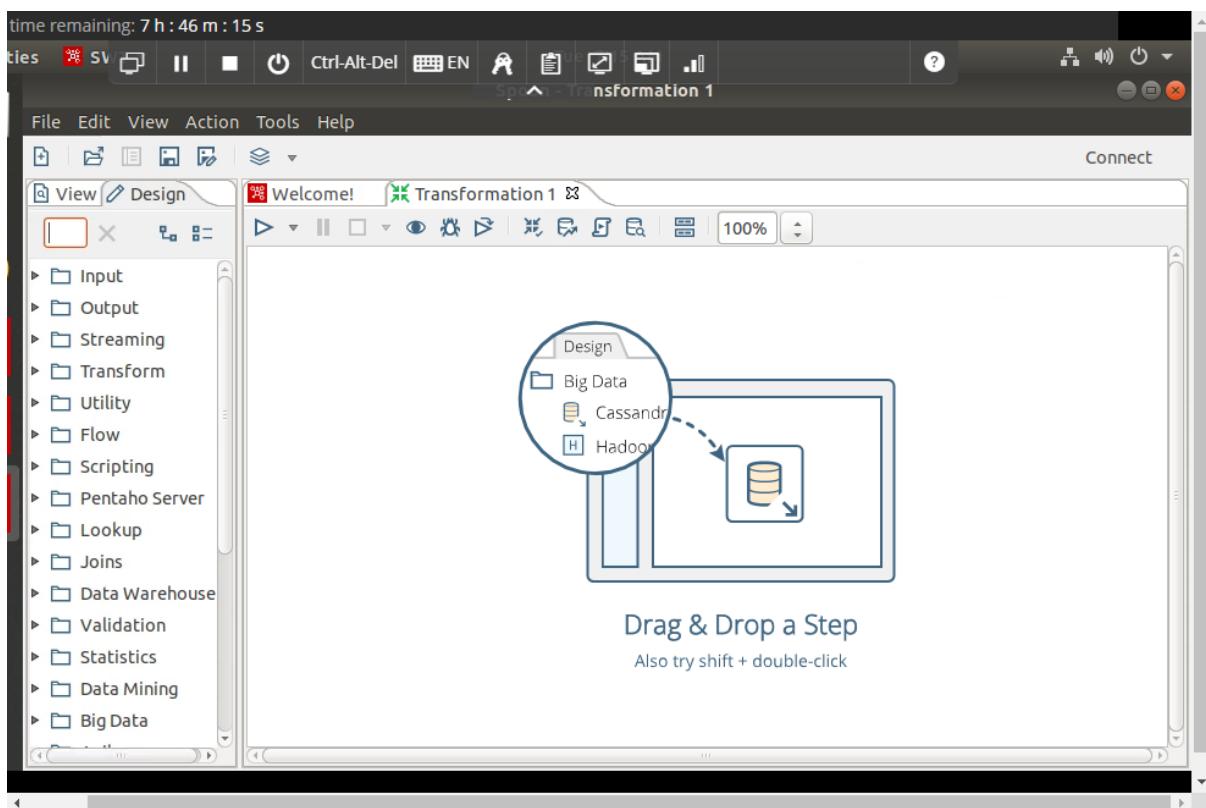
Langkah Pertama kita terlebih dahulu akan membaca inputan data , disini kita akan menggunakan data daftar customer. Sebelumnya kita dapat mengakses Spoon pada virtual machine yang sudah tersedia.



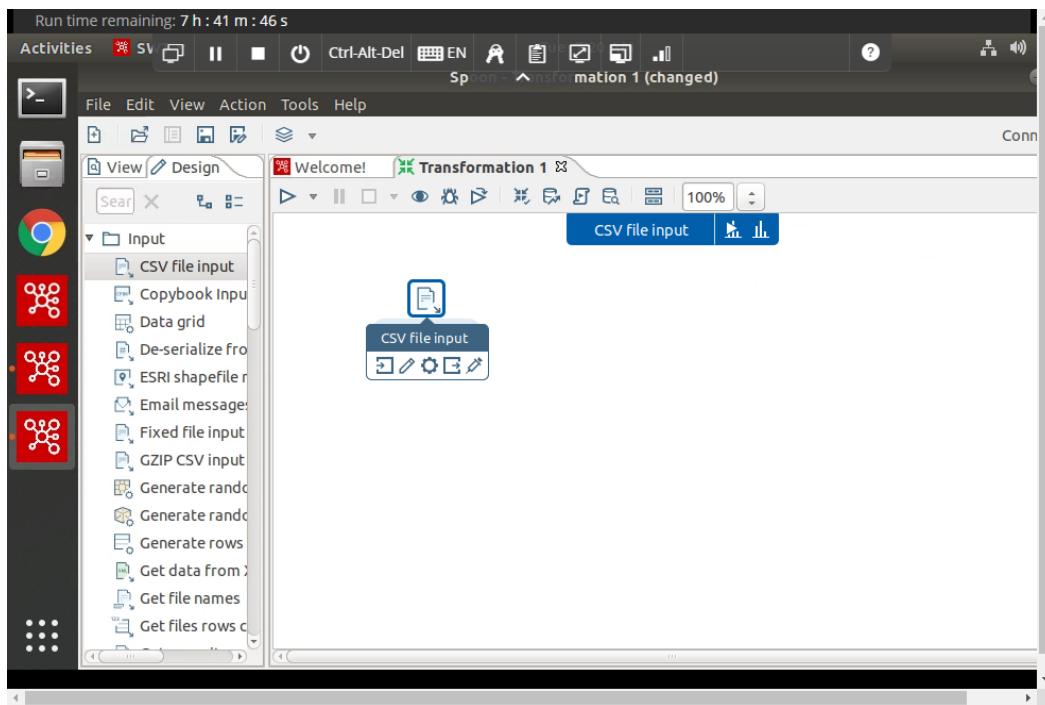
Aplikasi Spoon ini merupakan produk dari Hitachi yang bertujuan untuk menangani integrasi data, yang biasa dikenal sebagai Pentaho Data Integration. Selanjutnya kita akan melakukan transformasi data pelanggan dari data yang belum bersih menjadi data yang bersih dan data siap dipakai dalam database.



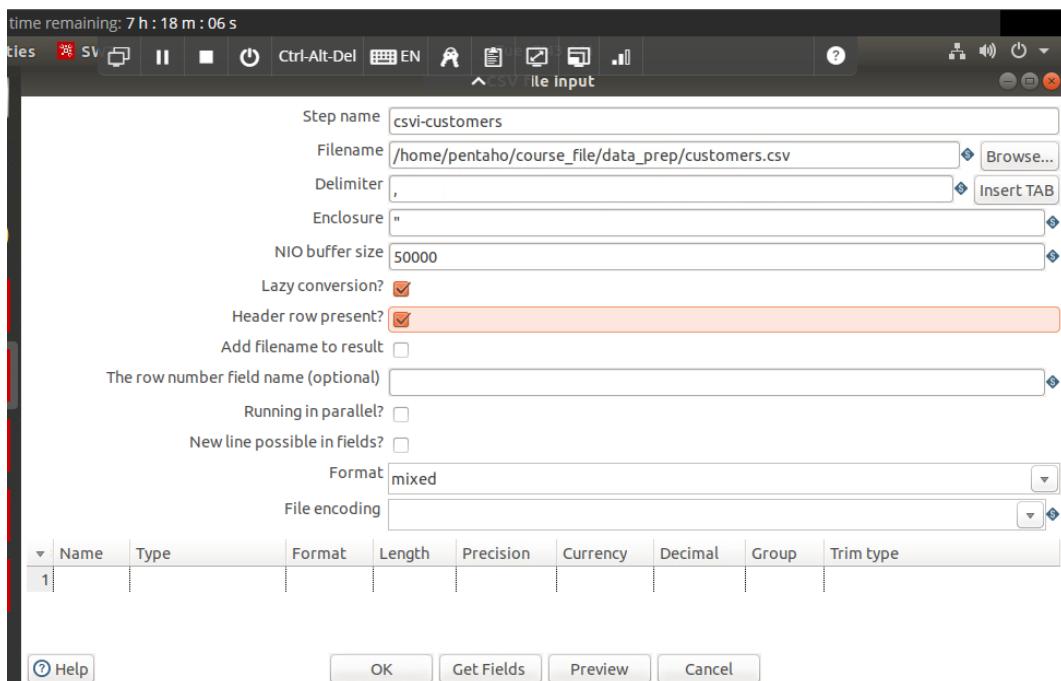
Akan muncul halaman baru untuk melakukan transformation data, halaman ini (Drag and Drop step) dikenal juga dengan canvas.



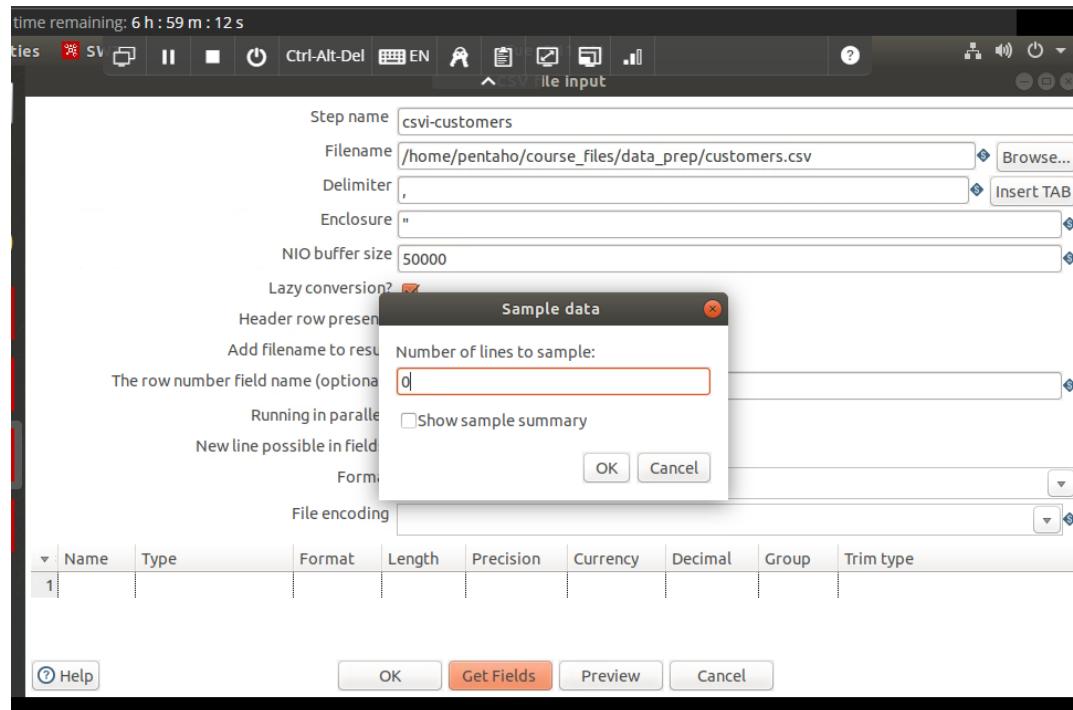
Selanjutnya kita membuat step untuk menginput customer dan dengan drag CSV file input pada halaman design canvas, double click pada step tersebut untuk dapat melakukan konfigurasi.



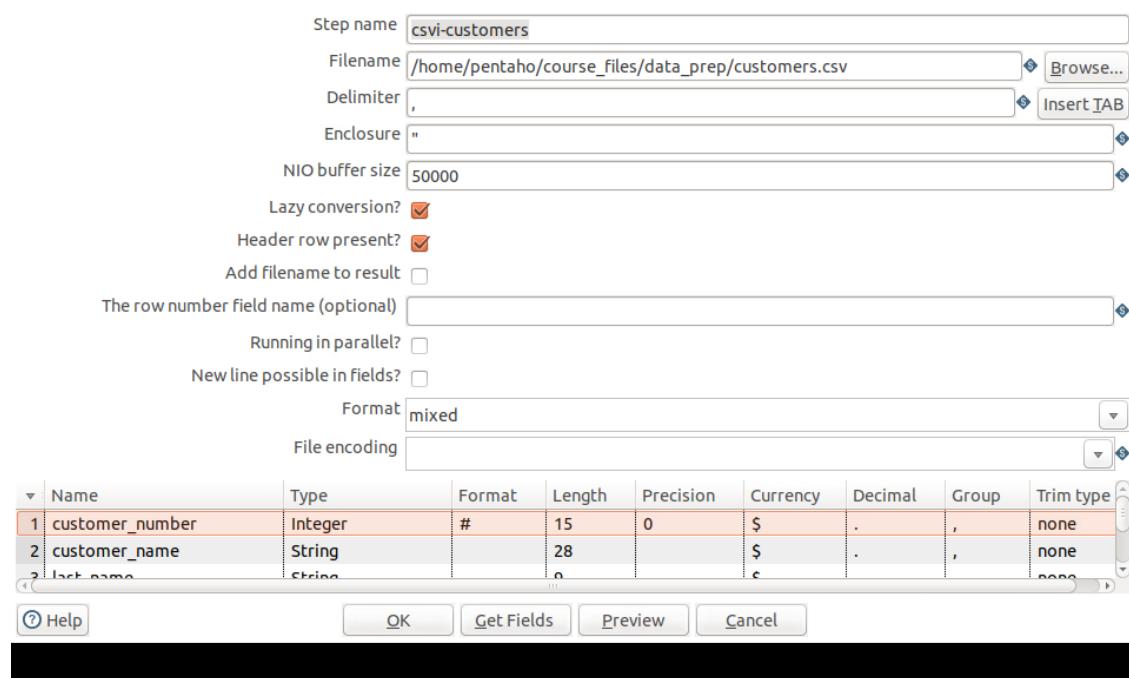
Konfigurasi yang dilakukan ini adalah step name, dengan filename sebagai sumber data, enclosure , delimiter dan sebagaimana yang sesuai untuk menginput data.



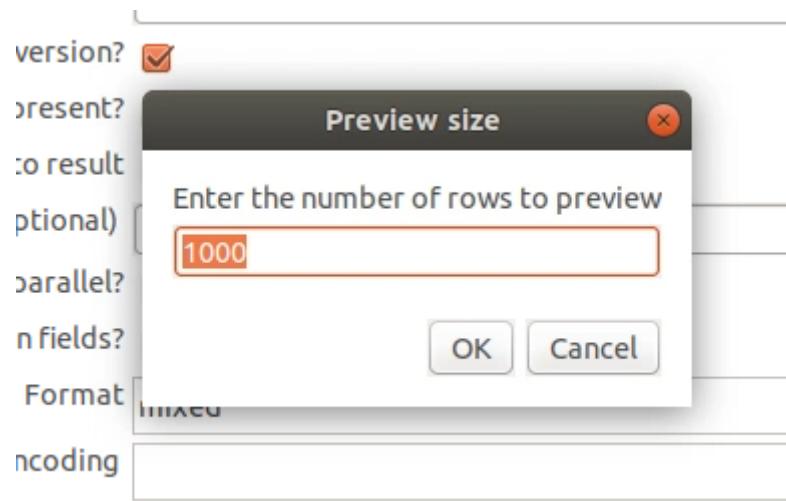
Selanjutnya klik bagian Get Fields untuk mendapatkan nama kolom yang berada pada input data tetapkan 0 dibagian Number of lines sample untuk mendapatkan nama kolom data secara keseluruhan.



Nama – nama kolom beserta propertinya akan muncul pada bagian bawah dialog konfigurasi step.



Agar dapat melihat contoh data yang ada dari file tersebut, klik priview pada bagian bawah dialog dan kita dapat memasukan preview size sesuai keinginan banyak data yang akan ditampilkan



Hasilnya muncul kotak dialog baru yang memuat data preview dari file input yang ditentukan. Dengan 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.

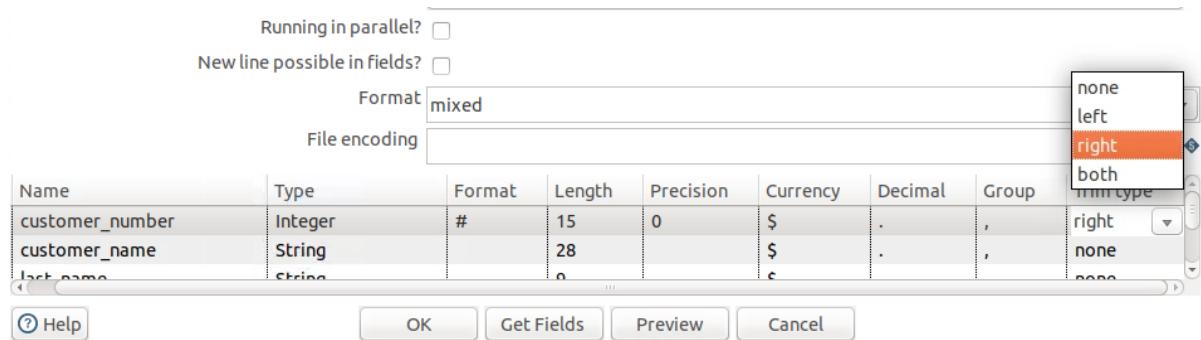
ties SWT tue 8:14 AM Examine preview data

Rows of step: csvi-customers (36 rows)

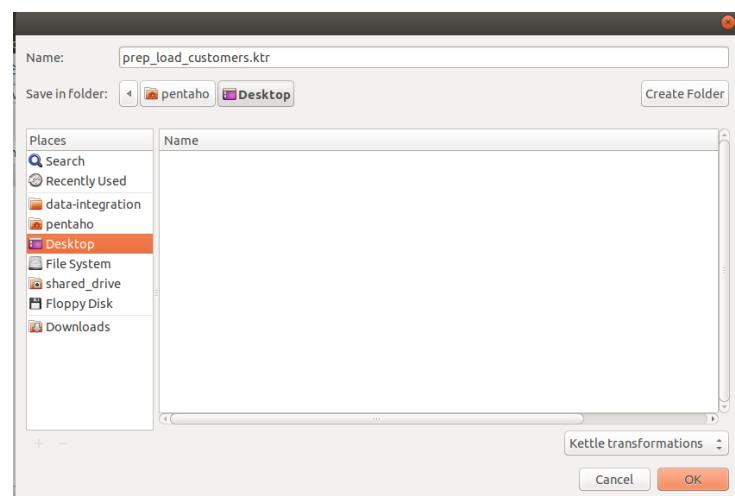
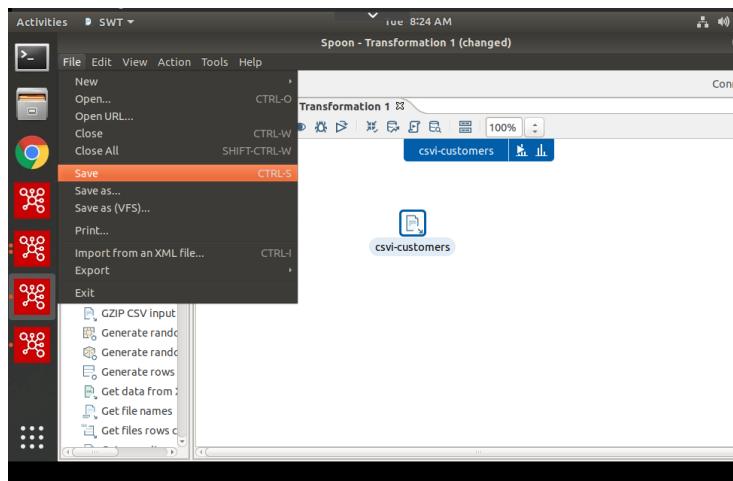
	customer_number	customer_name	last_name	first_name	address	city
1	112	Signal Gift Stores	King	Sue	245 Edgefield St.	Las Vegas
2	124	Mini Gifts Distributors Ltd.	Nelson	Valarie	982 Lilac Street	San Rafael
3	129	Mini Wheels Co.	Murphy	Julie	125 Augusta St.	San Francisco
4	131	Land of Toys Inc.	Yu	Kwai	9 South Summit Road	NYC
5	151	Muscle Machine Inc	Young	Jeff	9788 Union Avenue	NYC
6	157	Diecast Classics Inc.	Yu	Kyung	907 Windsor St.	Allentown
7	161	Technics Stores Inc.	Hirano	Juri	7473 Bishop Ave.	Burlingame
8	168	American Souvenirs Inc	Franco	Sue	541 East Halifax Street	New Haven
9	173	Cambridge Collectables Co.	Tseng	Kyung	271 South Moon Drive	Cambridge
10	175	Gift Depot Inc.	King	Julie	9920 Pennsylvania Lane	Bridgewater
11	181	Vitachrome Inc.	Frick	Michael	48 Tower Lane	NYC
12	198	Auto-Moto Classics Inc.	Taylor	Leslie	8132 East Riverside Street	Brickhaven
13	204	Online Mini Collectables	Barajas	Miguel	8093 Archer St	Brickhaven
14	205	Toys4GrownUps.com	Young	Julie	663 Blue Spring Lane	Pasadena
15	219	Boards & Toys Co.	Young	Leslie	377 Java Ave.	Glendale
16	239	Collectable Mini Designs Co.	Thompson	Valarie	341 South Lumber St.	San Diego
17	286	Marta's Replicas Co.	Hernandez	Marta	904 Meadowbrook Street	Cambridge
18	319	Mini Classics	Frick	Steve	429 Anchorage St.	White Plains

Close Show Log

Kita juga dapat melakukan pengaturan trim type pada kotak dialog konfigurasi agar melakukan white-space yang tidak perlu.



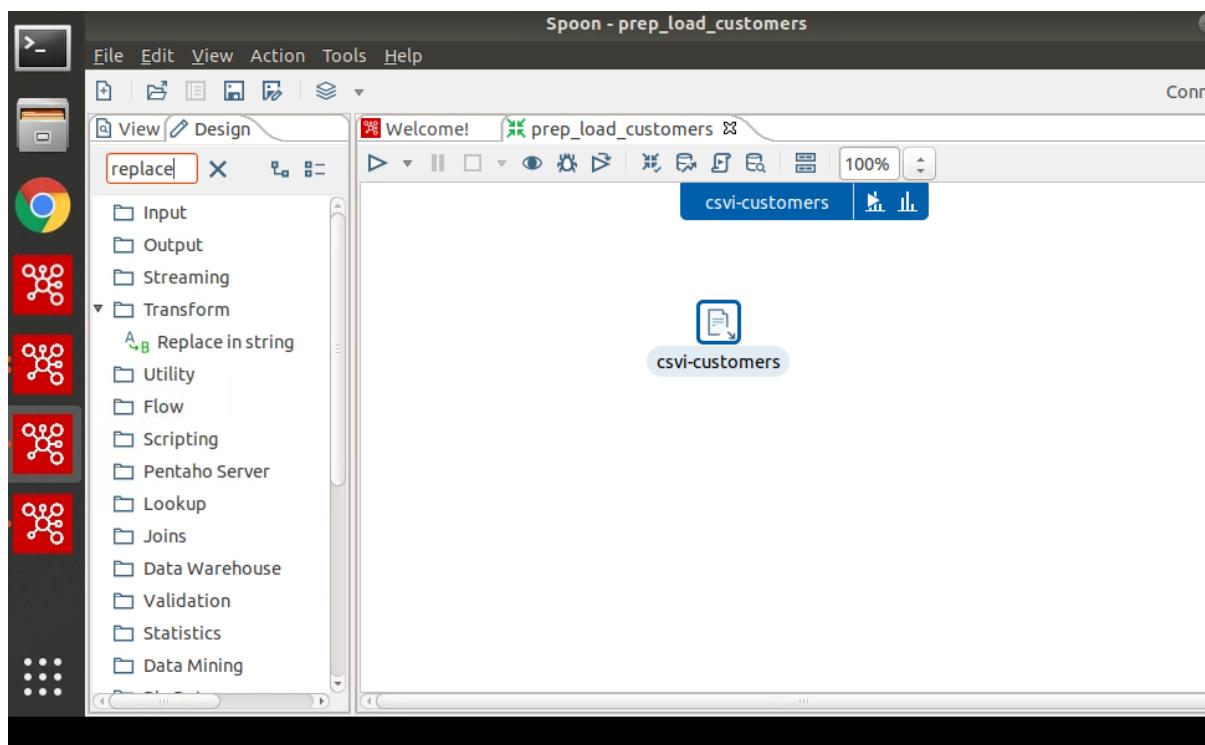
Etelah kita menyelesaikan konfigurasi pada file input klik button OK. Selanjutnya dengan menyimpan perubahan yang kita lakukan pada canvas pada halaman file pilih save dan simpan di lokasi yang diinginkan.

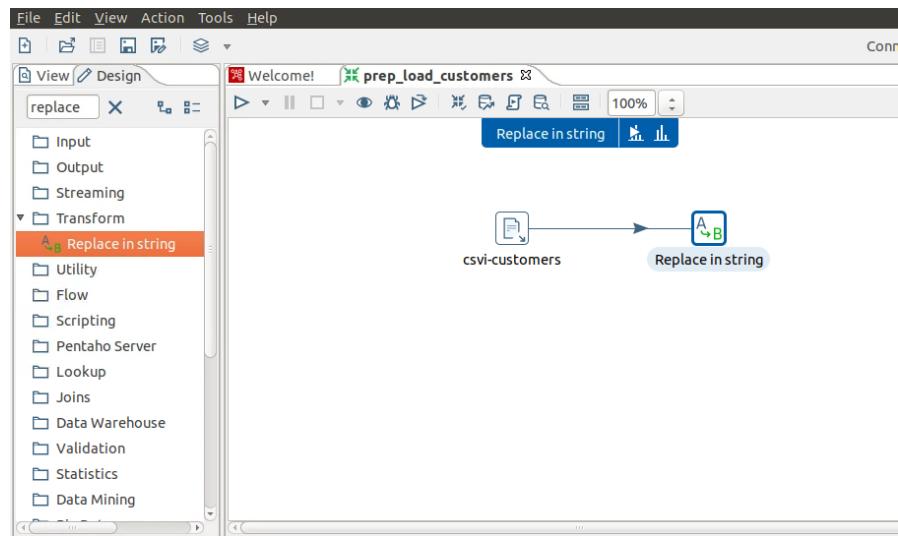


Filter Customers

(Filter Rows Step)

Langkah selanjutnya kita akan melakukan standarisasi format untuk kolom, sebelum ini kita mencari terlebih dahulu step replace in string pada tab design. Untuk dapat menghubungkan kedua step ini dapat klik step-csvi dan clik pada step replace in string .

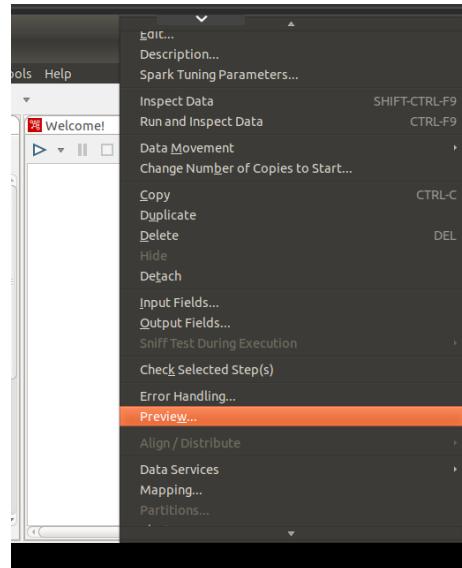


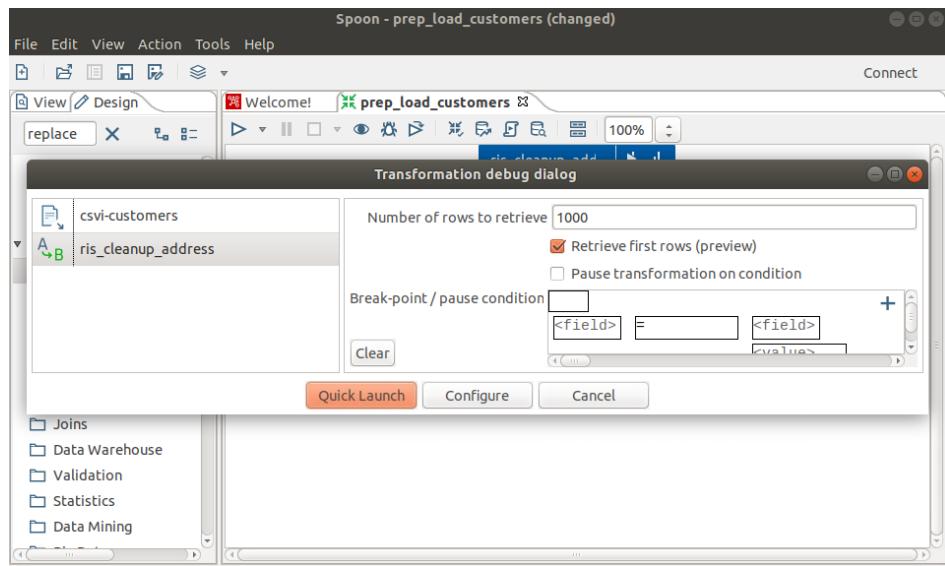


Setelah itu kita dapat melanjutkan standarisasi yang diterapkan berupa menghapus tanda titik di kolom address and state dan mengubah singkatan St dan Ave menjadi Street dan Avenue kemudian klik OK.

Fields string						
In stream field	Out stream field	use RegEx	Search	Replace with	Set empty string?	Replace with field
1 address			.	Street		
2 address			St			
3 address			Ave	Avenue		
4 state			.			

Klik kanan pada step tersebut dan pilih preview kemudian kita akan melukukan konfigurasi pada transformation debug dialog.





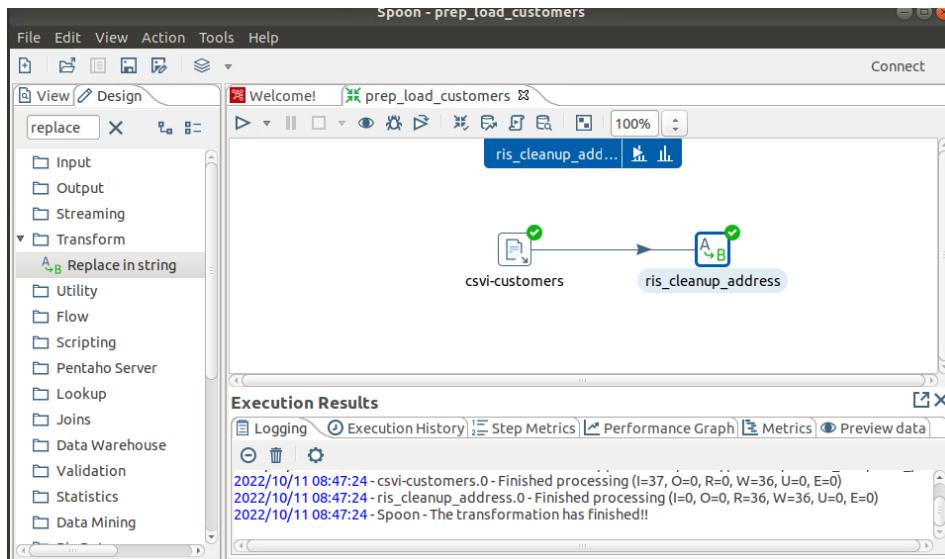
Maka kita dapat melihat hasil standarisasi data customer pada table dialog baru , simpan proges ini dengan klik ikon save pada bagian kiri layar.

Examine preview data

Rows of step: ris_cleanup_address (36 rows)

	customer_number	customer_name	last_name	first_name	address	city
1	112	Signal Gift Stores	King	Sue	245 Edgefield Street	Las Vegas
2	124	Mini Gifts Distributors Ltd.	Nelson	Valarie	982 Lilac Street	San Rafael
3	129	Mini Wheels Co.	Murphy	Julie	125 Augusta Street	San Francisco
4	131	Land of Toys Inc.	Yu	Kwai	9 South Summit Road	NYC
5	151	Muscle Machine Inc	Young	Jeff	9788 Union Avenue	NYC
6	157	Diecast Classics Inc.	Yu	Kyung	907 Windsor Street	Allentown
7	161	Technics Stores Inc.	Hirano	Juri	7473 Bishop Avenue	Burlingame
8	168	American Souvenirs Inc	Franco	Sue	541 East Halifax Street	New Haven
9	173	Cambridge Collectables Co.	Tseng	Kyung	271 South Moon Drive	Cambridge
10	175	Gift Depot Inc.	King	Julie	9920 Pennsylvania Lane	Bridgewater
11	181	Vitachrome Inc.	Frick	Michael	48 Tower Lane	NYC
12	198	Auto-Moto Classics Inc.	Taylor	Leslie	8132 East Riverside Street	Brickhaven
13	204	Online Mini Collectables	Barajas	Miguel	8093 Archer Street	Brickhaven
14	205	Toys4GrownUps.com	Young	Julie	663 Blue Spring Lane	Pasadena
15	219	Boards & Toys Co.	Young	Leslie	377 Java Avenue	Glendale
16	239	Collectable Mini Designs Co.	Thompson	Valarie	341 South Lumber Street	San Diego
17	286	Marta's Replicas Co.	Hernandez	Marta	904 Meadowbrook Street	Cambridge
18	319	Mini Classics	Frick	Steve	429 Anchorage Street	White Plains

Close



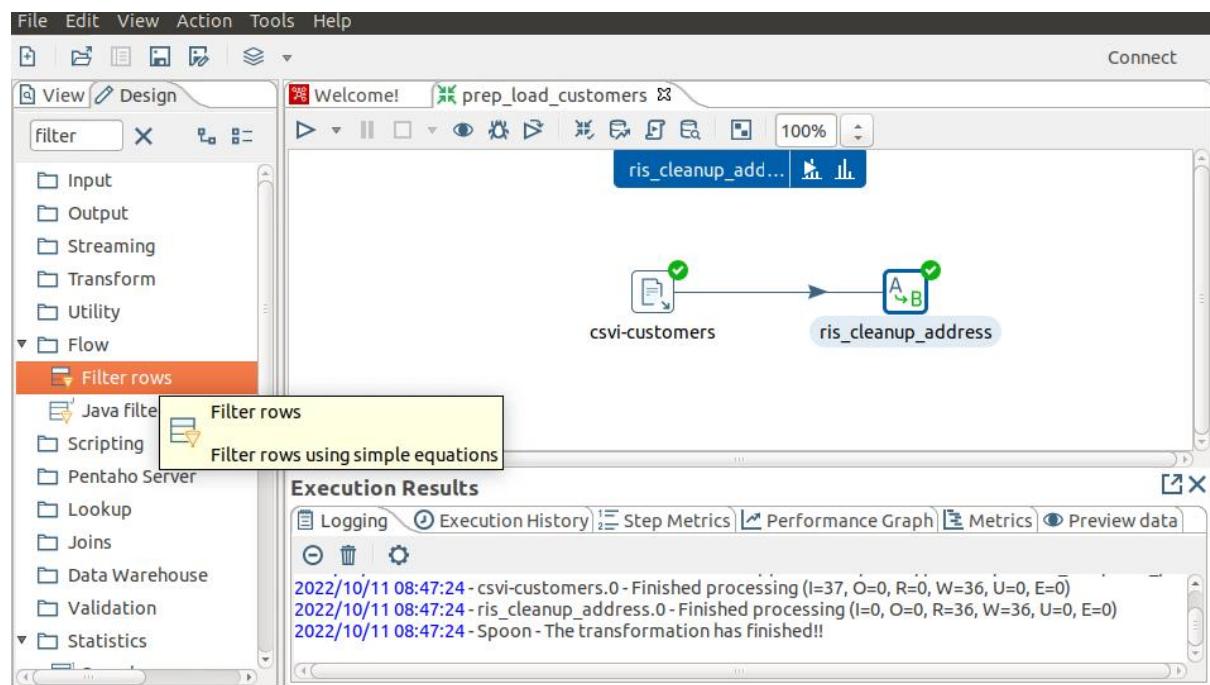
Data Preparation Scenario

HITACHI
Inspire the Next

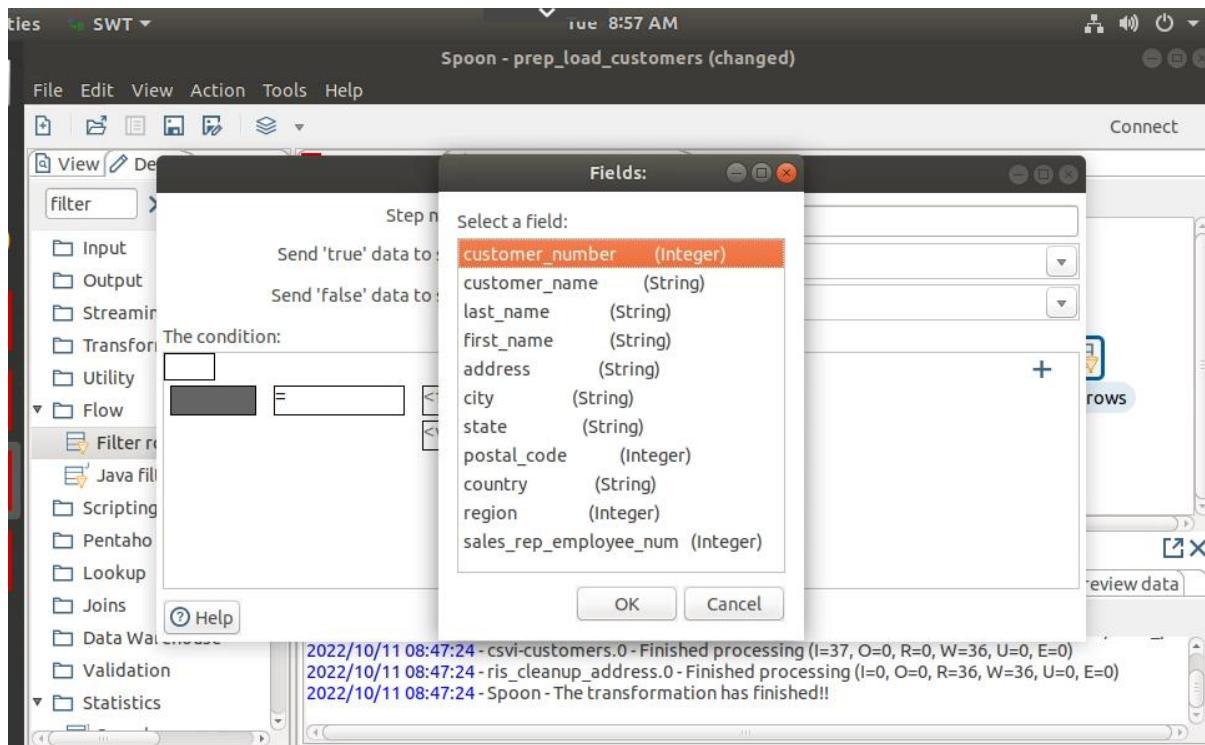
Filter Customers

(Filter Rows Step)

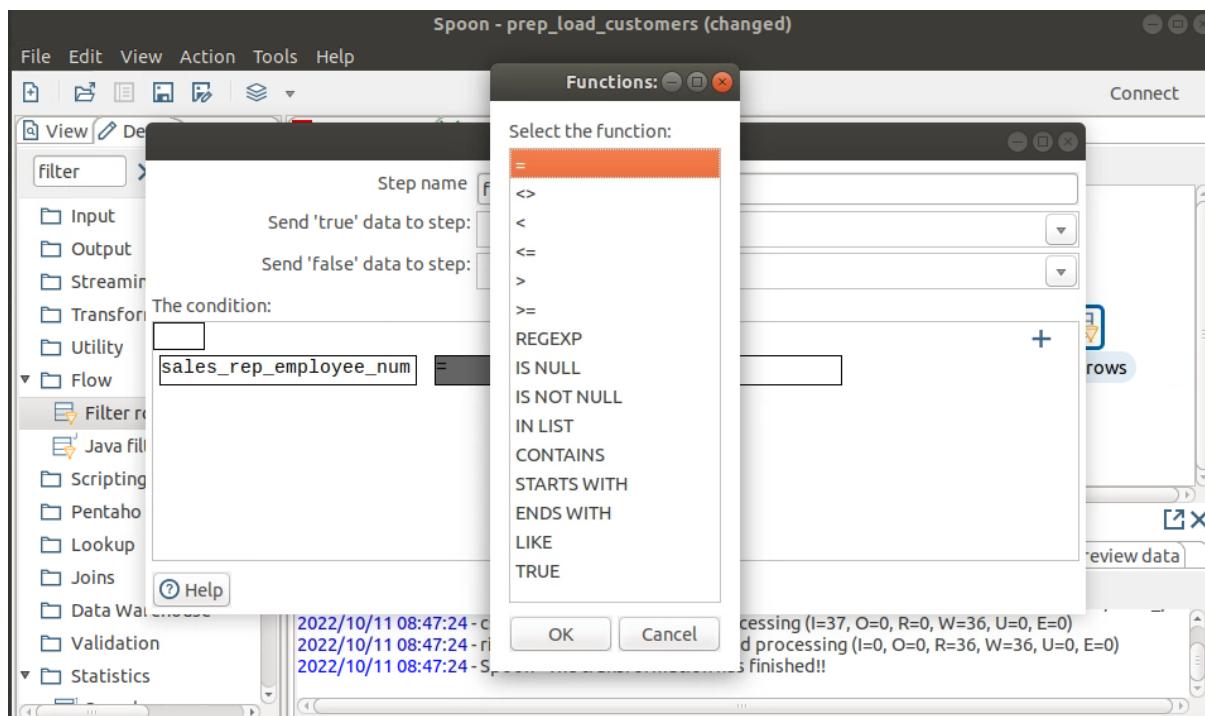
Selanjutnya melakukan filtrasi terhadap data terstandarisasi. Pada contoh penerapan kali ini kita akan memfiltrasi kolom sales_rep_employee_num.

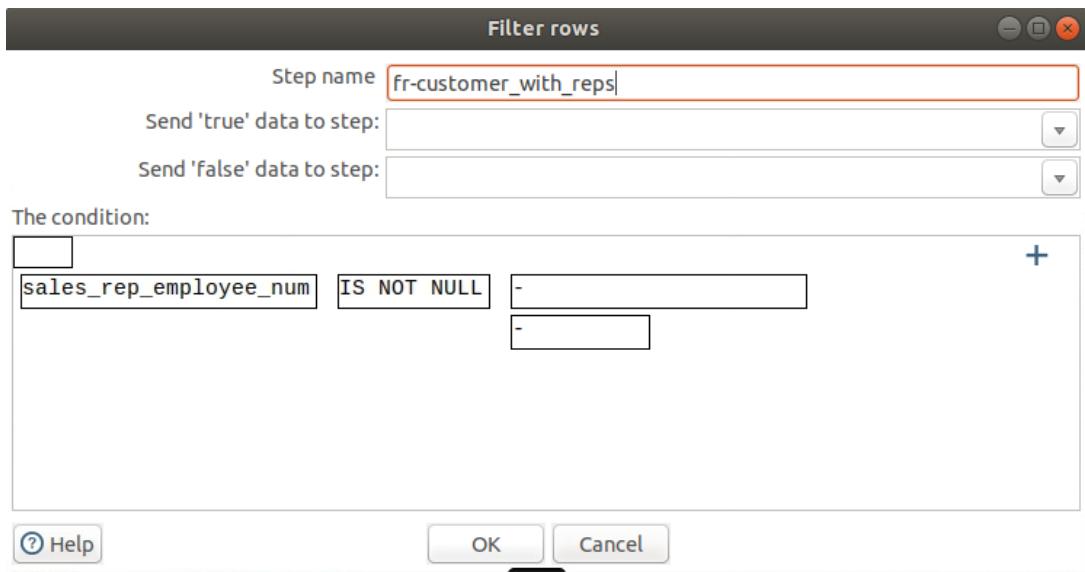


Pada bagian kotak pertama pilih sales_rep_employee_num

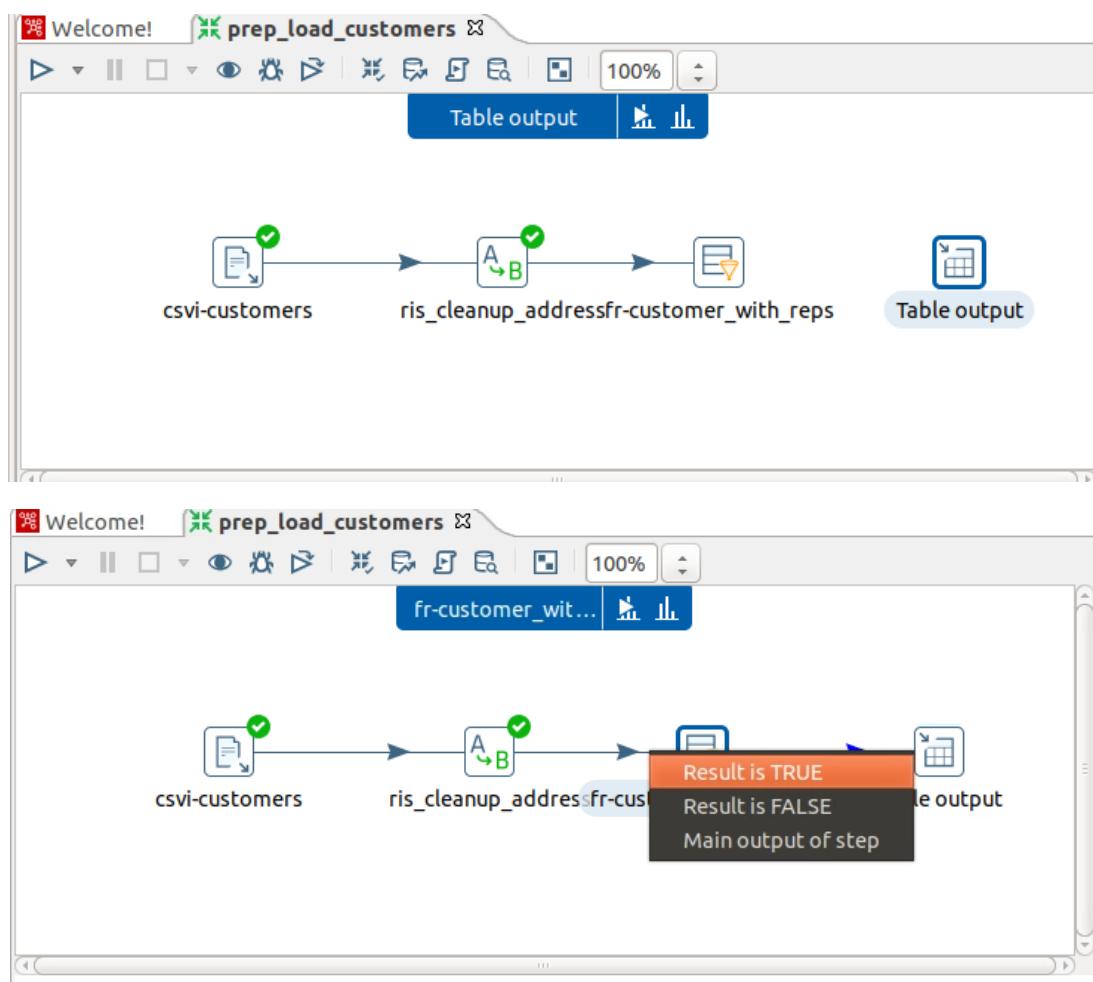


Pada kotak kedua pilih IS NOT NULL

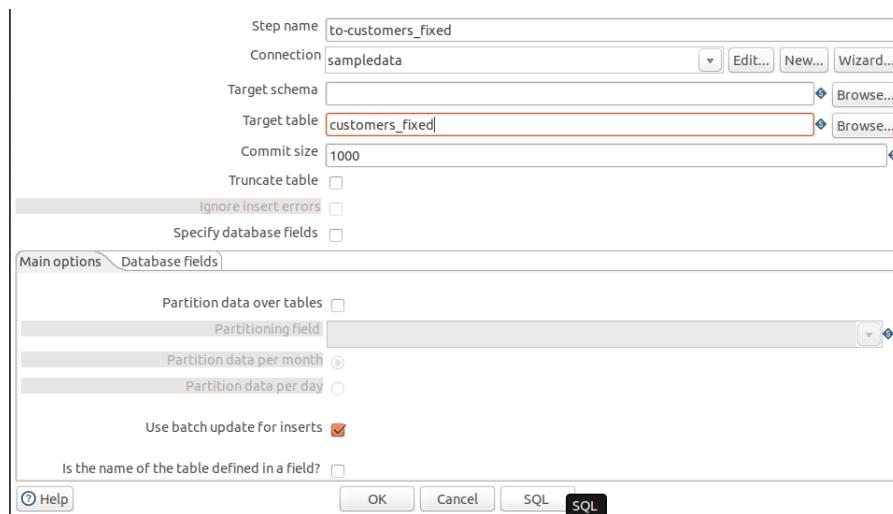




Klik OK, kemudian kita akan melanjutkan untuk penanganan kondisi true untuk filterisasi.



double-click pada Table output untuk memulai konfigurasi. Pada dialog konfigurasi ini, kita juga dapat melihat SQL yang tercipta dengan click pada tombol SQL



The screenshot shows two windows from the Talend interface. On the left, the 'Simple SQL editor' window displays the SQL code for creating the 'customers_fixed' table:

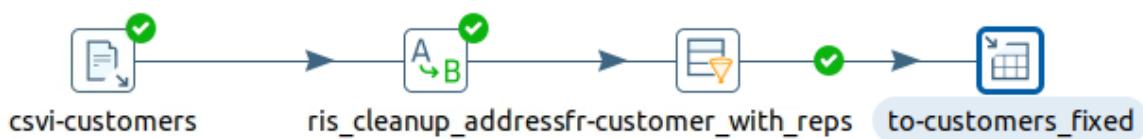
```
SQL statements, separated by semicolon';'
CREATE TABLE customers_fixed
(
    customer_number BIGINT
, customer_name VARCHAR(28)
, last_name VARCHAR(9)
, first_name VARCHAR(9)
, address VARCHAR(26)
, city VARCHAR(13)
, state VARCHAR(10)
, postal_code BIGINT
, country VARCHAR(3)
, region BIGINT
, sales_rep_employee_num BIGINT
)
;'
```

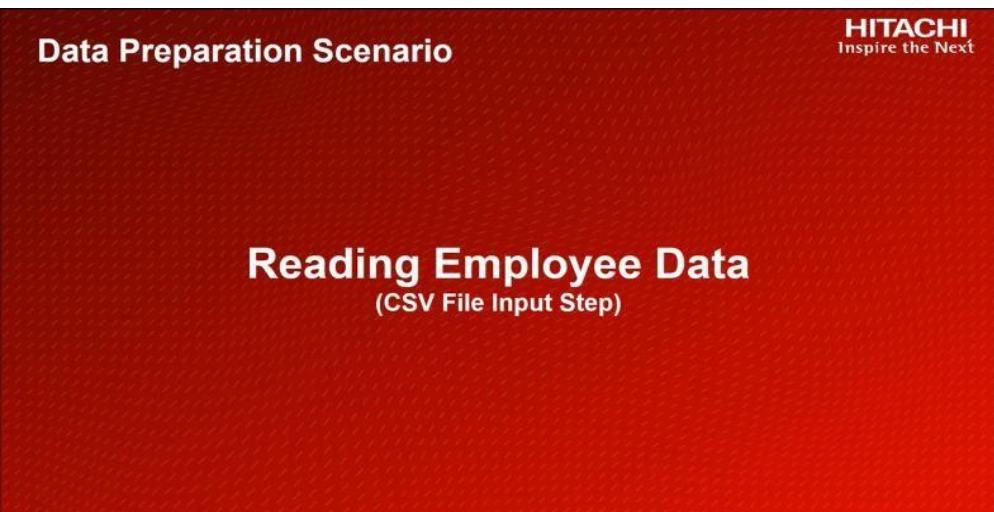
On the right, the 'Results of the SQL statements' window shows the executed SQL statement and its results:

```
The SQL statements had the following results
SQL executed: CREATE TABLE customers_fixed
(
    customer_number BIGINT
, customer_name VARCHAR(28)
, last_name VARCHAR(9)
, first_name VARCHAR(9)
, address VARCHAR(26)
, city VARCHAR(13)
, state VARCHAR(10)
, postal_code BIGINT
, country VARCHAR(3)
, region BIGINT
, sales_rep_employee_num BIGINT
)

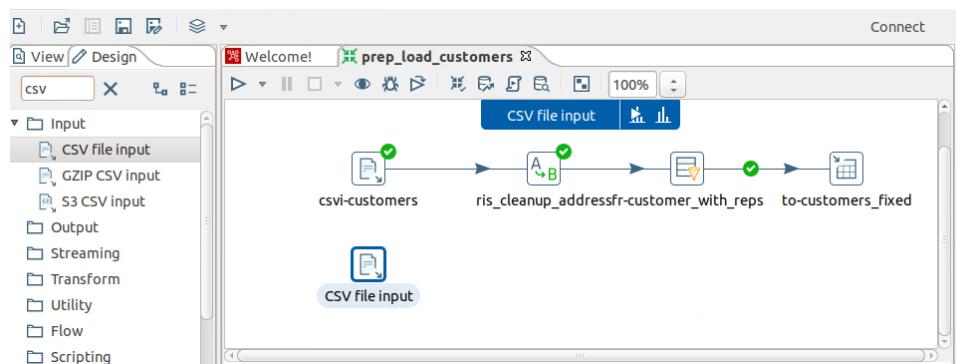
1 SQL statements executed
```

Tabel dan struktur yang terbentuk dapat dilihat pada penerapan selanjutnya. Klik ikon save untuk menyimpan progress.





Pada tahap ini kita akan menginput employee data pada filtrasu bernilai null. Caranya sama dengan step CSV file input .



Klik pada step ini dan get fields dan preview pada data employee ini.

CSV file input

Step name: csvi-employees

Filename: /home/pentaho/course_files/data_prep/employees.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 employee_number	Integer	#	15	0	\$.	,	none
2 last_name	String		9		\$.	,	none
3 first_name	String		8		\$.	,	none

OK Get Fields Preview Cancel

Examine preview data

Rows of step: csvi-employees (23 rows)

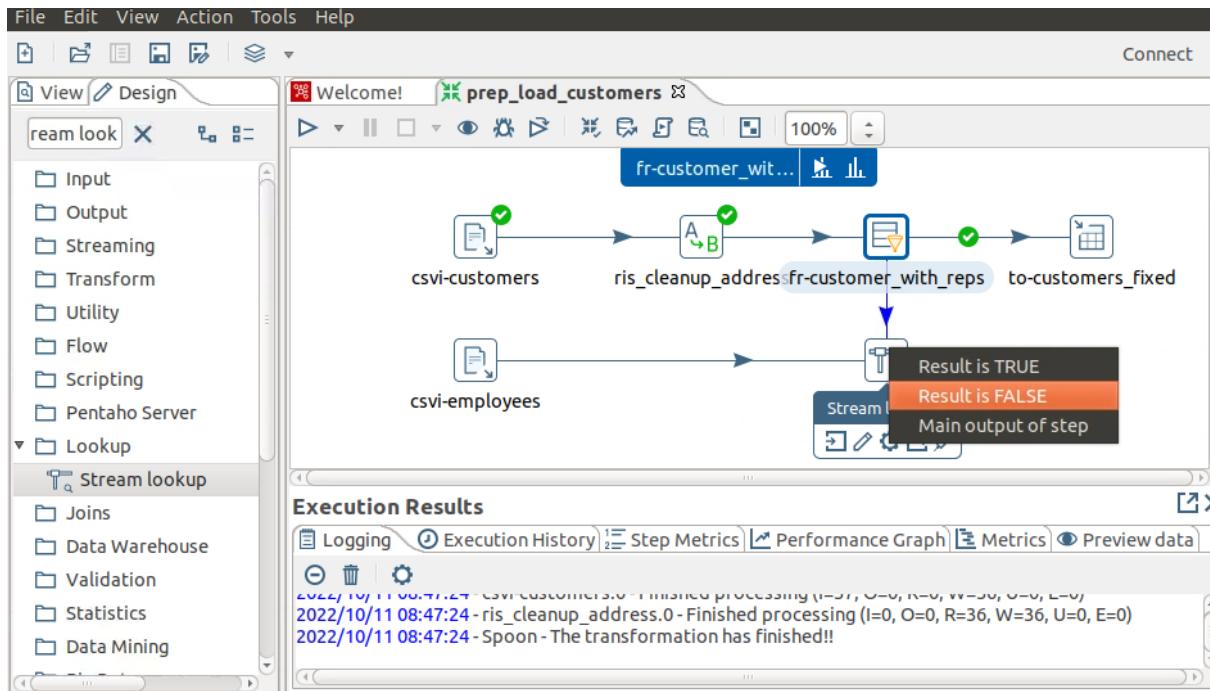
	employee_number	last_name	first_name	email	region	job_title
1	1002	Murphy	Diane	dmurphy@classicmodelcars.com	<null>	President
2	1056	Patterson	Mary	mpatterso@classicmodelcars.com	<null>	VP Sales
3	1076	Firrelli	Jeff	jfirrelli@classicmodelcars.com	<null>	VP Marketing
4	1088	Patterson	William	wpatterson@classicmodelcars.com	<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	jfirrelli@classicmodelcars.com	4	Sales Rep
10	1216	Patterson	Steve	spatterso@classicmodelcars.com	8	Sales Rep
11	1286	Tseng	Foon Yue	ftseng@classicmodelcars.com	14	Sales Rep
12	1323	Vanauf	George	gvanauf@classicmodelcars.com	3	Sales Rep
13	1337	Bondur	Loui	lbondur@classicmodelcars.com	17	Sales Rep
14	1370	Hernandez	Gerard	ghernande@classicmodelcars.com	9	Sales Rep
15	1401	Castillo	Pamela	pcastillo@classicmodelcars.com	13	Sales Rep
16	1501	Bott	Larry	lbott@classicmodelcars.com	16	Sales Rep
17	1504	Jones	Barry	bjones@classicmodelcars.com	7	Sales Rep
18	1611	Fixter	Andy	afixter@classicmodelcars.com	11	Sales Rep

Data Preparation Scenario

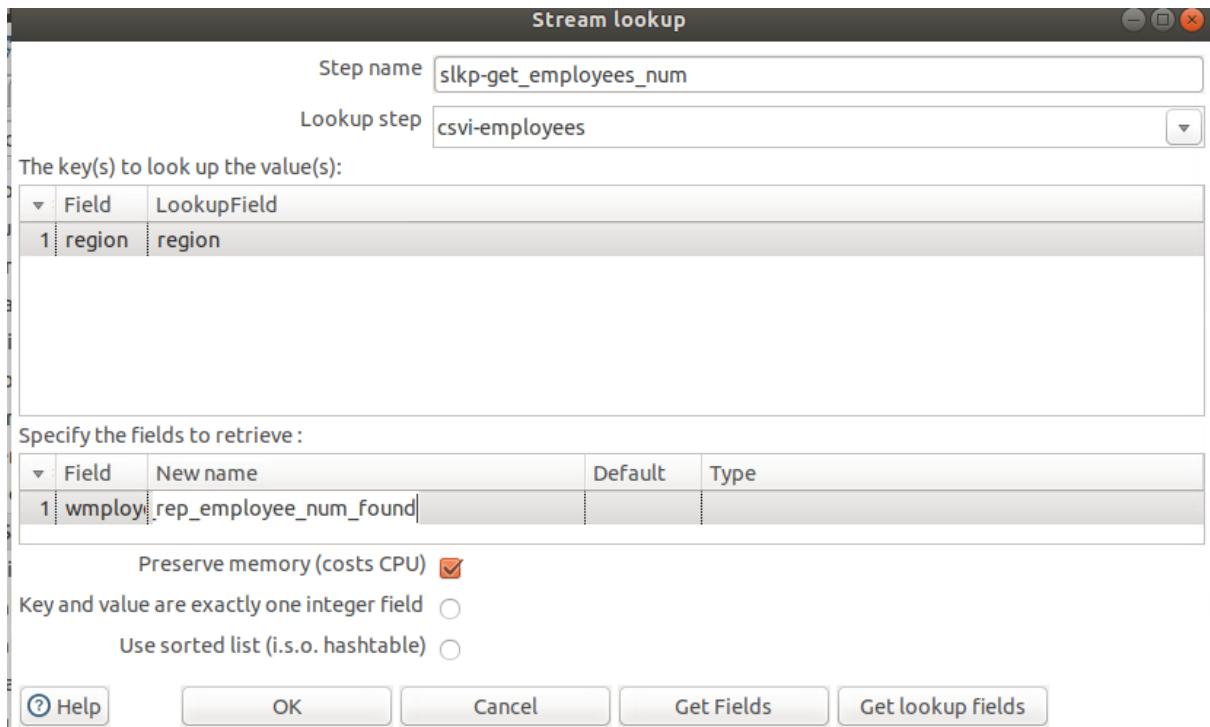
HITACHI
Inspire the Next

Lookup Employee Number (Stream Lookup Step)

Pada tahap ini kita mencari step Stream lookup pada tab design dan masukkan kedalam canvas dan hubungkan csvi-employees dengan step ini.



click pada step dan lakukan beberapa konfigurasi. Pada step ini, data akan mengacu pada kolom region sebagai lookup value untuk data employee dan Juga menetapkan employee_number dari data employee sebagai sales_rep_employee_num_found

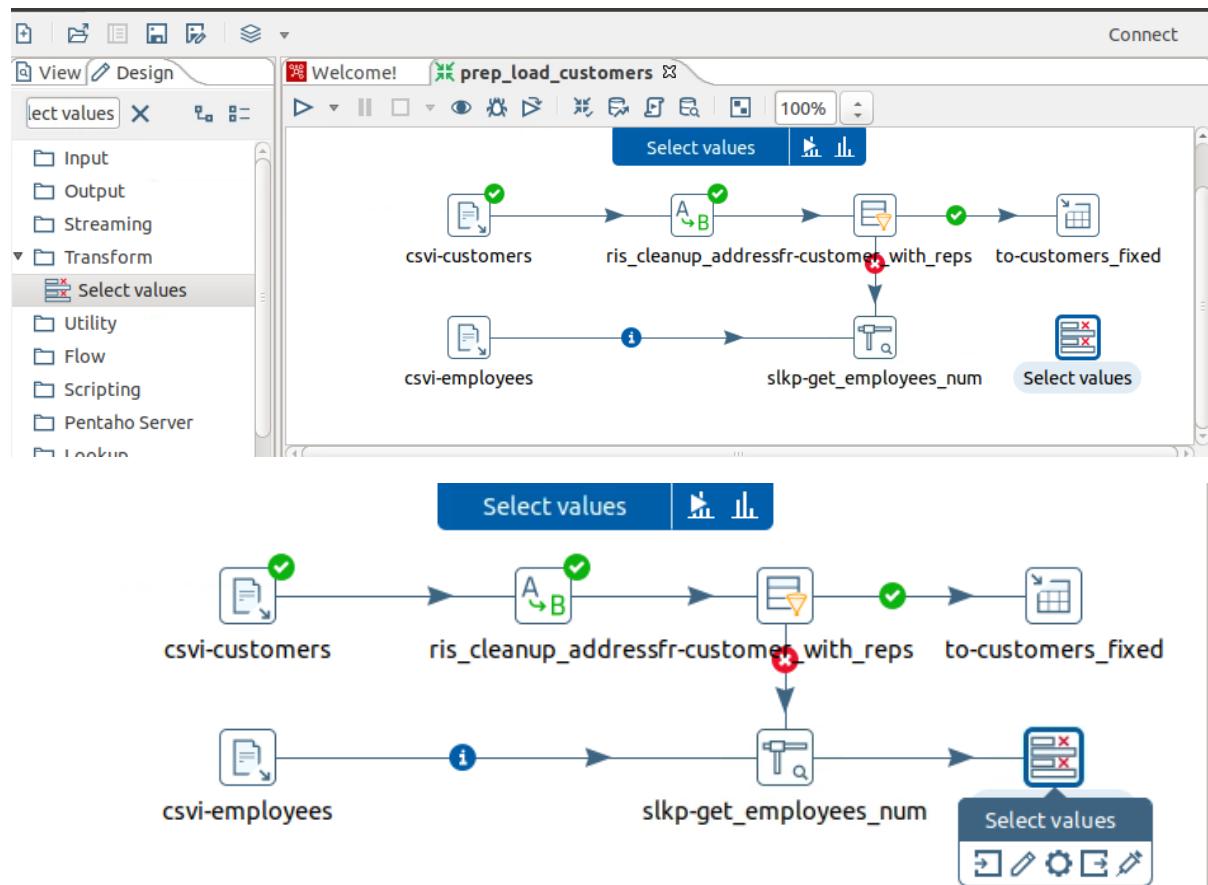


Data Preparation Scenario

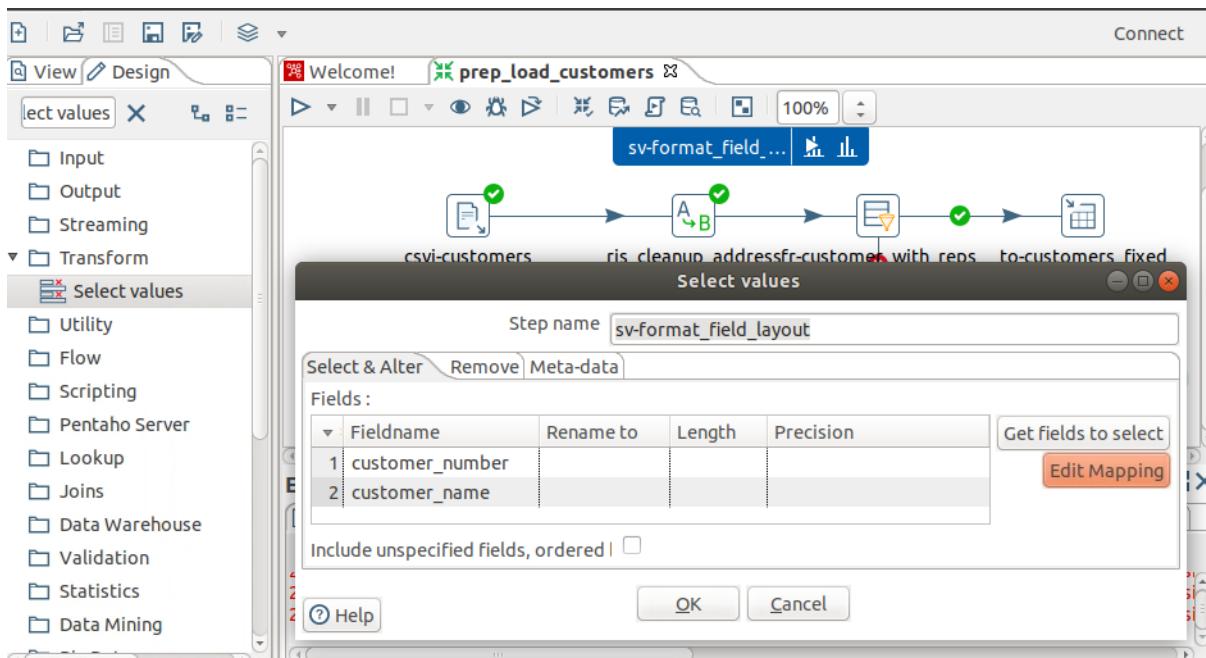
HITACHI
Inspire the Next

Format Field Layout (Select Values Step)

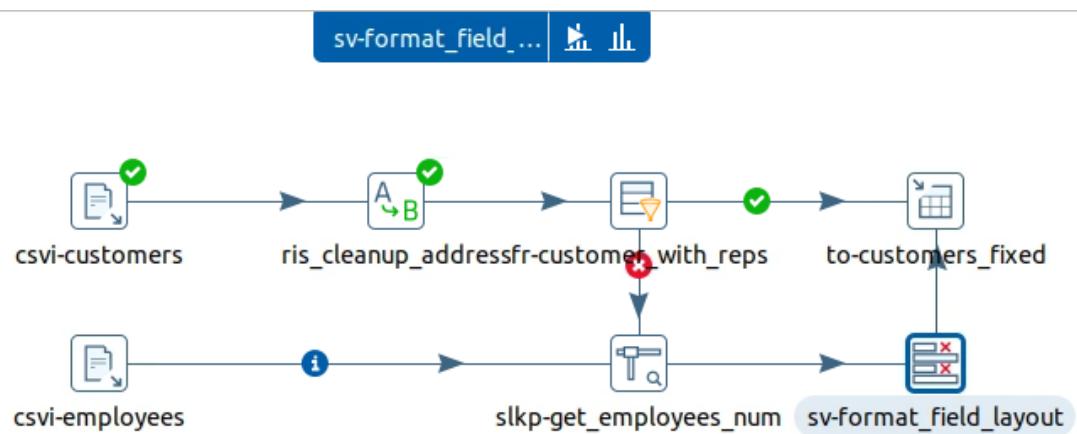
step Select Values pada tab Design dan drag pada canvas, serta hubungkan step stream lookup



Klik step untuk memulai konfigurasi



Susunan akhir canvas transformation:



Running Your Transformation



Untuk menjalankan hasil dari transformasi ini kita dapat klik ikon play pada bagian atas kiri canvas maka akan muncul kotak dialog Run Options.

The screenshot shows the Pentaho Spoon application interface. On the left, there is a sidebar with various transformation components like Input, Output, Streaming, Transform, and Select values. The 'Select values' component is currently selected. In the center, there is a main canvas area with a transformation diagram. A 'Run Options' dialog box is open over the canvas. The dialog has a title 'Run Options' and a sub-section 'Run configuration:' with a dropdown set to 'Pentaho local'. Below this are 'Options' settings: 'Clear log before running' (checked), 'Log level:' dropdown set to 'Basic', 'Enable safe mode' (unchecked), and 'Gather performance metrics' (checked). At the bottom of the dialog are tabs for 'Parameters' and 'Variables', both of which are currently empty. In the bottom right corner of the main window, there is a section titled 'Execution Results' with a list of log entries:

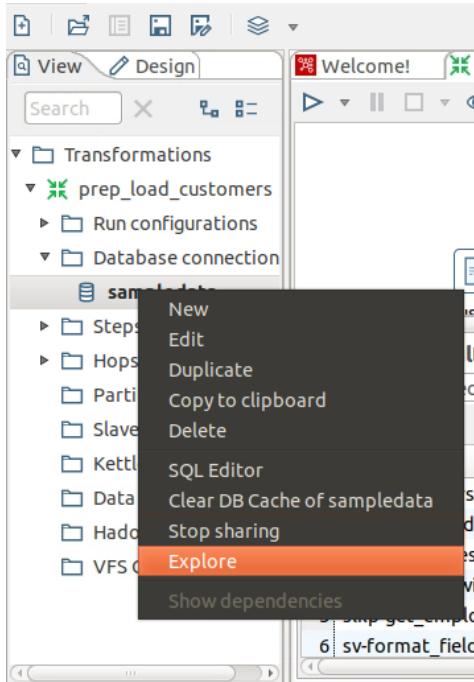
- 2022/10/11 10:25:15 - Spoon - Running transformation using the Kettle execution engine
- 2022/10/11 10:25:15 - Spoon - Transformation opened.
- 2022/10/11 10:25:15 - Spoon - Launching transformation [prep_load_customers]...
- 2022/10/11 10:25:15 - Spoon - Started the transformation execution.
- 2022/10/11 10:25:15 - prep_load_customers - Dispatching started for transformation [prep_load_customers]

Step Metrics untuk menunjukkan pemrosesan data pada tiap step

Execution Results

Stepname	Copynr	Read	Written	Input	Output	Updated
1 csvi-customers	0	0	0	0	0	0
2 ris_cleanup_address	0	0	0	0	0	0
3 csvi-employees	0	0	0	0	0	0
4 fr-customer_with_reps	0	0	0	0	0	0
5 slkp-get_employees_num	0	0	0	0	0	0
6 sv-format_field_layout	0	0	0	0	0	0

Untuk melihat hasil transformation pada table, pilih tab View dan expand folder Transformations, prep_load_customers



Hasil On Experience untuk pentaho data integration: Evaluation.

Rows of step: CUSTOMERS_FIXED (36 rows)

IMER_NAME	LAST_NAME	FIRST_NAME	ADDRESS	CITY	STATE	POSTAL_CODE	COUNTRY	REGION	SALES_REP_EMPLOYEE_NUM
Gift Stores	King	Sue	245 Edgefield Street	Las Vegas	NV	83030	USA	14	1166
if Toys Inc.	Yu	Kwai	9 South Summit Road	NYC	NY	10022	USA	10	1323
t Classics Inc.	Yu	Kyung	907 Windsor Street	Allentown	PA	70267	USA	4	1216
cs Stores Inc.	Hirano	Juri	7473 Bishop Avenue	Burlingame	CA	94217	USA	16	1165
can Souvenirs Inc	Franco	Sue	541 East Halifax Street	New Haven	CT	97823	USA	13	1286
idge Collectables Co.	Tseng	Kyung	271 South Moon Drive	Cambridge	MA	51247	USA	5	1188
spot Inc.	King	Julie	9920 Pennsylvania Lane	Bridgewater	CT	97562	USA	3	1323
Auto Classics Inc.	Taylor	Leslie	8132 East Riverside Street	Brickhaven	MA	58339	USA	2	1216
Mini Collectables	Barajas	Miguel	8093 Archer Street	Brickhaven	MA	58339	USA	2	1188
rownUps.com	Young	Julie	663 Blue Spring Lane	Pasadena	CA	90003	USA	16	1166
's Replicas Co.	Hernandez	Marta	904 Meadowbrook Street	Cambridge	MA	51247	USA	5	1216
lassics	Frick	Steve	429 Anchorage Street	White Plains	NY	24067	USA	11	1323
reations Ltd.	Tam	Wing C	357 John Drive	New Bedford	MA	50553	USA	7	1188
rate Gift Ideas Co.	Brown	Julie	908 Gray Street	San Francisco	CA	94217	USA	15	1165
Collectables Inc.	Brown	William	266 Court Drive	Newark	NJ	94019	USA	9	1323
: Gift Ideas, Inc.	Cervantes	Francisca	798 Pinnacle Avenue	Philadelphia	PA	71270	USA	12	1188
't US Retailers, Ltd.	Chandler	Michael	529 Carriage Street	Los Angeles	CA	91003	USA	16	1166
AllAges.com	Yoshido	Juri	222 Green Lake Street	Boston	MA	51003	USA	1	1216
Diecast Creations Co.	Young	Valarie	7987 Manchester Street	Nashua	NH	62005	USA	8	1216
Friends Inc	Hernandez	Maria	8847 West Oceanview Street	NYC	NY	10027	USA	10	1286

The screenshot shows a course completion summary for 'Getting Started'. At the top, it says '100% COMPLETE' with a progress bar. Below that is a list of completed modules: 'Course Introduction', 'Data Preparation Scenario', 'Your First Look at PDI', 'Take a Peak at the Solution', and 'Hands-On: Building Your Transformation'. Each module has a red checkmark next to it. At the bottom, there's a copyright notice: '©2021 HITACHI VANTARA LLC. ALL RIGHTS RESERVED.'

Overachiever?

If you want to do even more on your own, consider using your lab environment to enhance your transformation. Here are some things you could do:

- Add notes on the canvas to tell others what the transformation does and how it works
- Change the transformation so that it outputs the fixed customers to a text file instead of a database table
- Browse the list of available steps and maybe even try out a few. Some good ones are Sort Rows, Group By, and Calculator. If you know how to code, check out JavaScript, Java Expression, or even the R Script Executor steps.
- Create a job that will run the transformation

Congratulations!

You've completed PDI Evaluation: Getting Started.

~ Terima Kasih ~