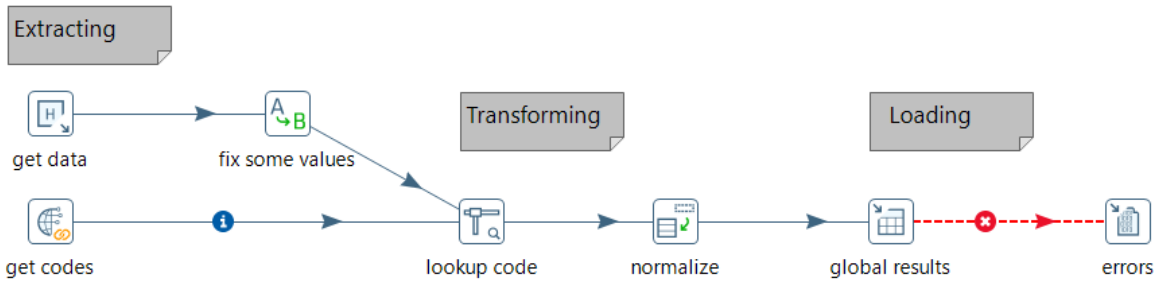
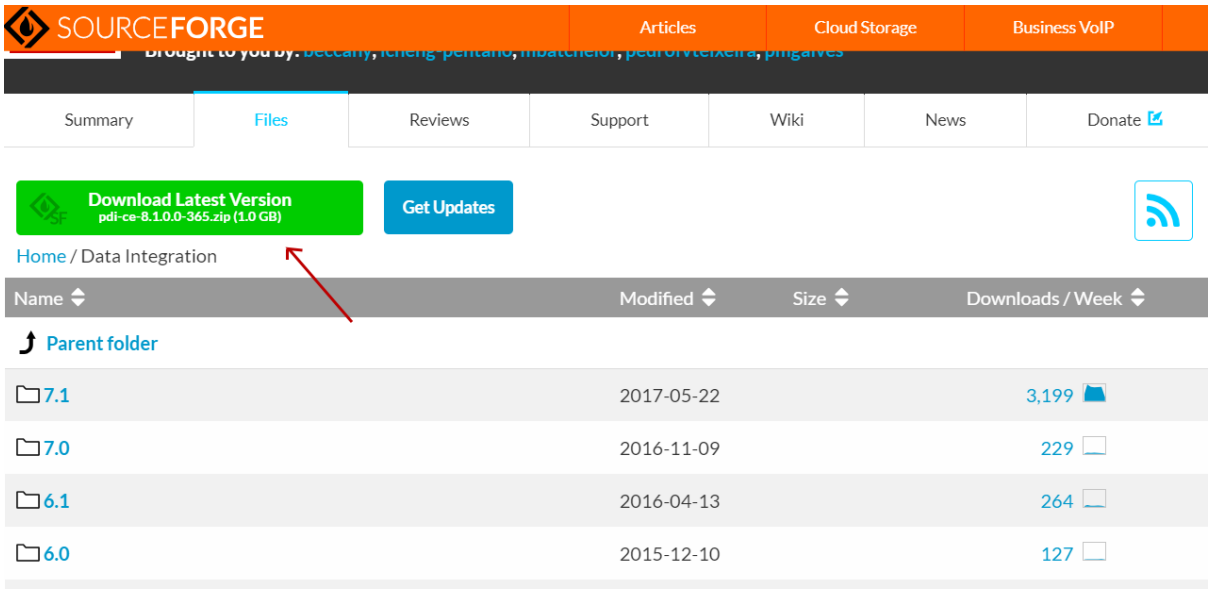
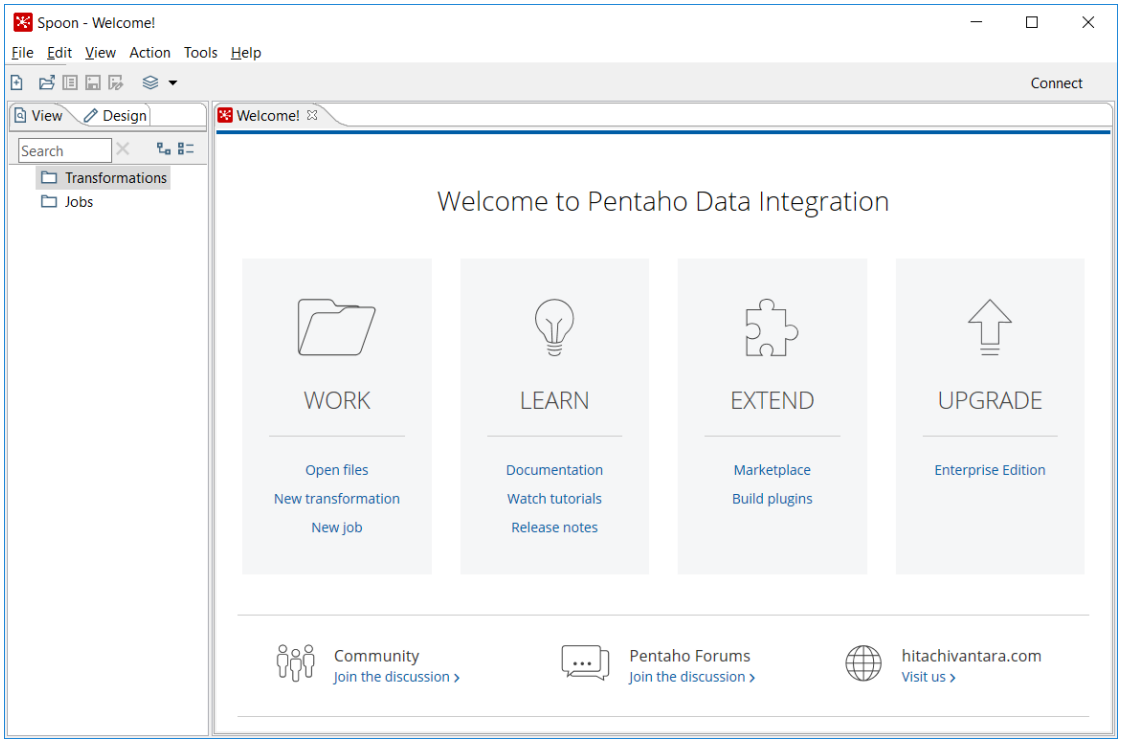
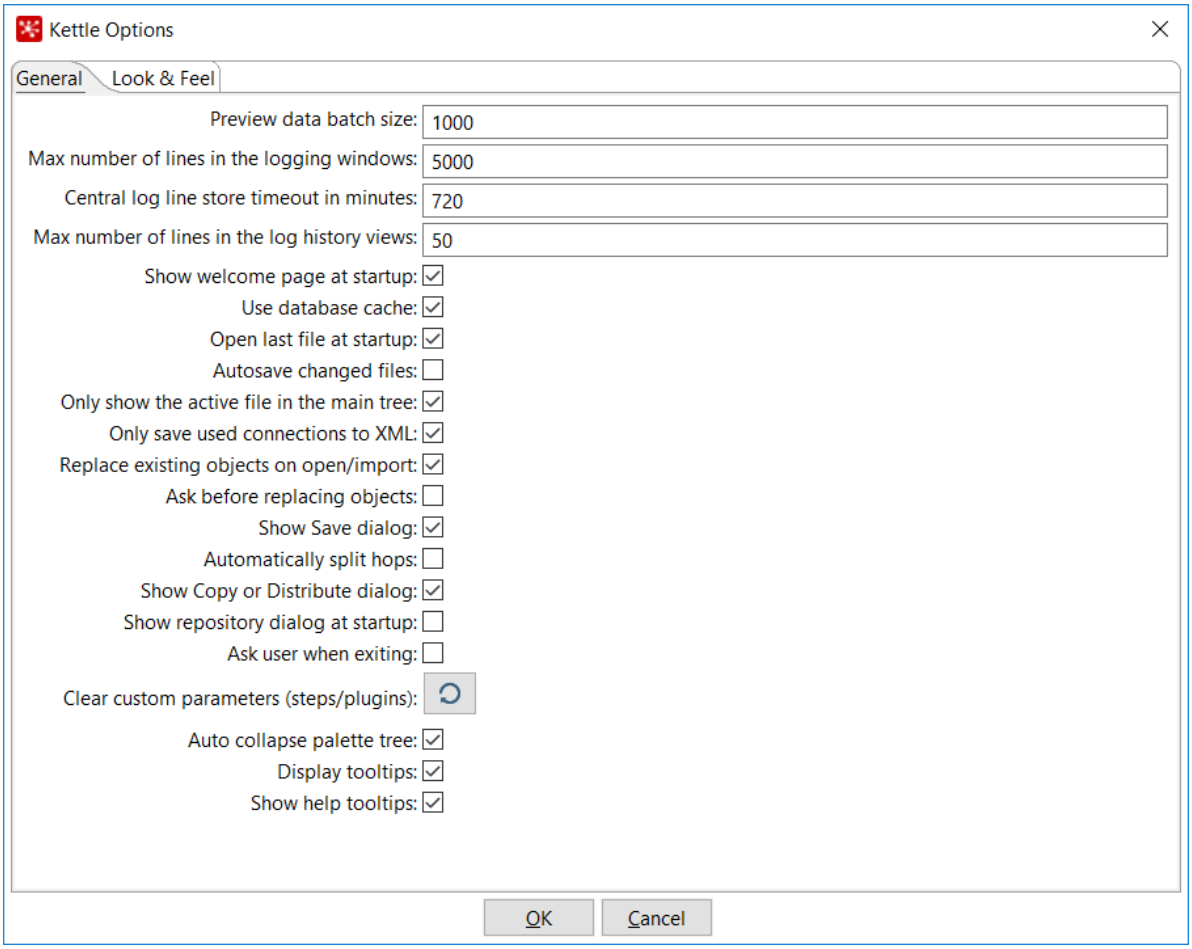
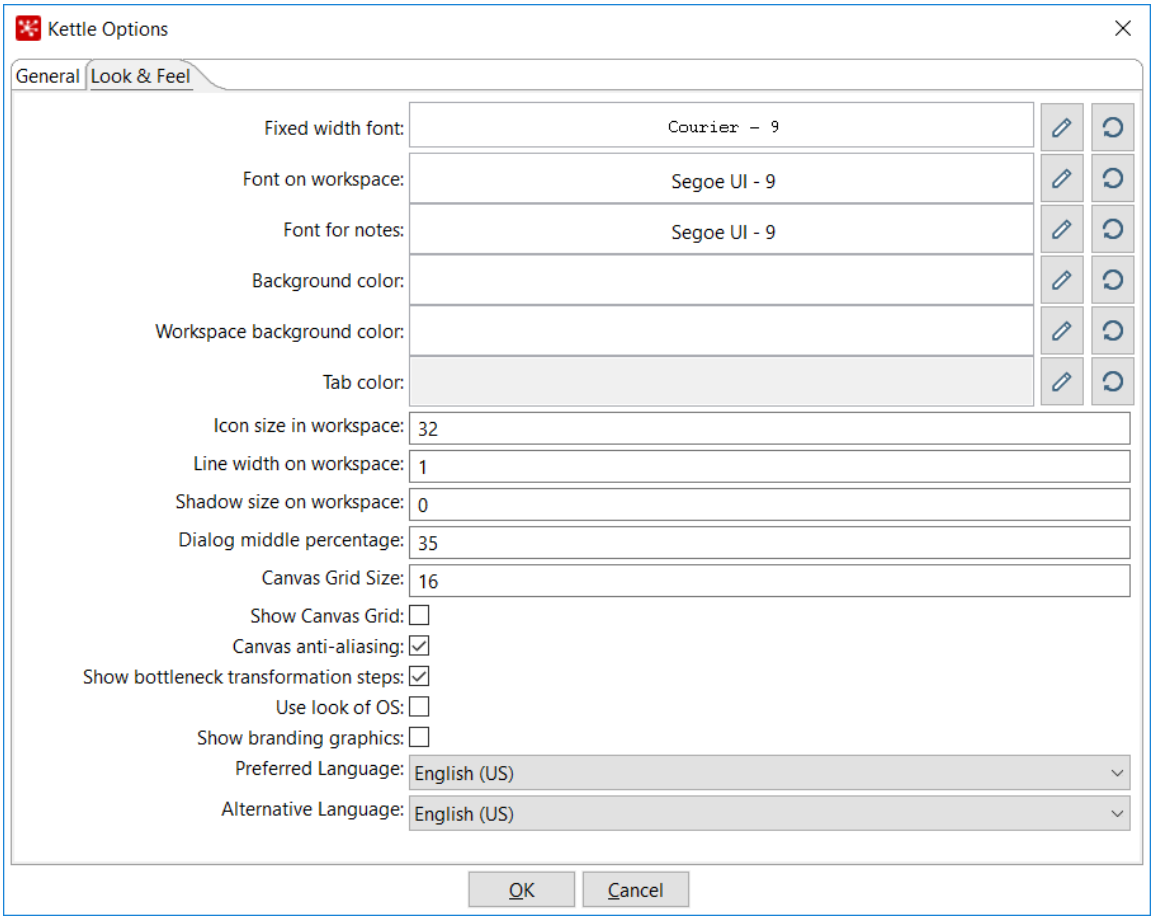
Chapter 1: Getting Started with PDI

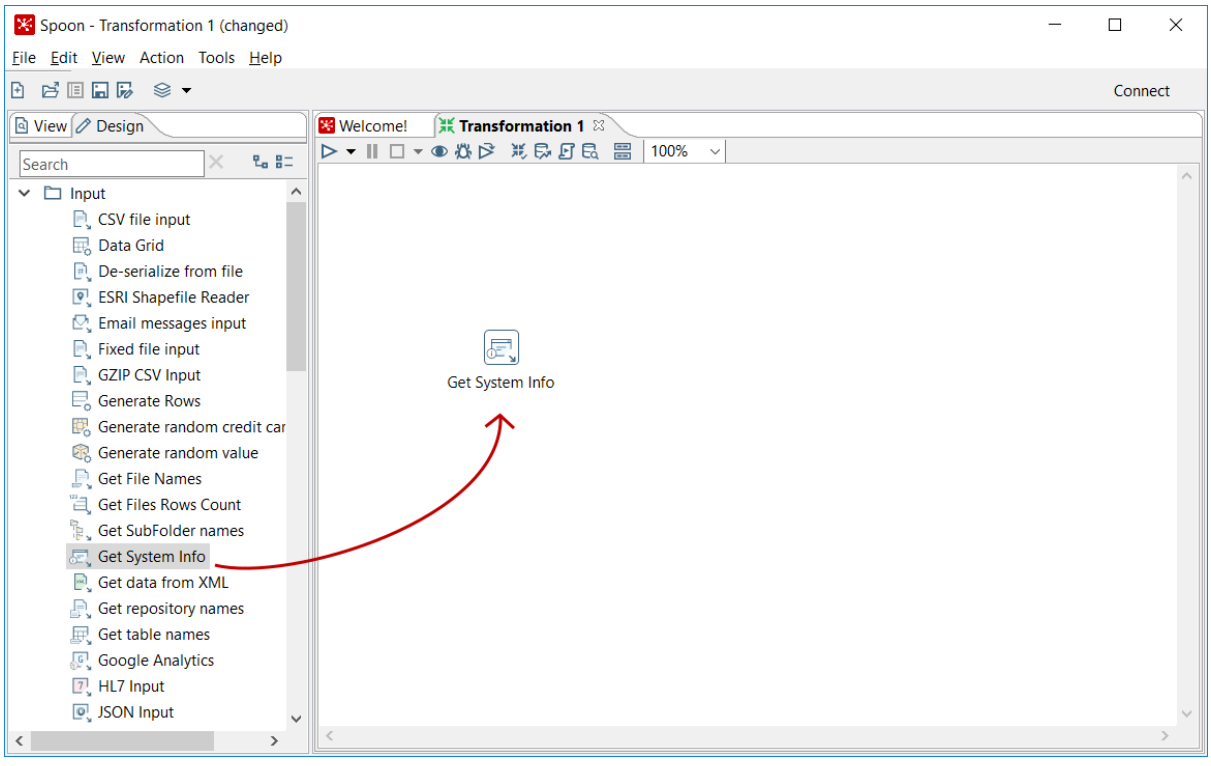


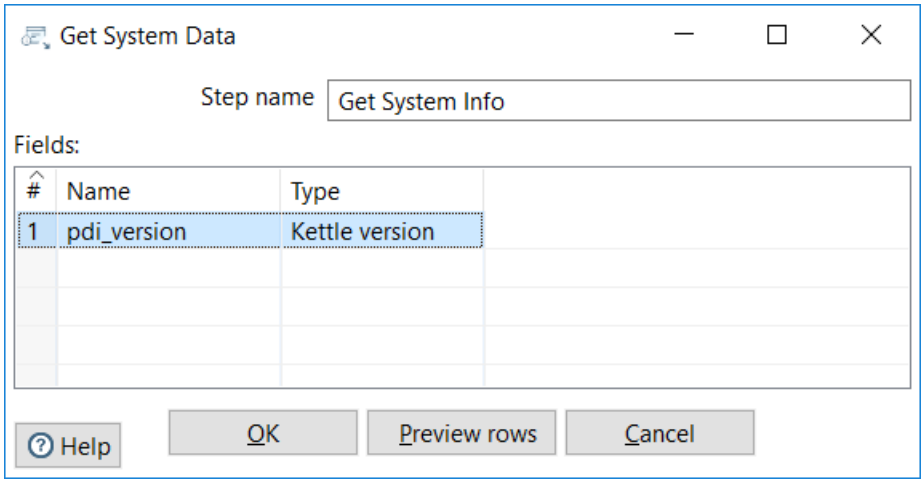



Spoon - Welcome!
File Edit View Action Tools Help
Connect
View
Design
Search
Transformations
Jobs
Welcome!
Welcome to Pentaho Data Integration
WORK
Open files
New transformation
New job
LEARN
Documentation
Watch tutorials
Release notes
EXTEND
Marketplace
Build plugins
UPGRADE
Enterprise Edition
Community
Join the discussion ›
Pentaho Forums
Join the discussion ›
hitachivantara.com
Visit us ›


Kettle Options
General
Look & Feel
Preview data batch size: 1000
Max number of lines in the logging windows: 5000
Central log line store timeout in minutes: 720
Max number of lines in the log history views: 50
Show welcome page at startup:
Use database cache:
Open last file at startup:
Autosave changed files:
Only show the active file in the main tree:
Only save used connections to XML:
Replace existing objects on open/import:
Ask before replacing objects:
Show Save dialog:
Automatically split hops:
Show Copy or Distribute dialog:
Show repository dialog at startup:
Ask user when exiting:
Clear custom parameters (steps/plugins):
Auto collapse palette tree:
Display tooltips:
Show help tooltips:
OK
Cancel

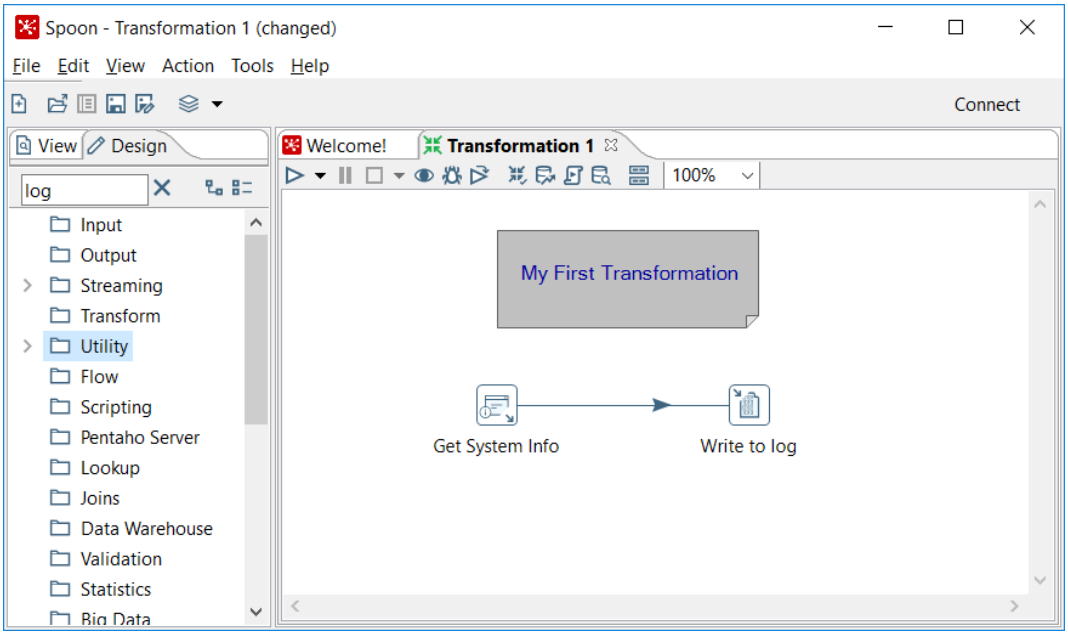

Kettle Options
General
Look & Feel
Fixed width font:
Courier - 9
Font on workspace:
Segoe UI - 9
Font for notes:
Segoe UI - 9
Background color:
Workspace background color:
Tab color:
Icon size in workspace:
32
Line width on workspace:
1
Shadow size on workspace:
0
Dialog middle percentage:
35
Canvas Grid Size:
16
Show Canvas Grid:
Canvas anti-aliasing:
Show bottleneck transformation steps:
Use look of OS:
Show branding graphics:
Preferred Language:
English (US)
Alternative Language:
English (US)
OK
Cancel

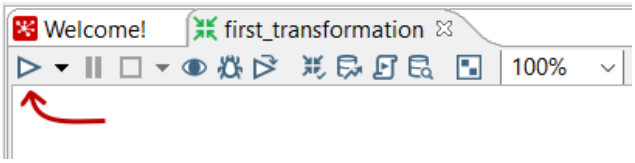

Spoon - Transformation 1 (changed)
File Edit View Action Tools Help
Connect
View
Design
Search
Input
CSV file input
Data Grid
De-serialize from file
ESRI Shapefile Reader
Email messages input
Fixed file input
GZIP CSV Input
Generate Rows
Generate random credit car
Generate random value
Get File Names
Get Files Rows Count
Get SubFolder names
Get System Info
Get data from XML
Get repository names
Get table names
Google Analytics
HL7 Input
JSON Input
Welcome!
Transformation 1
100%
Get System Info

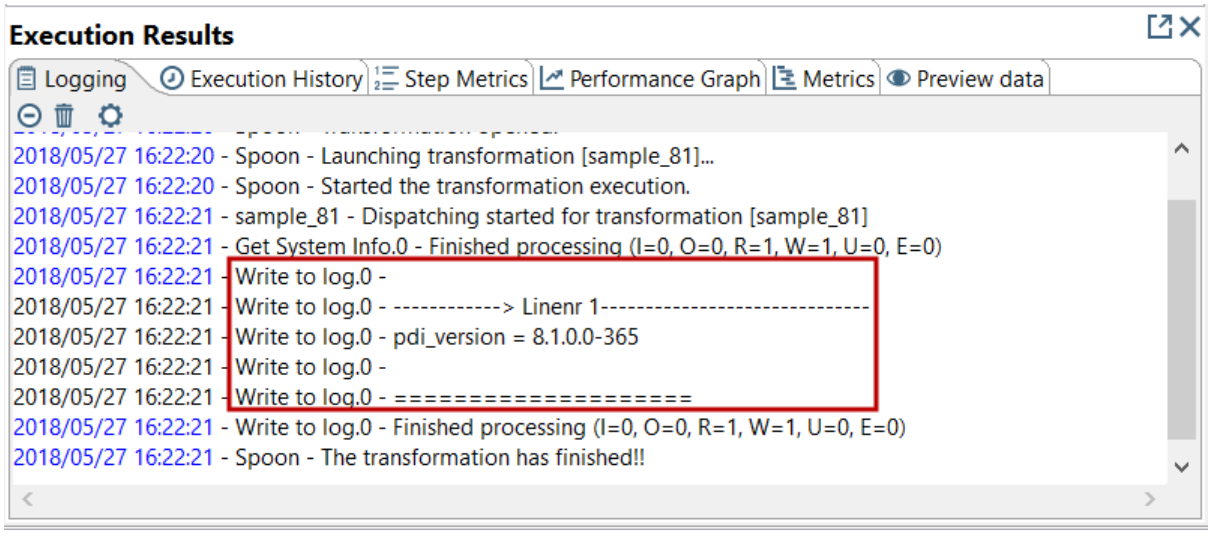

Get System Data
Step name
Get System Info
Fields:
Name Type
1 pdi_version Kettle version
Help
OK
Preview rows
Cancel

Get System Info

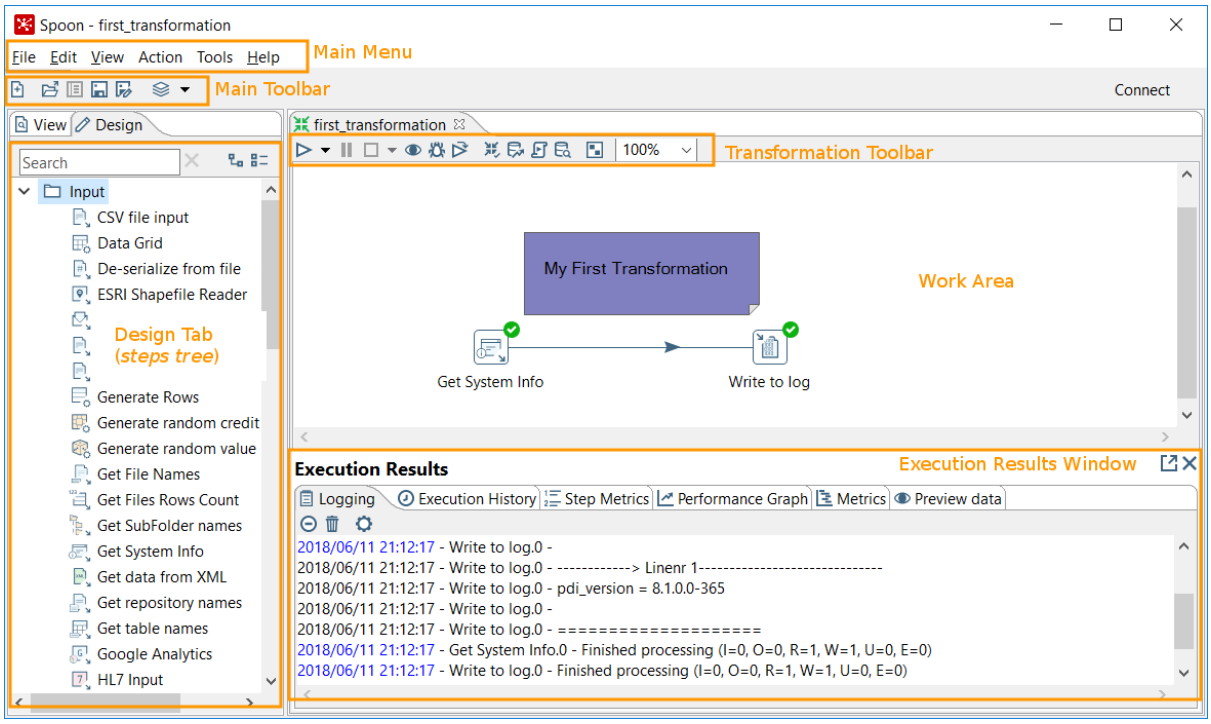
Get System Info
Write to log

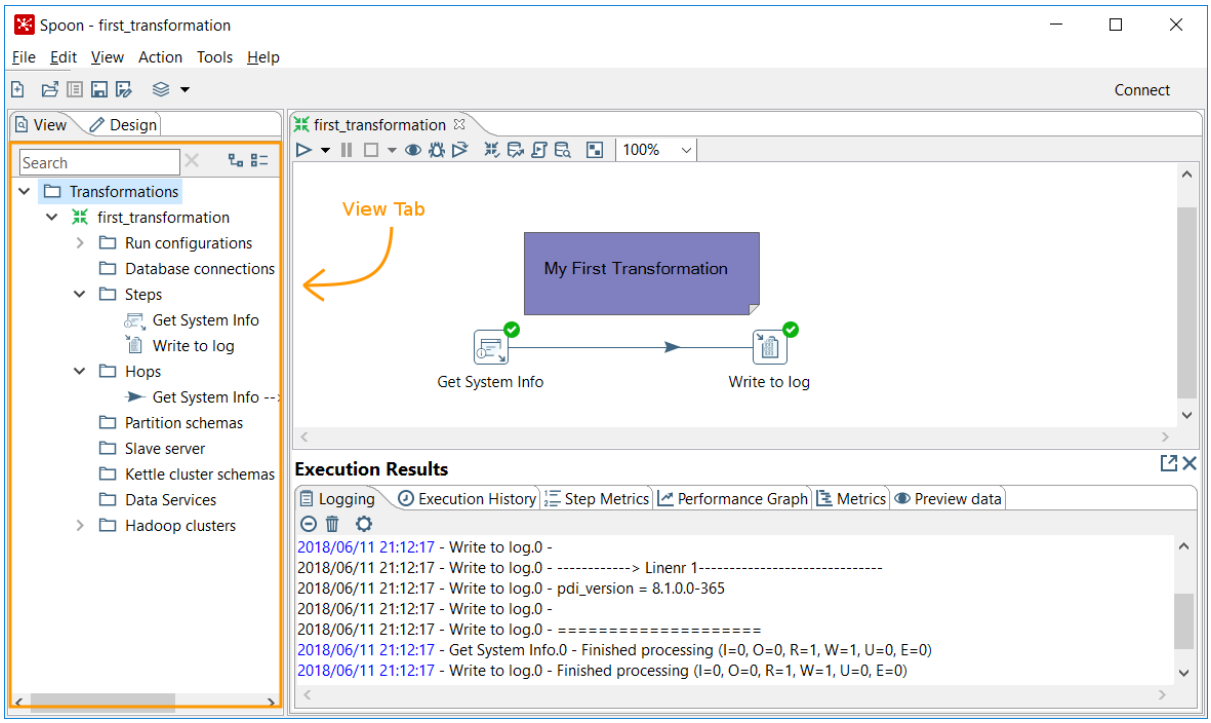

Spoon - Transformation 1 (changed)
File Edit View Action Tools Help
Connect
View Design
log
Input
Output
Streaming
Transform
Utility
Flow
Scripting
Pentaho Server
Lookup
Joins
Data Warehouse
Validation
Statistics
Big Data
Welcome!
Transformation 1
100%
My First Transformation
Get System Info
Write to log

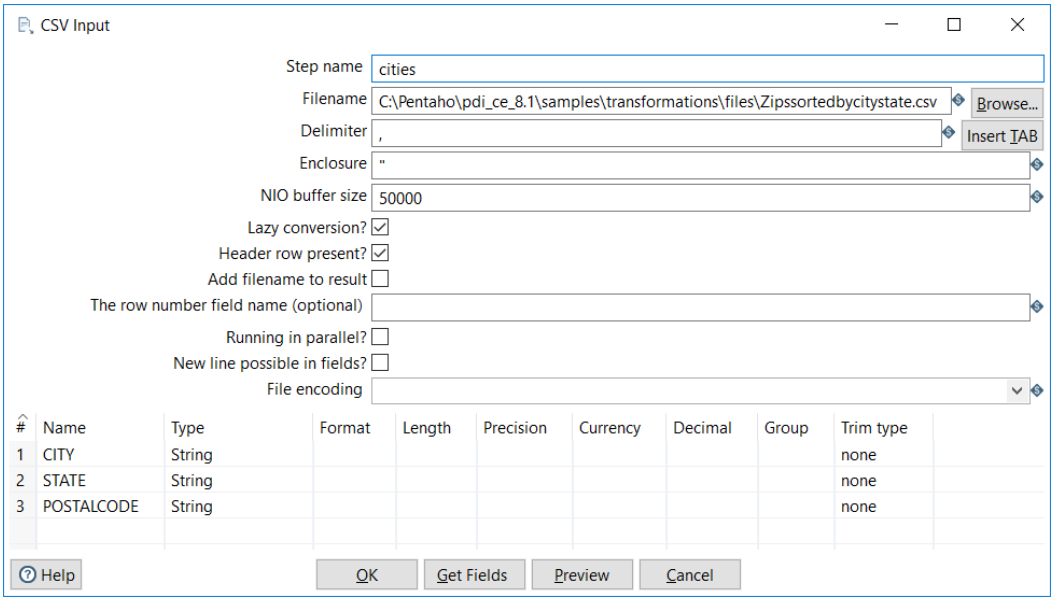

Welcome!
first_transformation
100%

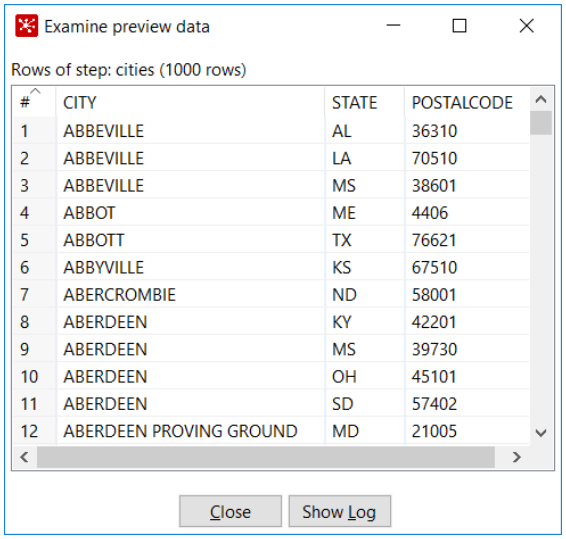

Execution Results
Logging
Execution History
Step Metrics
Performance Graph
Metrics
Preview data
2018/05/27 16:22:20 - Spoon - Launching transformation [sample_81]...
2018/05/27 16:22:20 - Spoon - Started the transformation execution.
2018/05/27 16:22:21 - sample_81 - Dispatching started for transformation [sample_81]
2018/05/27 16:22:21 - Get System Info.0 - Finished processing (I=0, O=0, R=1, W=1, U=0, E=0)
2018/05/27 16:22:21 - Write to log.0 -
2018/05/27 16:22:21 - Write to log.0 - ------------> Linenr 1----------
2018/05/27 16:22:21 - Write to log.0 - pdi_version = 8.1.0.0-365
2018/05/27 16:22:21 - Write to log.0 -
2018/05/27 16:22:21 - Write to log.0 - ====================
2018/05/27 16:22:21 - Write to log.0 - Finished processing (I=0, O=0, R=1, W=1, U=0, E=0)
2018/05/27 16:22:21 - Spoon - The transformation has finished!!

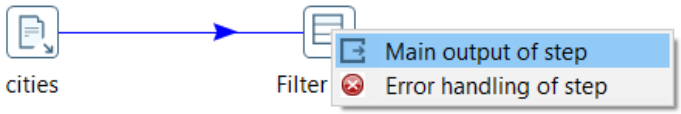
Chapter 2: Getting Familiar with Spoon

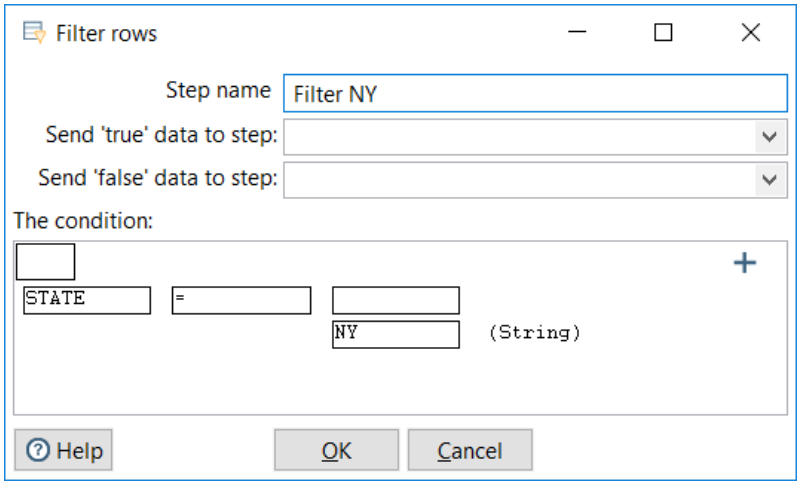




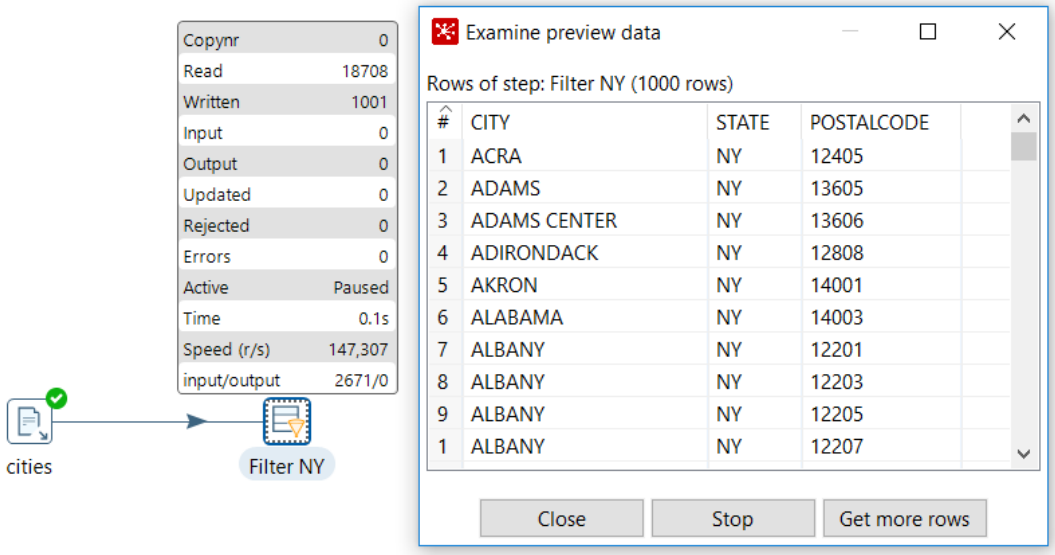

CSV Input
Step name
cities
Filename
C:\Pentaho\pdi_ce_8.1\samples\transformations\files\Zipssortedbycitystate.csv
Browse...
Delimiter
,
Insert TAB
Enclosure
"
NIO buffer size
50000
Lazy conversion?
Header row present?
Add filename to result
The row number field name (optional)
Running in parallel?
New line possible in fields?
File encoding
Name Type Format Length Precision Currency Decimal Group Trim type
1 CITY String none
2 STATE String none
3 POSTALCODE String none
Help
OK
Get Fields
Preview
Cancel

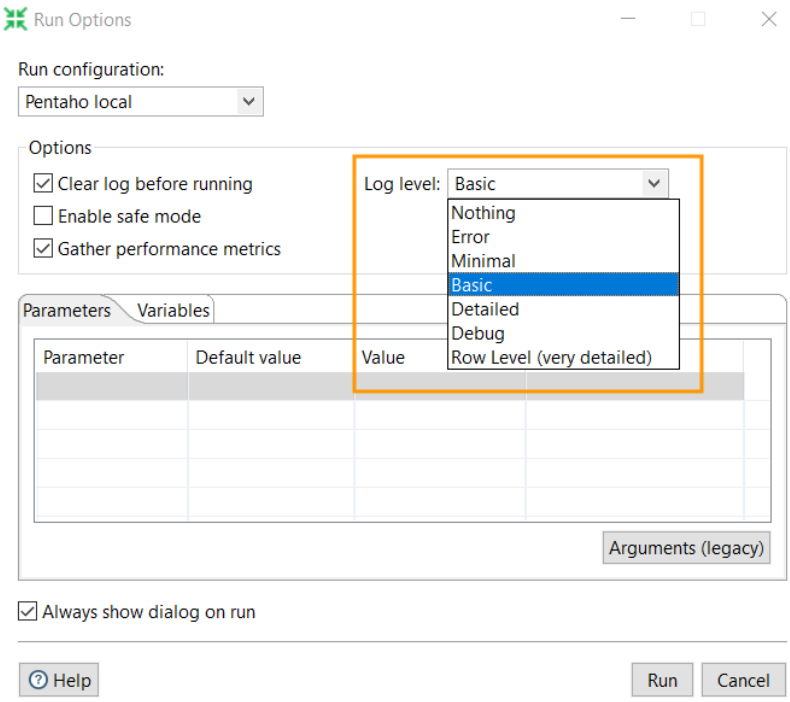

Examine preview data
Rows of step: cities (1000 rows)
CITY STATE POSTALCODE
1 ABBEVILLE AL 36310
2 ABBEVILLE LA 70510
3 ABBEVILLE MS 38601
4 ABBOT ME 4406
5 ABBOTT TX 76621
6 ABBYVILLE KS 67510
7 ABERCROMBIE ND 58001
8 ABERDEEN KY 42201
9 ABERDEEN MS 39730
10 ABERDEEN OH 45101
11 ABERDEEN SD 57402
12 ABERDEEN PROVING GROUND MD 21005
Close
Show Log

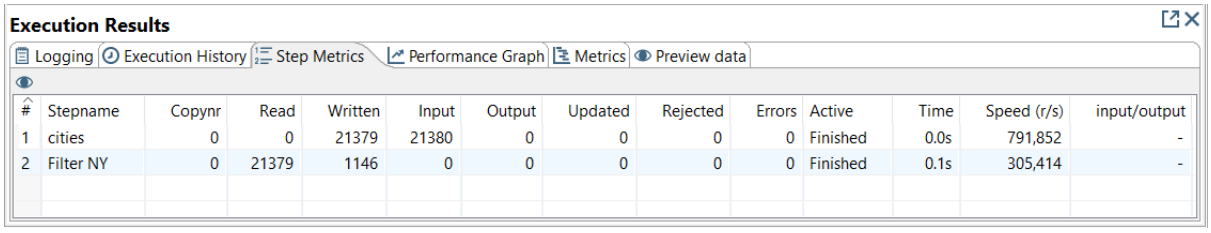

cities
Filter
Main output of step
Error handling of step


Filter rows
Step name
Filter NY
Send 'true' data to step:
Send 'false' data to step:
The condition:
STATE
=
NY
(String)
Help
OK
Cancel

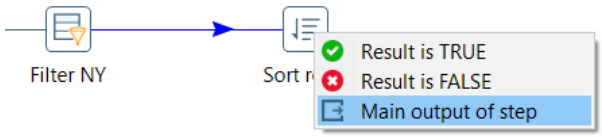
cities
Filter NY

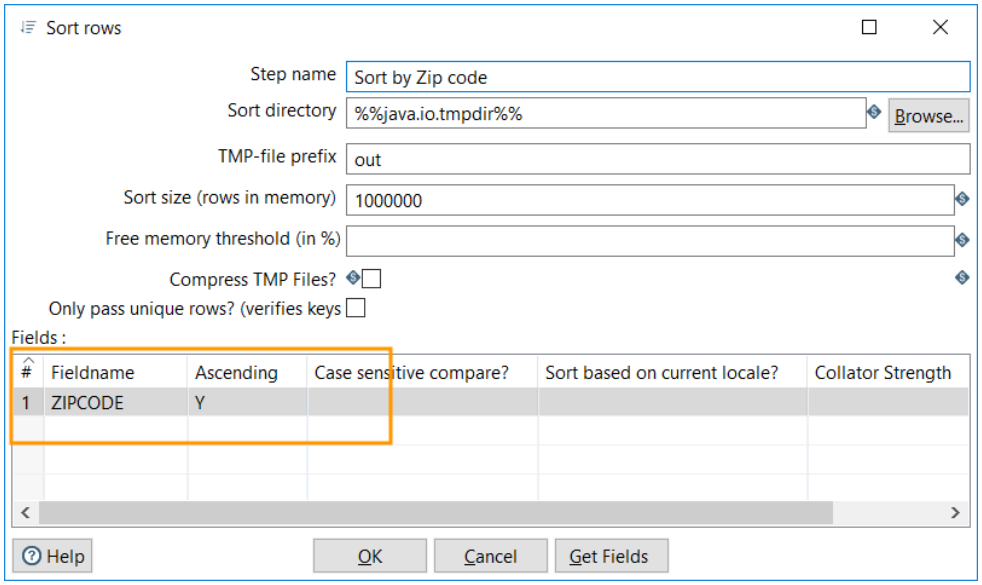

Copynr 0
Read 18708
Written 1001
Input 0
Output 0
Updated 0
Rejected 0
Errors 0
Active Paused
Time 0.1s
Speed (r/s) 147,307
input/output 2671/0
cities
Filter NY
Examine preview data
Rows of step: Filter NY (1000 rows)
CITY STATE POSTALCODE
1 ACRA NY 12405
2 ADAMS NY 13605
3 ADAMS CENTER NY 13606
4 ADIRONDACK NY 12808
5 AKRON NY 14001
6 ALABAMA NY 14003
7 ALBANY NY 12201
8 ALBANY NY 12203
9 ALBANY NY 12205
1 ALBANY NY 12207
Close
Stop
Get more rows

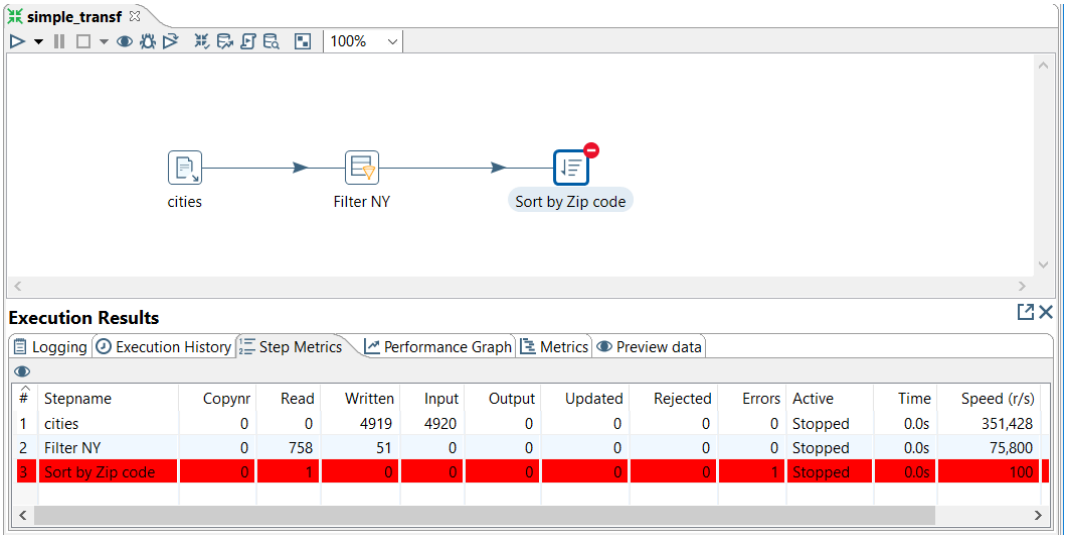


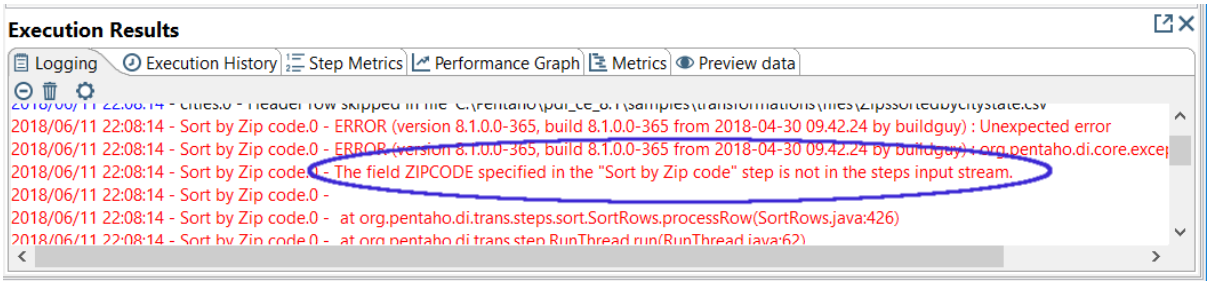


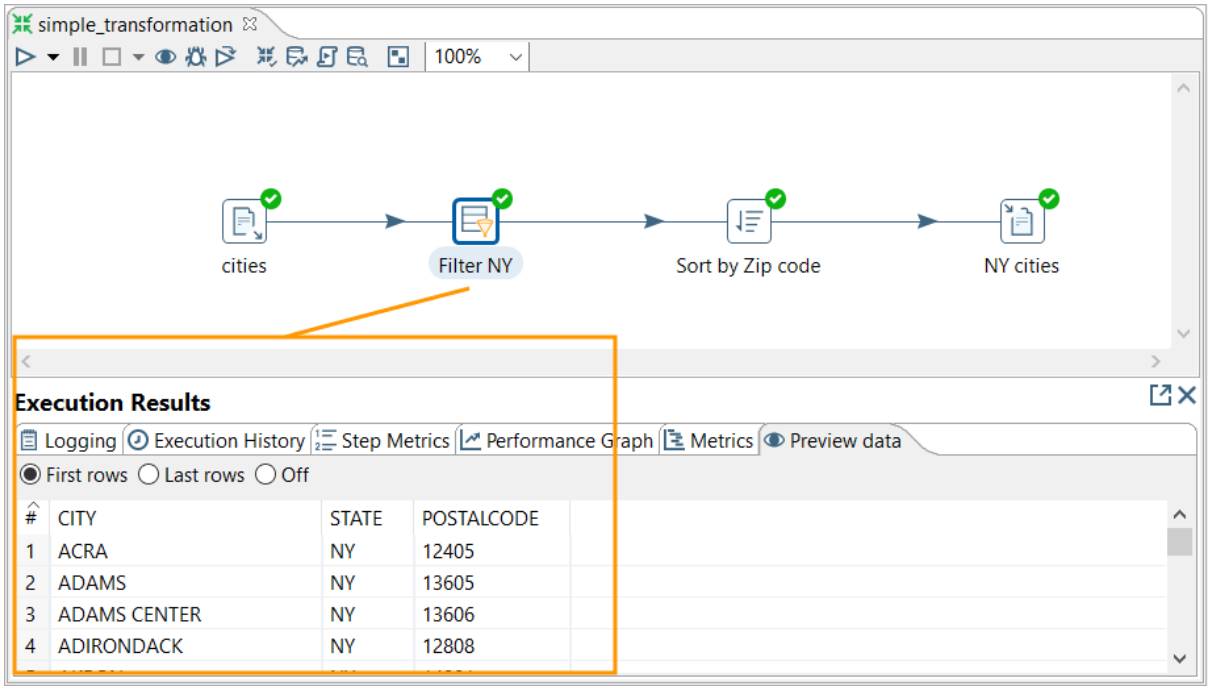
#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	input/output
1	cities	0	0	21379	21380	0	0	0	0	Finished	0.0s	791,852	-
2	Filter NY	0	21379	1146	0	0	0	0	0	Finished	0.1s	305,414	-

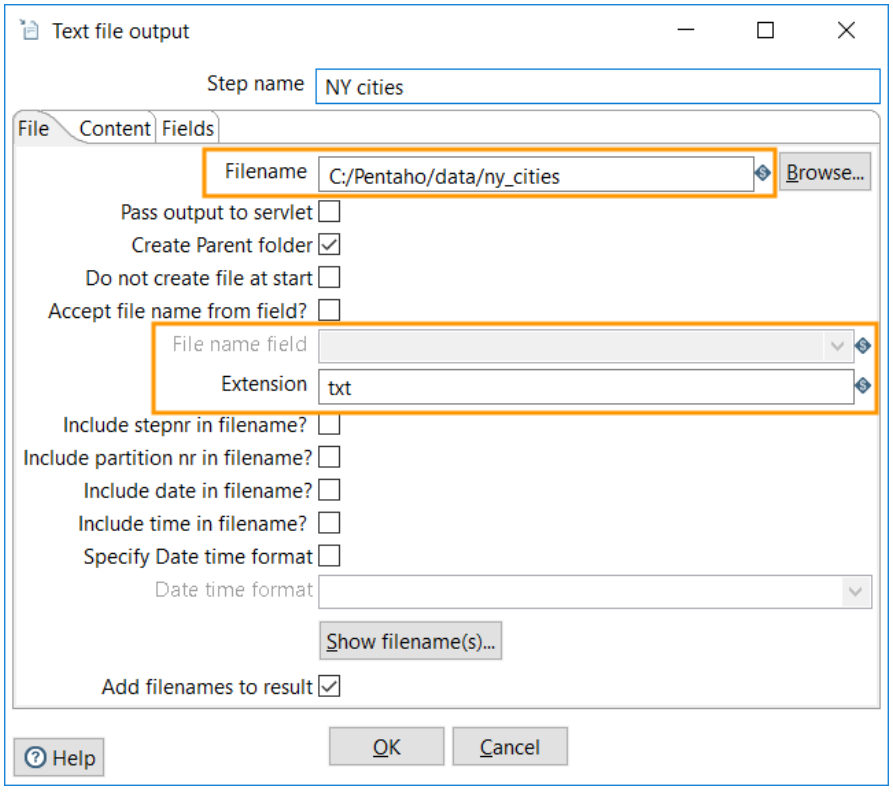


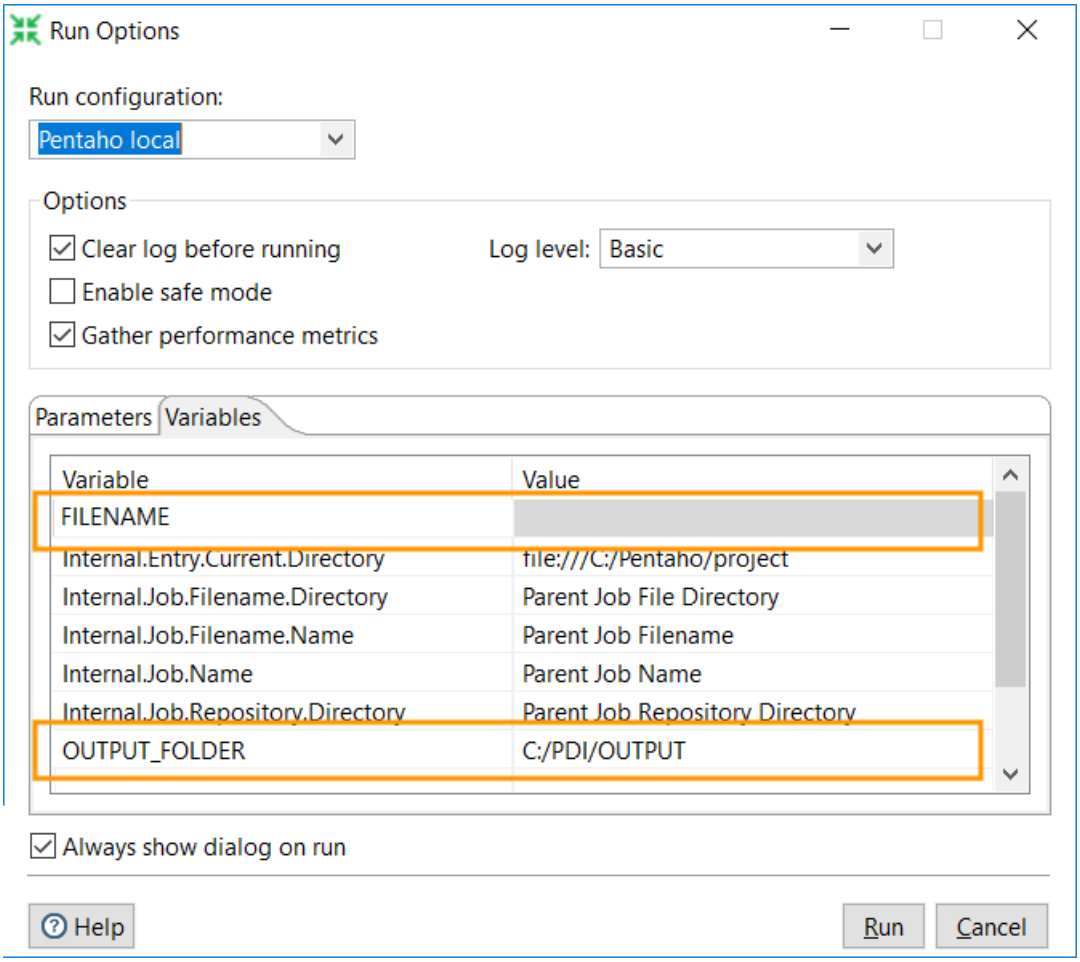

Sort rows
Step name
Sort by Zip code
Sort directory
%%java.io.tmpdir%%
Browse...
TMP-file prefix
out
Sort size (rows in memory)
1000000
Free memory threshold (in %)
Compress TMP Files?
Only pass unique rows? (verifies keys
Fields :
Fieldname Ascending Case sensitive compare? Sort based on current locale? Collator Strength
1 ZIPCODE Y
Help
OK
Cancel
Get Fields

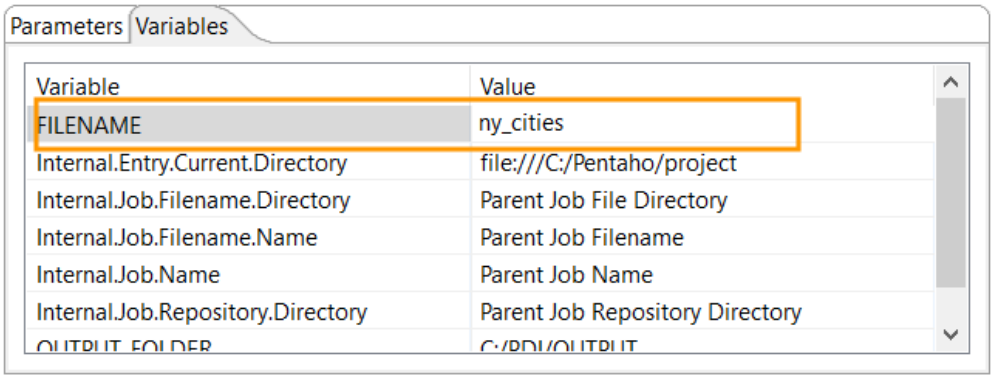

simple_transf
100%
cities
Filter NY
Sort by Zip code
Execution Results
Logging
Execution History
Step Metrics
Performance Graph
Metrics
Preview data
Stepname Copynr Read Written Input Output Updated Rejected Errors Active Time Speed (r/s)
1 cities 0 0 4919 4920 0 0 0 0 Stopped 0.0s 351,428
2 Filter NY 0 758 51 0 0 0 0 0 Stopped 0.0s 75,800
3 Sort by Zip code 0 1 0 0 0 0 0 1 Stopped 0.0s 100

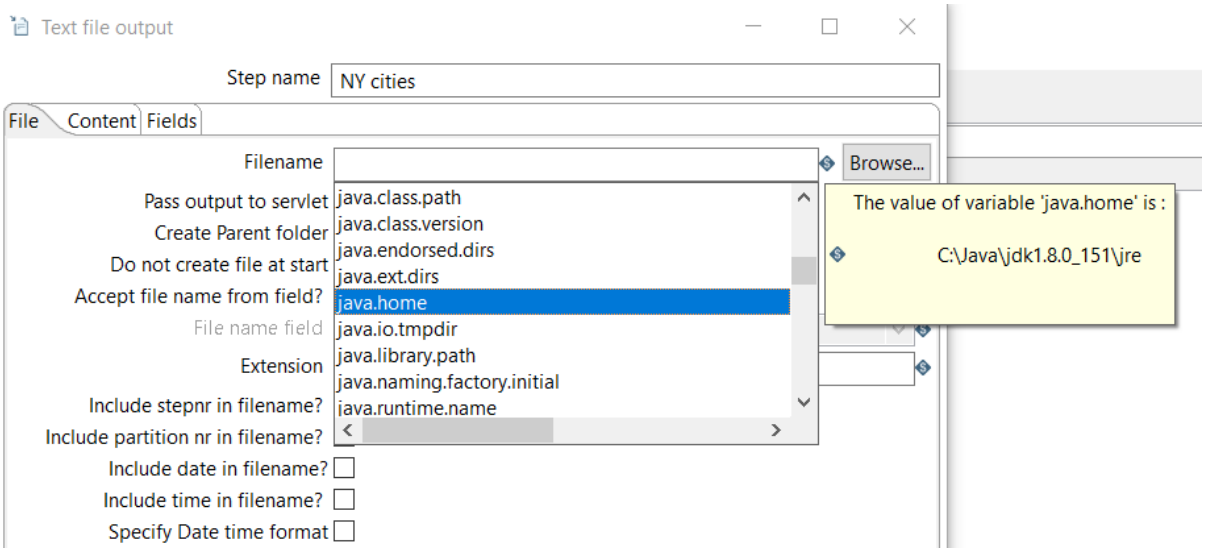

Execution Results
Logging
Execution History
Step Metrics
Performance Graph
Metrics
Preview data
2018/06/11 22:08:14 - Sort by Zip code.0 - ERROR (version 8.1.0.0-365, build 8.1.0.0-365 from 2018-04-30 09.42.24 by buildguy) : Unexpected error
2018/06/11 22:08:14 - Sort by Zip code.0 - The field ZIPCODE specified in the "Sort by Zip code" step is not in the steps input stream.
2018/06/11 22:08:14 - Sort by Zip code.0 -
2018/06/11 22:08:14 - Sort by Zip code.0 - at org.pentaho.di.trans.steps.sort.SortRows.processRow(SortRows.java:426)

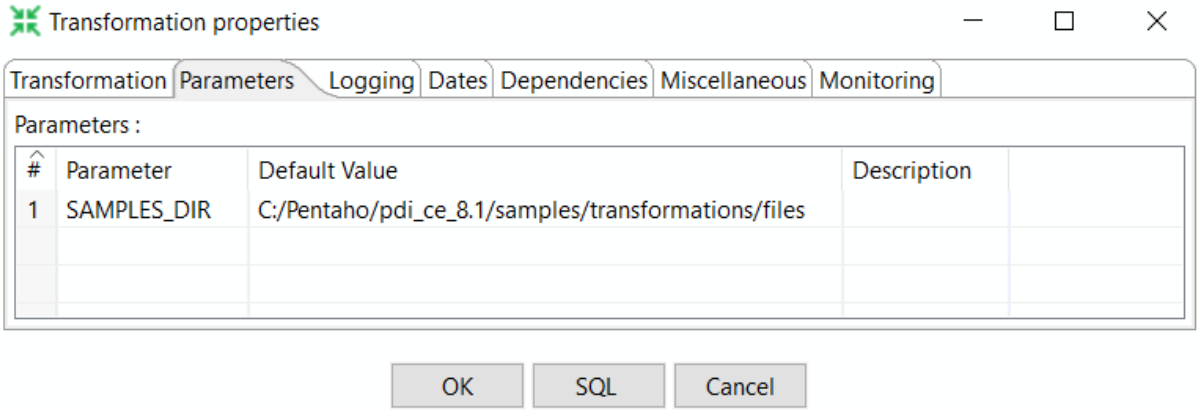

simple_transformation
100%
cities
Filter NY
Sort by Zip code
NY cities
Execution Results
Logging
Execution History
Step Metrics
Performance Graph
Metrics
Preview data
First rows
Last rows
Off
CITY STATE POSTALCODE
1 ACRA NY 12405
2 ADAMS NY 13605
3 ADAMS CENTER NY 13606
4 ADIRONDACK NY 12808

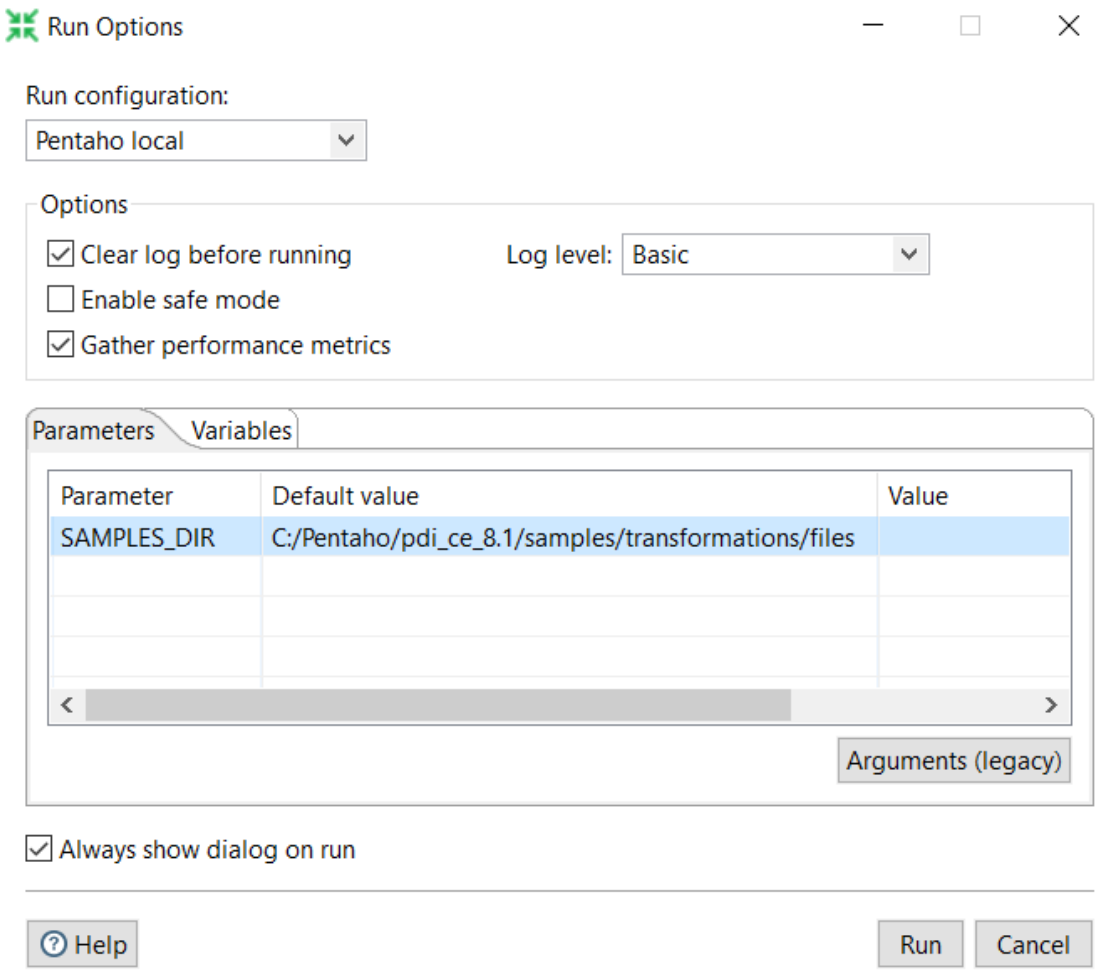

Text file output
Step name
NY cities
File
Content
Fields
Filename
C:/Pentaho/data/ny_cities
Browse...
Pass output to servlet
Create Parent folder
Do not create file at start
Accept file name from field?
File name field
Extension
txt
Include stepnr in filename?
Include partition nr in filename?
Include date in filename?
Include time in filename?
Specify Date time format
Date time format
Show filename(s)...
Add filenames to result
Help
OK
Cancel

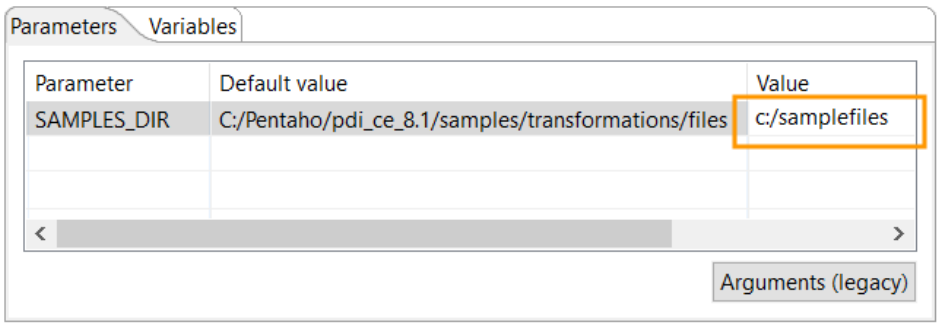

Run Options
Run configuration:
Pentaho local
Options
Clear log before running
Log level:
Basic
Enable safe mode
Gather performance metrics
Parameters
Variables
Variable
Value
FILENAME
Internal.Entry.Current.Directory
file:///C:/Pentaho/project
Internal.Job.Filename.Directory
Parent Job File Directory
Internal.Job.Filename.Name
Parent Job Filename
Internal.Job.Name
Parent Job Name
Internal.Job.Repository.Directory
Parent Job Repository Directory
OUTPUT_FOLDER
C:/PDI/OUTPUT
Always show dialog on run
Help
Run
Cancel


Parameters
Variables
Variable
Value
FILENAME
ny_cities
Internal.Entry.Current.Directory
file:///C:/Pentaho/project
Internal.Job.Filename.Directory
Parent Job File Directory
Internal.Job.Filename.Name
Parent Job Filename
Internal.Job.Name
Parent Job Name
Internal.Job.Repository.Directory
Parent Job Repository Directory

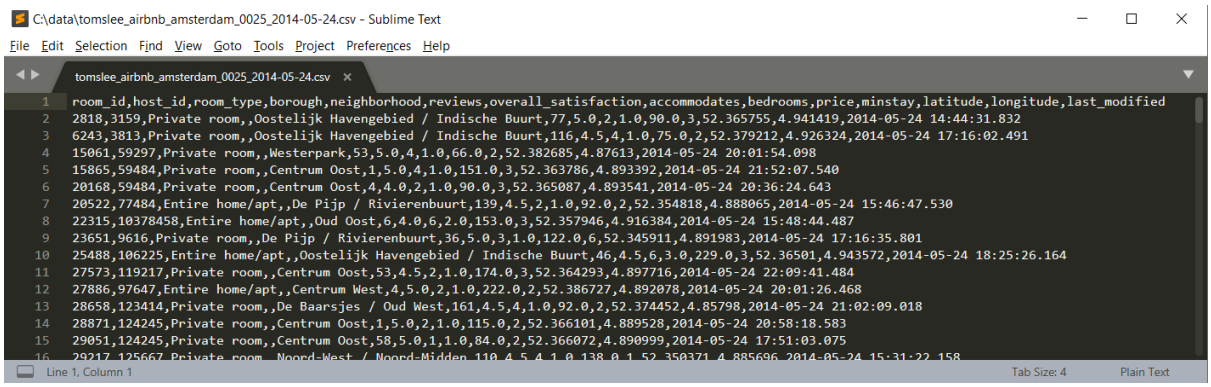

Text file output
Step name
NY cities
File
Content
Fields
Filename
Browse...
Pass output to servlet
Create Parent folder
Do not create file at start
Accept file name from field?
File name field
Extension
Include stepnr in filename?
Include partition nr in filename?
Include date in filename?
Include time in filename?
Specify Date time format
java.class.path
java.class.version
java.endorsed.dirs
java.ext.dirs
java.home
java.io.tmpdir
java.library.path
java.naming.factory.initial
java.runtime.name
The value of variable 'java.home' is :
C:\Java\jdk1.8.0_151\jre

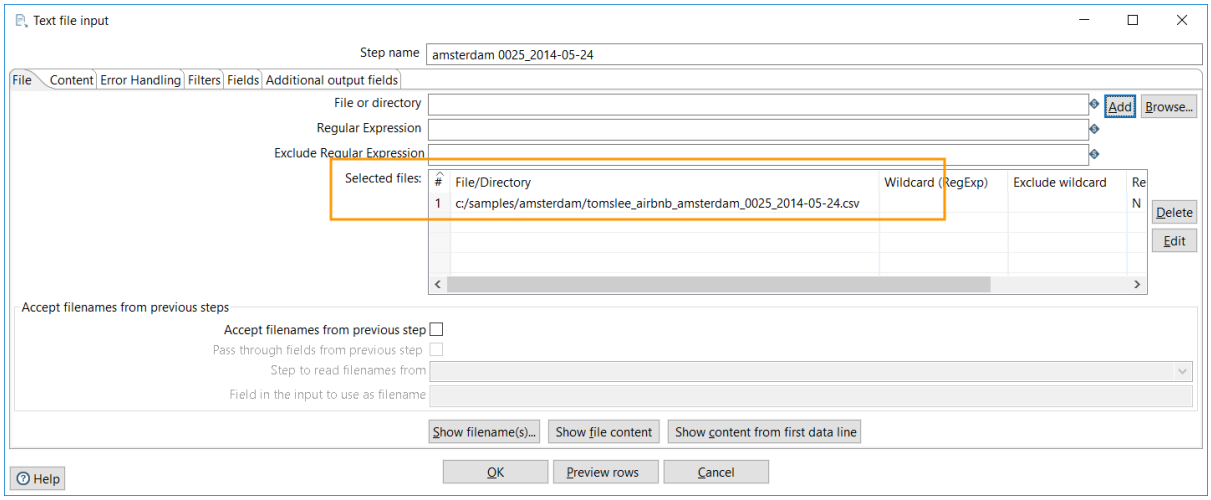

Transformation properties
Transformation
Parameters
Logging
Dates
Dependencies
Miscellaneous
Monitoring
Parameters :
Parameter Default Value Description
1 SAMPLES_DIR C:/Pentaho/pdi_ce_8.1/samples/transformations/files
OK
SQL
Cancel

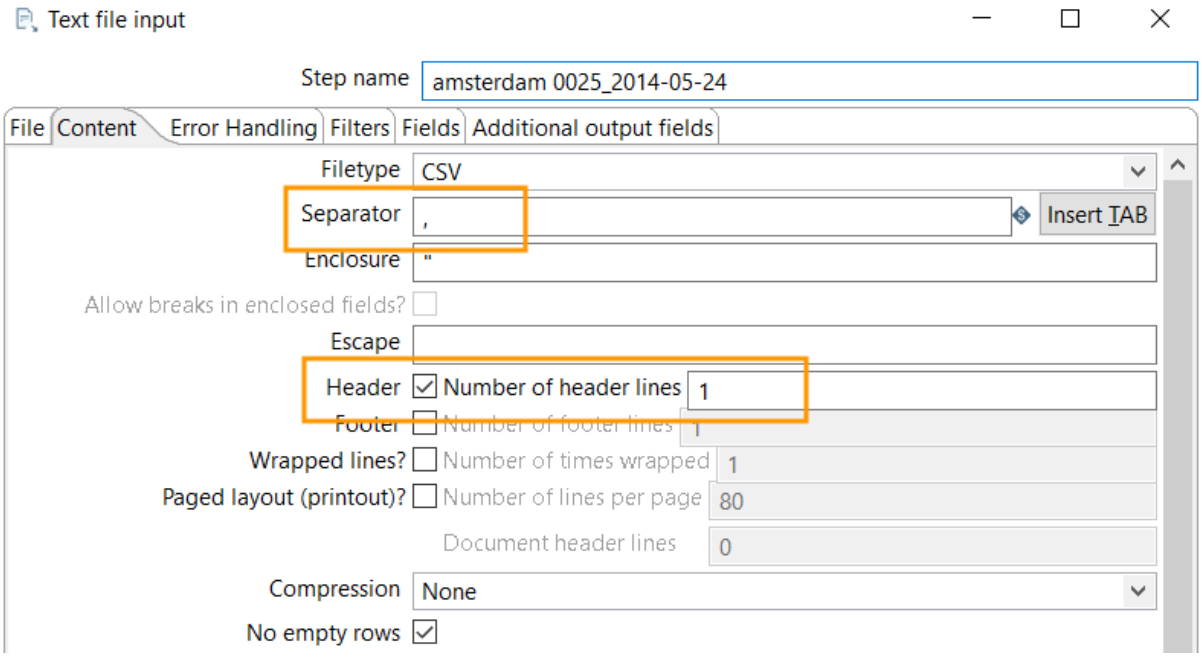

Run Options
Run configuration:
Pentaho local
Options
Clear log before running
Log level:
Basic
Enable safe mode
Gather performance metrics
Parameters
Variables
Parameter
Default value
Value
SAMPLES_DIR
C:/Pentaho/pdi_ce_8.1/samples/transformations/files
Arguments (legacy)
Always show dialog on run
Help
Run
Cancel

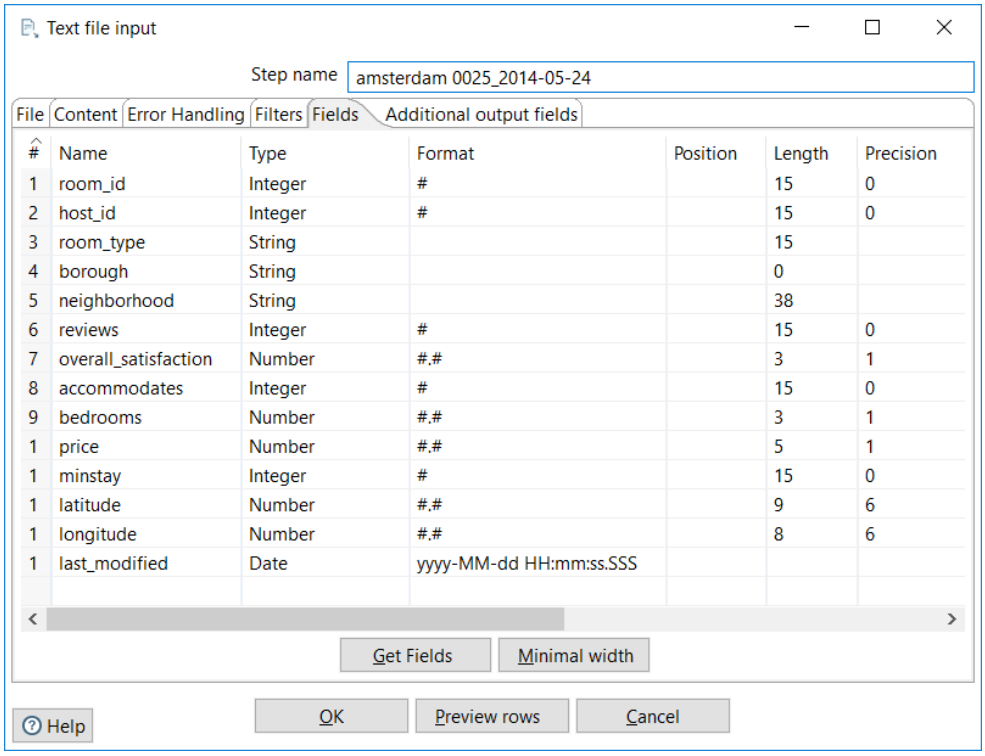

Parameters
Variables
Parameter
Default value
Value
SAMPLES_DIR
C:/Pentaho/pdi_ce_8.1/samples/transformations/files
c:/samplefiles
Arguments (legacy)

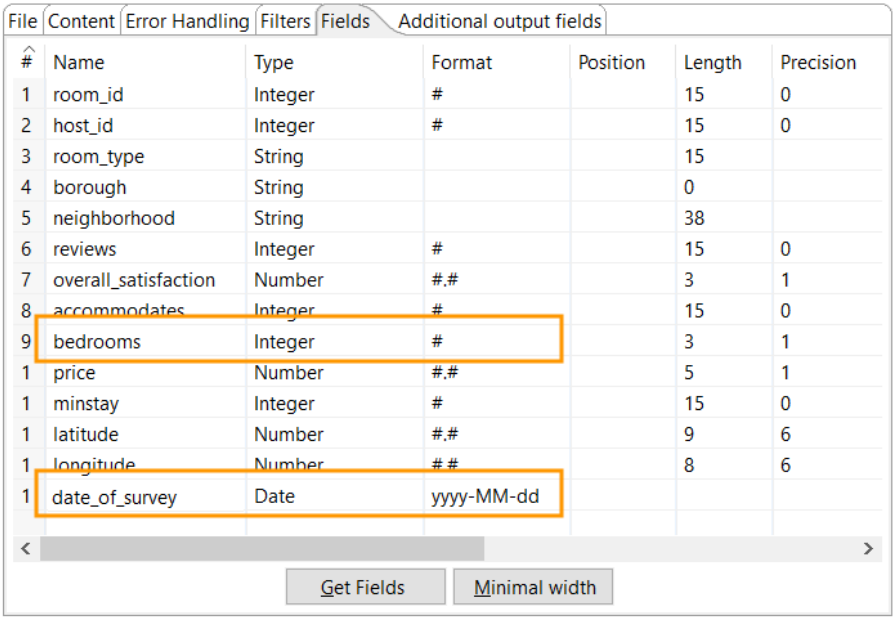
Chapter 3: Extracting Data

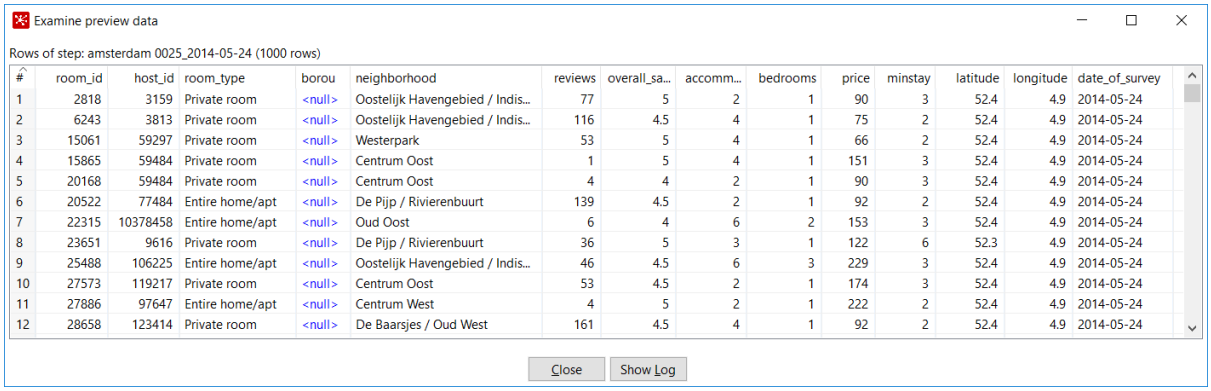




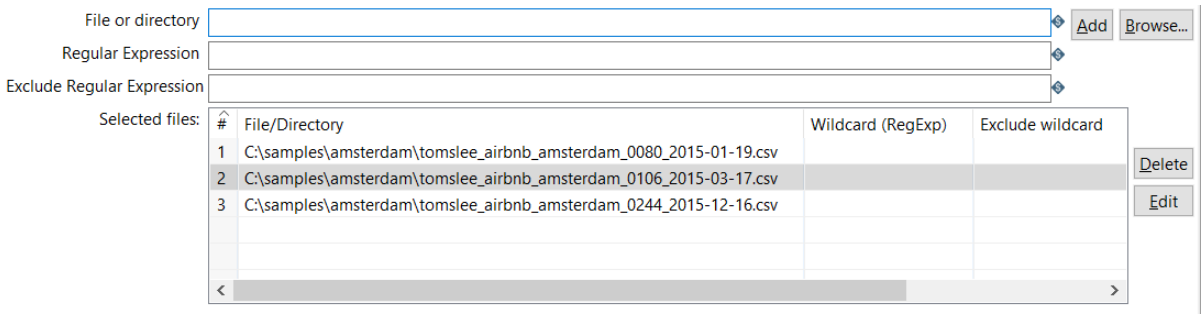


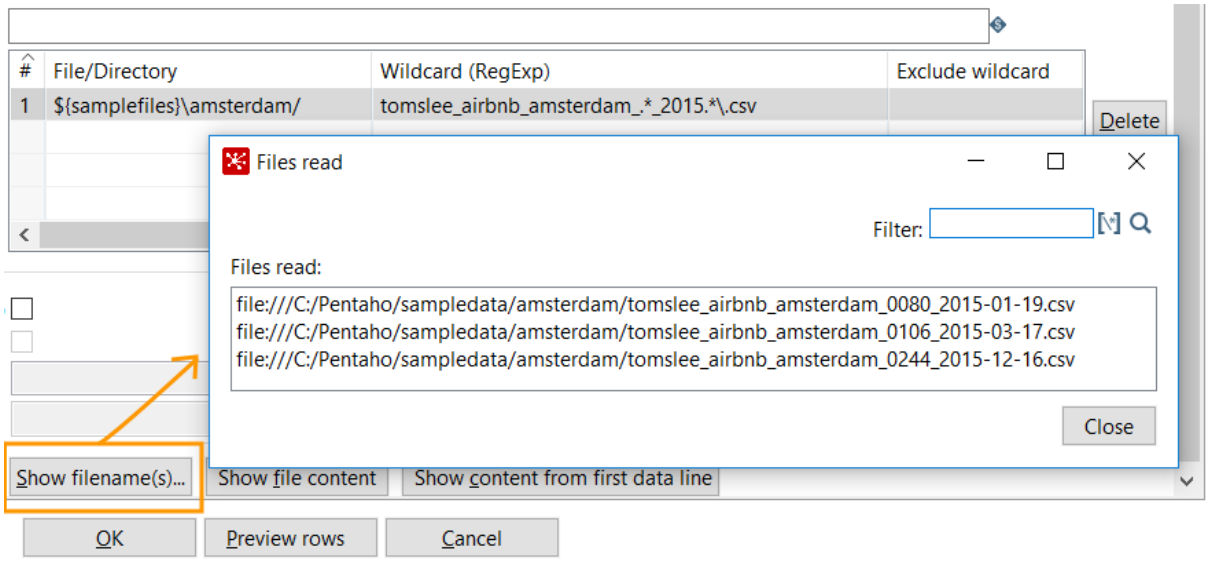

Text file input
Step name
amsterdam 0025_2014-05-24
File
Content
Error Handling
Filters
Fields
Additional output fields
Name Type Format Position Length Precision
1 room_id Integer # 15 0
2 host_id Integer # 15 0
3 room_type String 15
4 borough String 0
5 neighborhood String 38
6 reviews Integer # 15 0
7 overall_satisfaction Number #.# 3 1
8 accommodates Integer # 15 0
9 bedrooms Number #.# 3 1
1 price Number #.# 5 1
1 minstay Integer # 15 0
1 latitude Number #.# 9 6
1 longitude Number #.# 8 6
1 last_modified Date yyyy-MM-dd HH:mm:ss.SSS
Get Fields
Minimal width
Help
OK
Preview rows
Cancel

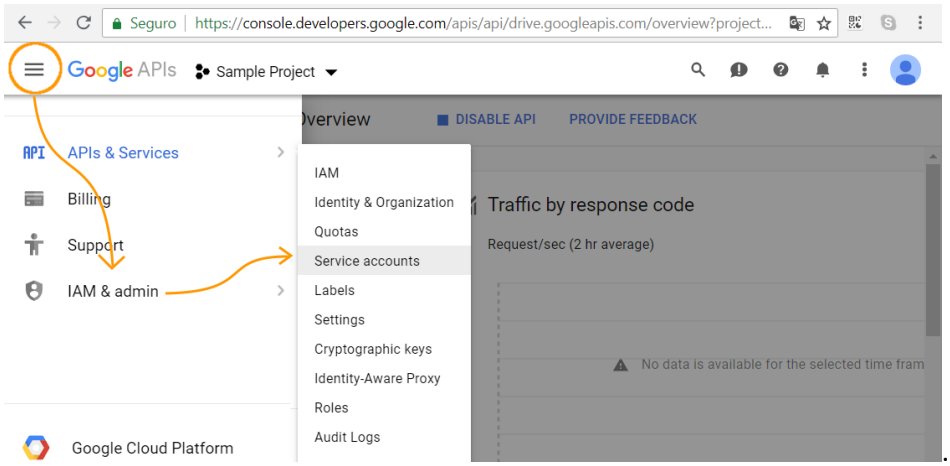

File
Content
Error Handling
Filters
Fields
Additional output fields
Name Type Format Position Length Precision
1 room_id Integer # 15 0
2 host_id Integer # 15 0
3 room_type String 15
4 borough String 0
5 neighborhood String 38
6 reviews Integer # 15 0
7 overall_satisfaction Number #.# 3 1
8 accommodates Integer # 15 0
9 bedrooms Integer # 3 1
1 price Number #.# 5 1
1 minstay Integer # 15 0
1 latitude Number #.# 9 6
1 longitude Number #.# 8 6
1 date_of_survey Date yyyy-MM-dd
Get Fields
Minimal width

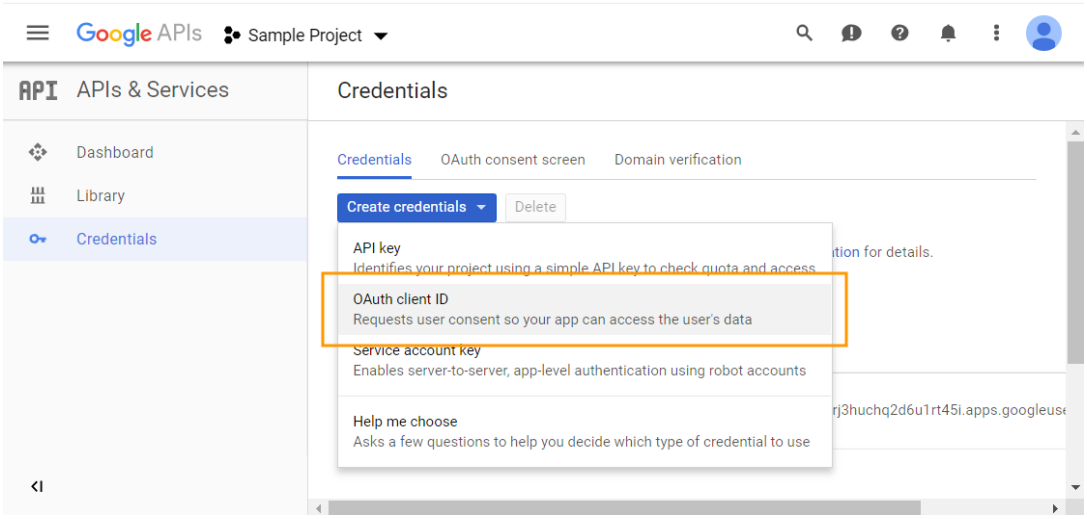


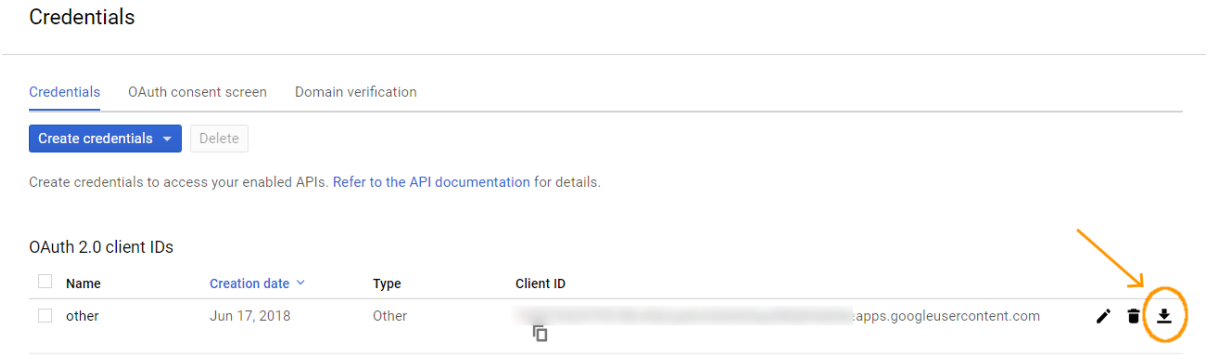
#	room_id	host_id	room_type	borou	neighborhood	reviews	overall_sa...	accomm...	bedrooms	price	minstay	latitude	longitude	date_of_survey
1	2818	3159	Private room	<null>	Oostelijk Havengebied / Indis...	77	5	2	1	90	3	52.4	4.9	2014-05-24
2	6243	3813	Private room	<null>	Oostelijk Havengebied / Indis...	116	4.5	4	1	75	2	52.4	4.9	2014-05-24
3	15061	59297	Private room	<null>	Westerpark	53	5	4	1	66	2	52.4	4.9	2014-05-24
4	15865	59484	Private room	<null>	Centrum Oost	1	5	4	1	151	3	52.4	4.9	2014-05-24
5	20168	59484	Private room	<null>	Centrum Oost	4	4	2	1	90	3	52.4	4.9	2014-05-24
6	20522	77484	Entire home/apt	<null>	De Pijp / Rivierenbuurt	139	4.5	2	1	92	2	52.4	4.9	2014-05-24
7	22315	10378458	Entire home/apt	<null>	Oud Oost	6	4	6	2	153	3	52.4	4.9	2014-05-24
8	23651	9616	Private room	<null>	De Pijp / Rivierenbuurt	36	5	3	1	122	6	52.3	4.9	2014-05-24
9	25488	106225	Entire home/apt	<null>	Oostelijk Havengebied / Indis...	46	4.5	6	3	229	3	52.4	4.9	2014-05-24
10	27573	119217	Private room	<null>	Centrum Oost	53	4.5	2	1	174	3	52.4	4.9	2014-05-24
11	27886	97647	Entire home/apt	<null>	Centrum West	4	5	2	1	222	2	52.4	4.9	2014-05-24
12	28658	123414	Private room	<null>	De Baarsjes / Oud West	161	4.5	4	1	92	2	52.4	4.9	2014-05-24

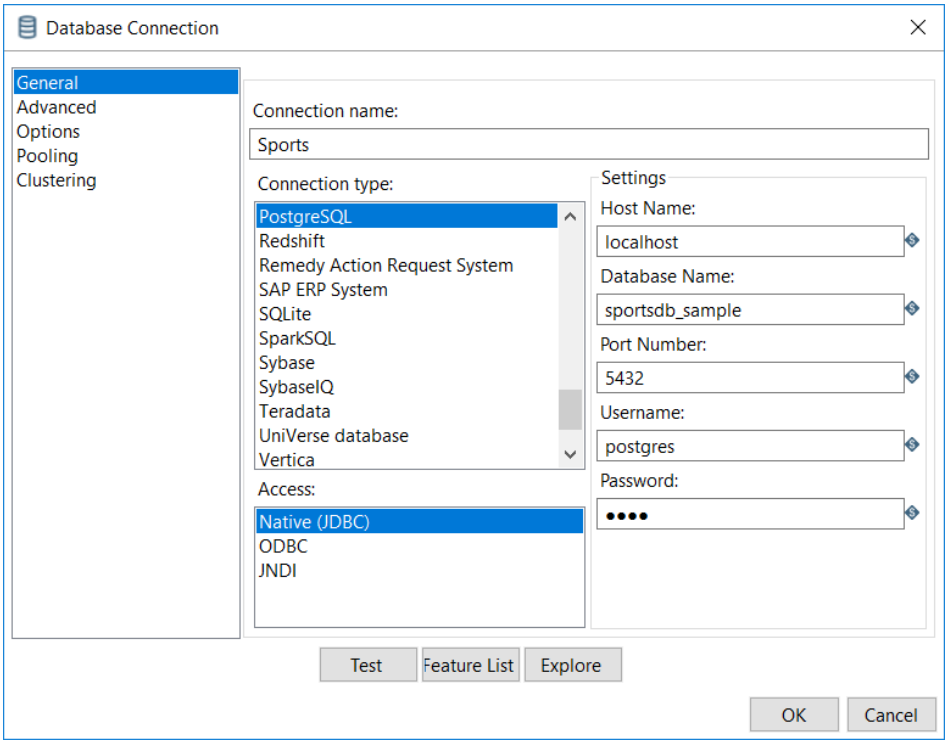


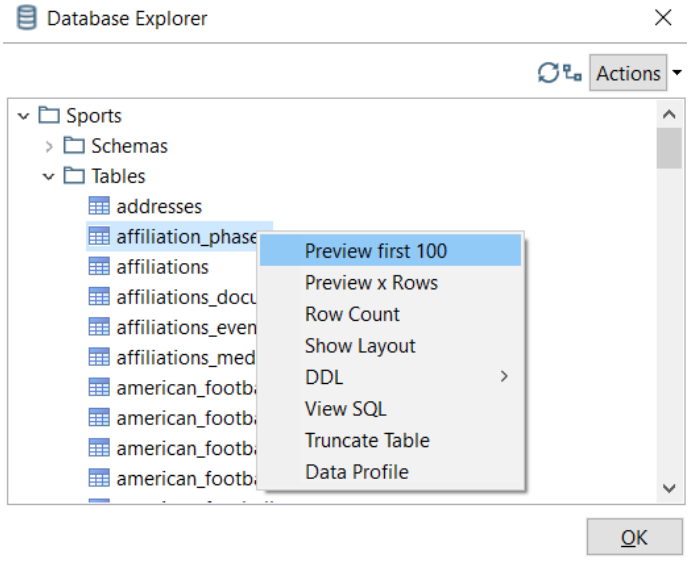


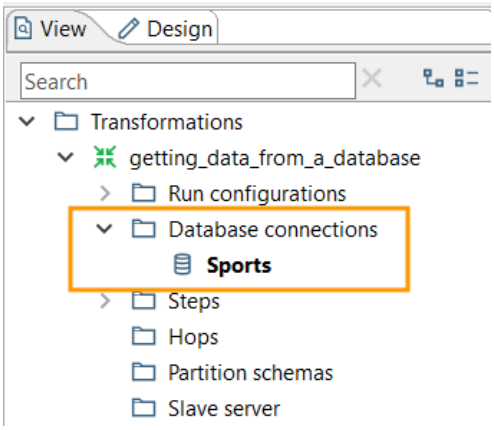

Seguro | https://console.developers.google.com/apis/api/drive.googleapis.com/overview?project...
Google APIs
Sample Project
APIs & Services
Billing
Support
IAM & admin
Google Cloud Platform
Overview
DISABLE API
PROVIDE FEEDBACK
IAM
Identity & Organization
Quotas
Service accounts
Labels
Settings
Cryptographic keys
Identity-Aware Proxy
Roles
Audit Logs
Traffic by response code
Request/sec (2 hr average)
No data is available for the selected time fram


Google APIs
Sample Project
APIs & Services
Dashboard
Library
Credentials
Credentials
Credentials
OAuth consent screen
Domain verification
Create credentials
Delete
API key
Identifies your project using a simple API key to check quota and access
OAuth client ID
Requests user consent so your app can access the user's data
Service account key
Enables server-to-server, app-level authentication using robot accounts
Help me choose
Asks a few questions to help you decide which type of credential to use
for details.


Credentials
Credentials
OAuth consent screen
Domain verification
Create credentials
Delete
Create credentials to access your enabled APIs. Refer to the API documentation for details.
OAuth 2.0 client IDs
Name
Creation date
Type
Client ID
other
Jun 17, 2018
Other
apps.googleusercontent.com


Database Connection
General
Advanced
Options
Pooling
Clustering
Connection name:
Sports
Connection type:
PostgreSQL
Redshift
Remedy Action Request System
SAP ERP System
SQLite
SparkSQL
Sybase
SybaseIQ
Teradata
UniVerse database
Vertica
Access:
Native (JDBC)
ODBC
JNDI
Settings
Host Name:
localhost
Database Name:
sportsdb_sample
Port Number:
5432
Username:
postgres
Password:
Test
Feature List
Explore
OK
Cancel


Database Explorer
Actions
Sports
Schemas
Tables
addresses
affiliation_phase
affiliations
affiliations_docu
affiliations_even
affiliations_med
american_footb
american_footb
american_footb
american_footb
Preview first 100
Preview x Rows
Row Count
Show Layout
DDL
View SQL
Truncate Table
Data Profile
OK


View
Design
Search
Transformations
getting_data_from_a_database
Run configurations
Database connections
Sports
Steps
Hops
Partition schemas
Slave server

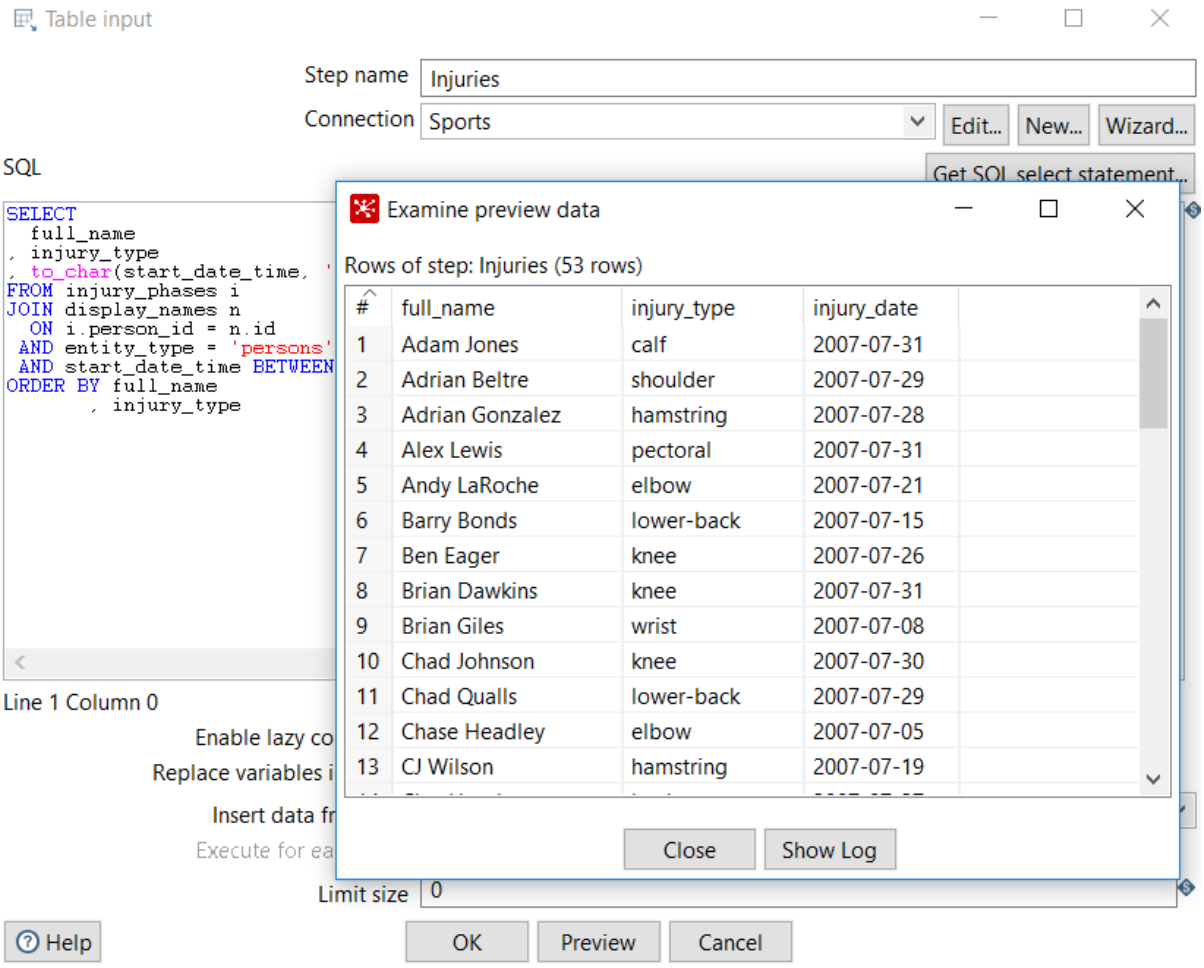


Table input
Step name
Injuries
Connection
Sports
Edit...
New...
Wizard...
SQL
Get SQL select statement...
SELECT
full_name
, injury_type
, to_char(start_date_time,
FROM injury_phases i
JOIN display_names n
ON i.person_id = n.id
AND entity_type = 'persons'
AND start_date_time BETWEEN
ORDER BY full_name
, injury_type
Line 1 Column 0
Enable lazy co
Replace variables i
Insert data fr
Execute for ea
Limit size
0
Help
OK
Preview
Cancel
Examine preview data
Rows of step: Injuries (53 rows)
full_name injury_type injury_date
1 Adam Jones calf 2007-07-31
2 Adrian Beltre shoulder 2007-07-29
3 Adrian Gonzalez hamstring 2007-07-28
4 Alex Lewis pectoral 2007-07-31
5 Andy LaRoche elbow 2007-07-21
6 Barry Bonds lower-back 2007-07-15
7 Ben Eager knee 2007-07-26
8 Brian Dawkins knee 2007-07-31
9 Brian Giles wrist 2007-07-08
10 Chad Johnson knee 2007-07-30
11 Chad Qualls lower-back 2007-07-29
12 Chase Headley elbow 2007-07-05
13 CJ Wilson hamstring 2007-07-19
Close
Show Log

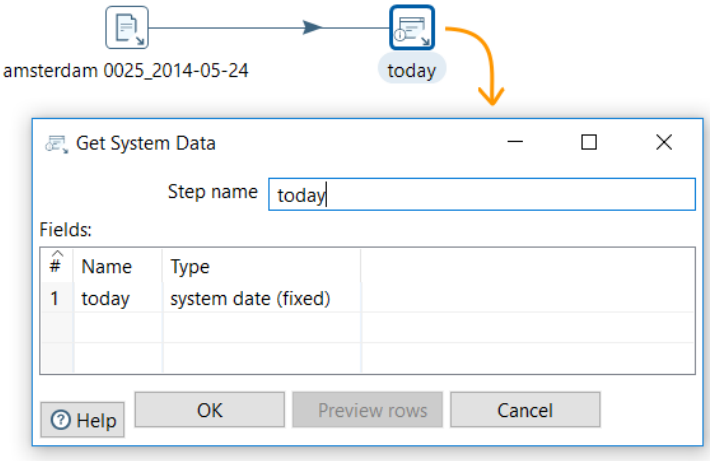
Step fields and their origin

Step name: Injuries

Fields:

#	Fieldname	Type	Length	Precision	Step origin	Storage
1	full_name	String	100	-	Injuries	normal
2	injury_type	String	100	-	Injuries	normal
3	injury_date	String	2147483647	-	Injuries	normal

Edit origin step | Cancel

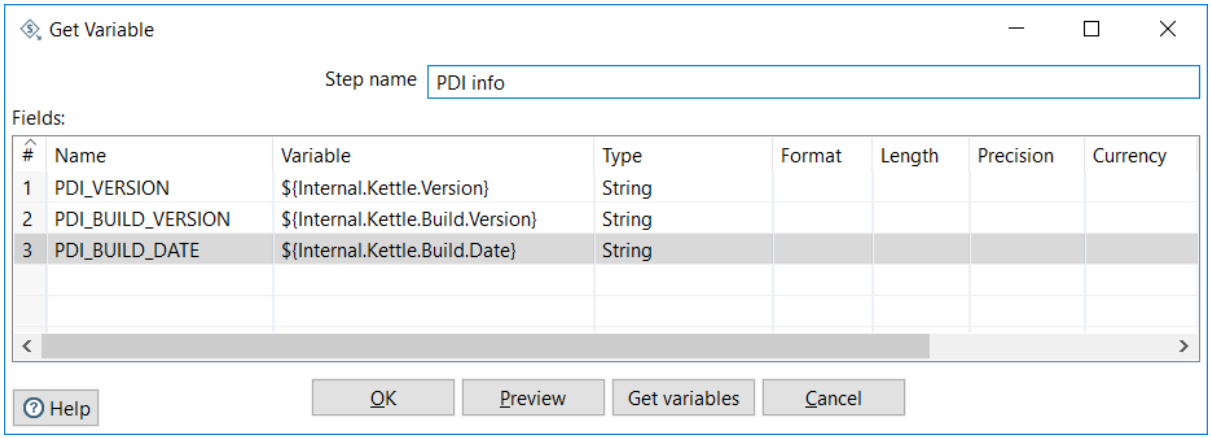


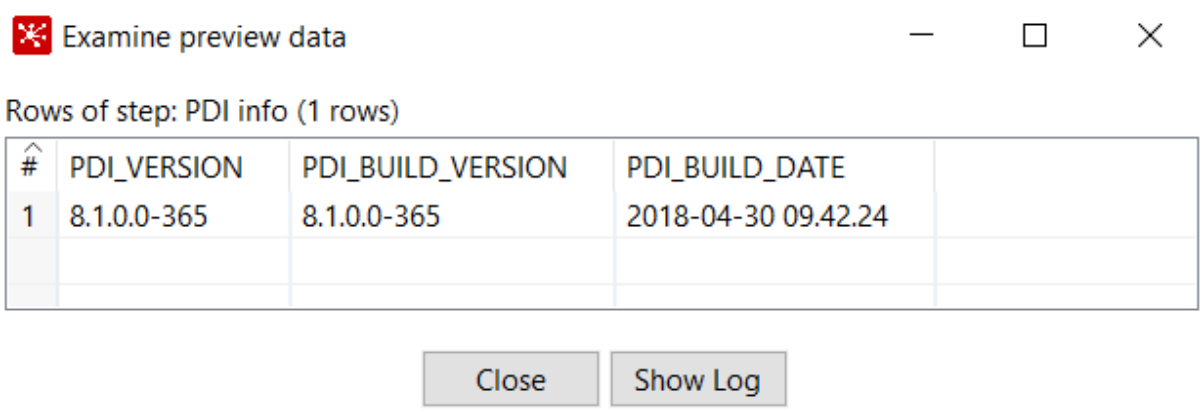
Examine preview data

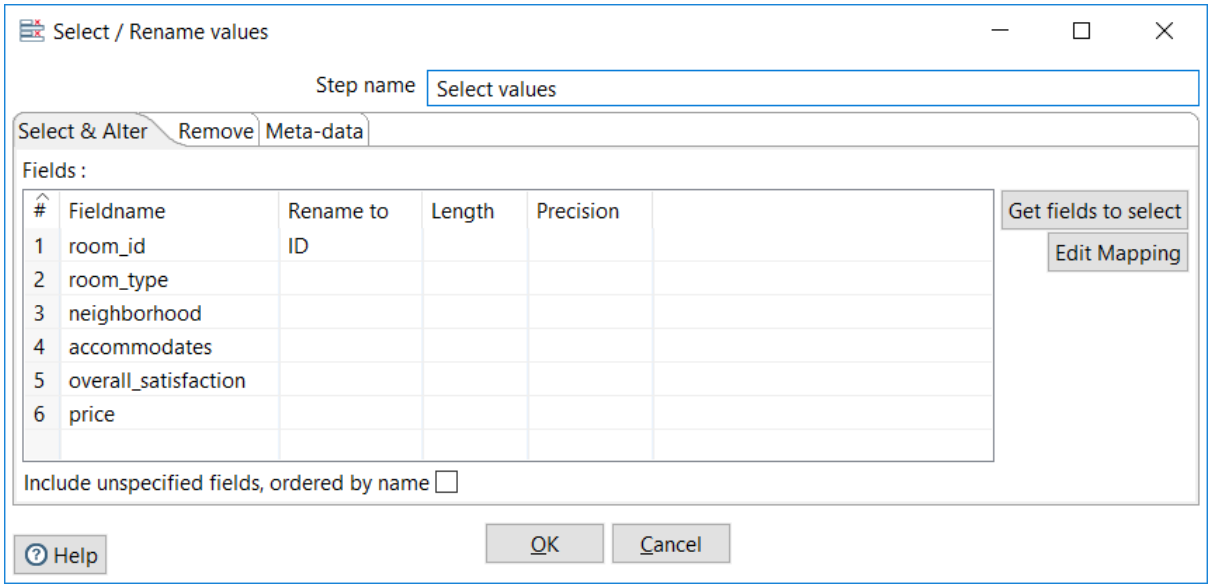
Rows of step: today (1000 rows)

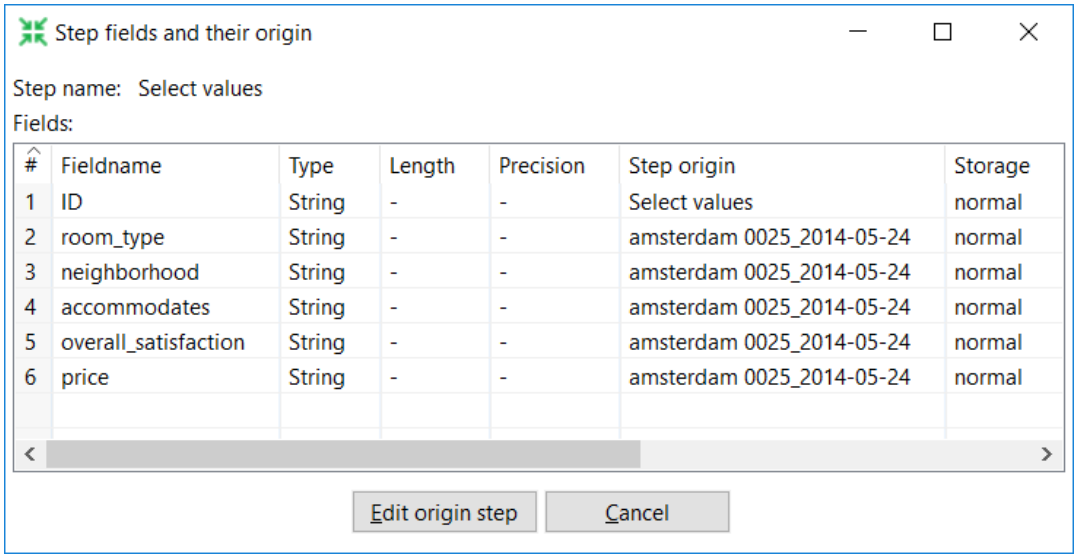
#	roo...	host_id	room_type	bo...	neighborhood	reviews	ov...	acco...	be...	price	m...	latit	longi...	last_modified	today
1	2818	3159	Private ro	<...	Oostelijk Haven...	77	5.0	2	1.0	90.0	3	52.3...	4.94...	2014-05-24 14:...	2018/06/18 10:41:11.920
2	6243	3813	Private ro	<...	Oostelijk Haven...	116	4.5	4	1.0	75.0	2	52.3...	4.92...	2014-05-24 17:...	2018/06/18 10:41:11.920
3	150	59297	Private ro	<...	Westerpark	53	5.0	4	1.0	66.0	2	52.3...	4.87	2014-05-24 20:...	2018/06/18 10:41:11.920
4	158	59484	Private ro	<...	Centrum Oost	1	5.0	4	1.0	151.0	3	52.3...	4.89...	2014-05-24 21:...	2018/06/18 10:41:11.920
5	201	59484	Private ro	<...	Centrum Oost	4	4.0	2	1.0	90.0	3	52.3...	4.89...	2014-05-24 20:...	2018/06/18 10:41:11.920
6	205	77484	Entire ho...	<...	De Pijp / Rivier...	139	4.5	2	1.0	92.0	2	52.3...	4.88...	2014-05-24 15:...	2018/06/18 10:41:11.920
7	223	10378	Entire ho...	<...	Oud Oost	6	4.0	6	2.0	153.0	3	52.3...	4.91...	2014-05-24 15:...	2018/06/18 10:41:11.920
8	236	9616	Private ro	<...	De Pijp / Rivier...	36	5.0	3	1.0	122.0	6	52.3...	4.89...	2014-05-24 17:...	2018/06/18 10:41:11.920
9	254	106225	Entire ho...	<...	Oostelijk Haven...	46	4.5	6	3.0	229.0	3	52.3...	4.94...	2014-05-24 18:...	2018/06/18 10:41:11.920
1	275	119217	Private ro	<...	Centrum Oost	53	4.5	2	1.0	174.0	3	52.3...	4.89...	2014-05-24 22:...	2018/06/18 10:41:11.920

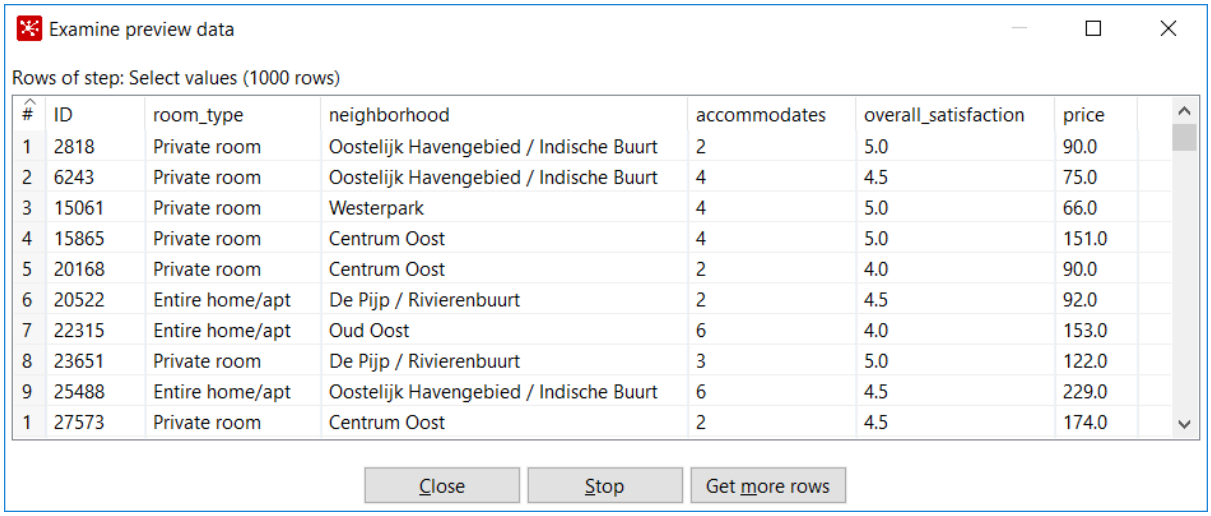
Close | Stop | Get more rows

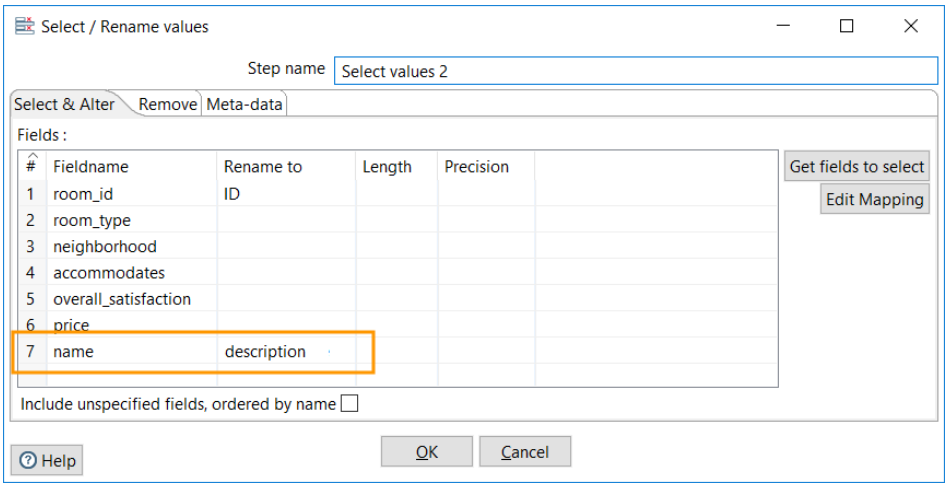

Get Variable
Step name
PDI info
Fields:
Name Variable Type Format Length Precision Currency
1 PDI_VERSION ${Internal.Kettle.Version} String
2 PDI_BUILD_VERSION ${Internal.Kettle.Build.Version} String
3 PDI_BUILD_DATE ${Internal.Kettle.Build.Date} String
Help
OK
Preview
Get variables
Cancel

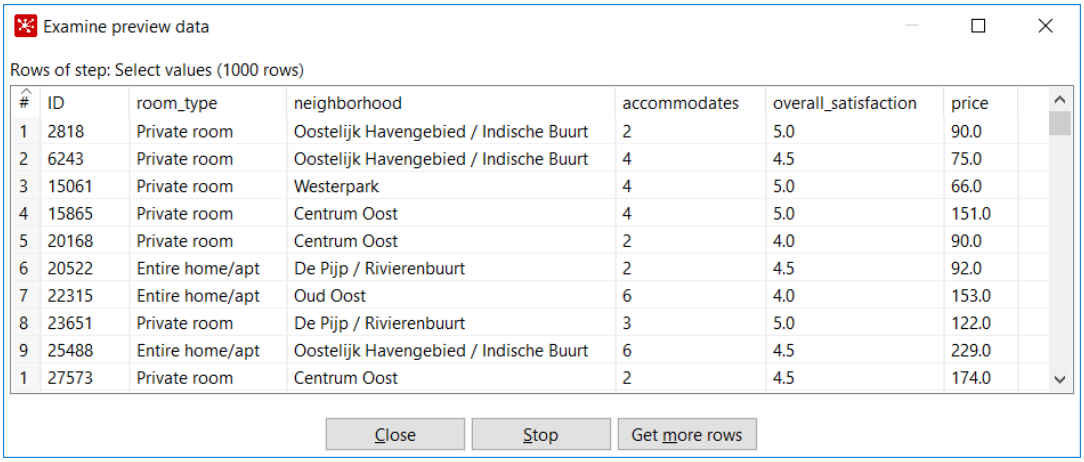

Examine preview data
Rows of step: PDI info (1 rows)
PDI_VERSION PDI_BUILD_VERSION PDI_BUILD_DATE
1 8.1.0.0-365 8.1.0.0-365 2018-04-30 09.42.24
Close
Show Log


Select / Rename values
Step name
Select values
Select & Alter
Remove
Meta-data
Fields :
Fieldname Rename to Length Precision
1 room_id ID
2 room_type
3 neighborhood
4 accommodates
5 overall_satisfaction
6 price
Get fields to select
Edit Mapping
Include unspecified fields, ordered by name
Help
OK
Cancel

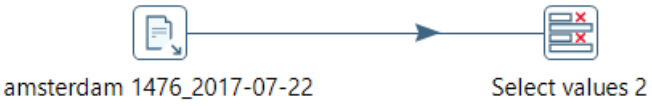

Step fields and their origin
Step name: Select values
Fields:
Fieldname Type Length Precision Step origin Storage
1 ID String - - Select values normal
2 room_type String - - amsterdam 0025_2014-05-24 normal
3 neighborhood String - - amsterdam 0025_2014-05-24 normal
4 accommodates String - - amsterdam 0025_2014-05-24 normal
5 overall_satisfaction String - - amsterdam 0025_2014-05-24 normal
6 price String - - amsterdam 0025_2014-05-24 normal
Edit origin step
Cancel

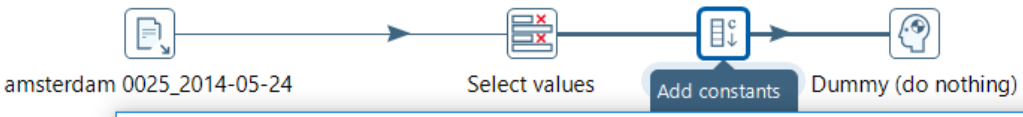

Examine preview data
Rows of step: Select values (1000 rows)
ID room_type neighborhood accommodates overall_satisfaction price
1 2818 Private room Oostelijk Havengebied / Indische Buurt 2 5.0 90.0
2 6243 Private room Oostelijk Havengebied / Indische Buurt 4 4.5 75.0
3 15061 Private room Westerpark 4 5.0 66.0
4 15865 Private room Centrum Oost 4 5.0 151.0
5 20168 Private room Centrum Oost 2 4.0 90.0
6 20522 Entire home/apt De Pijp / Rivierenbuurt 2 4.5 92.0
7 22315 Entire home/apt Oud Oost 6 4.0 153.0
8 23651 Private room De Pijp / Rivierenbuurt 3 5.0 122.0
9 25488 Entire home/apt Oostelijk Havengebied / Indische Buurt 6 4.5 229.0
1 27573 Private room Centrum Oost 2 4.5 174.0
Close
Stop
Get more rows

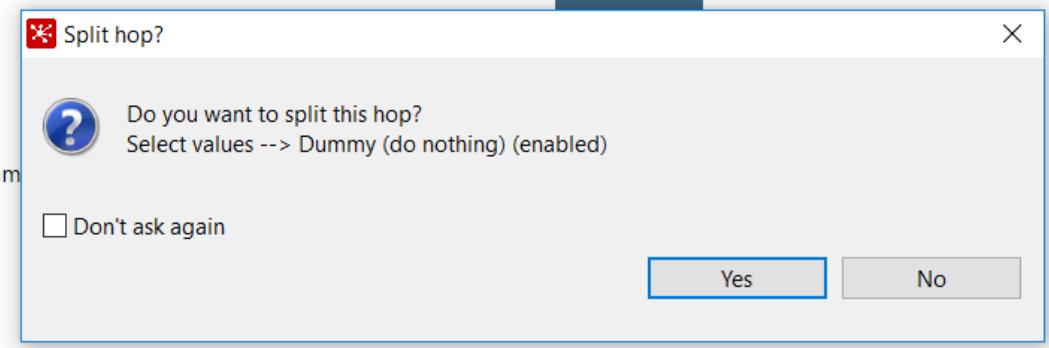

Select / Rename values
Step name
Select values 2
Select & Alter
Remove
Meta-data
Fields :
Fieldname Rename to Length Precision
1 room_id ID
2 room_type
3 neighborhood
4 accommodates
5 overall_satisfaction
6 price
7 name description
Get fields to select
Edit Mapping
Include unspecified fields, ordered by name
Help
OK
Cancel

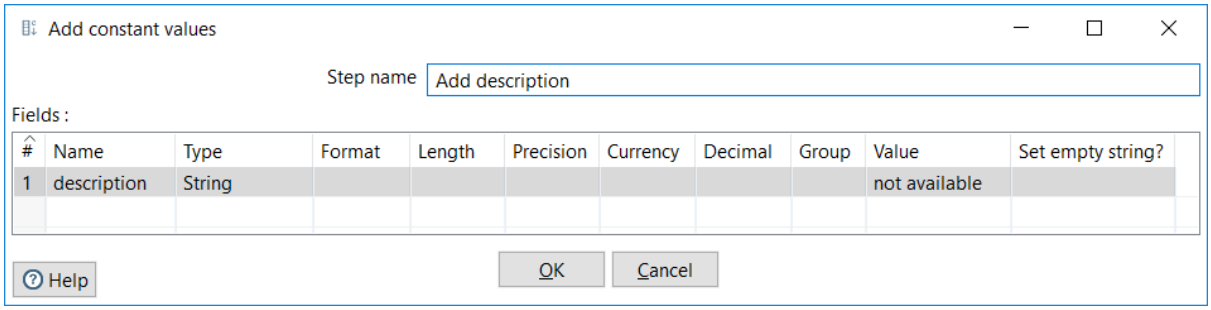

Examine preview data
Rows of step: Select values (1000 rows)
ID room_type neighborhood accommodates overall_satisfaction price
1 2818 Private room Oostelijk Havengebied / Indische Buurt 2 5.0 90.0
2 6243 Private room Oostelijk Havengebied / Indische Buurt 4 4.5 75.0
3 15061 Private room Westerpark 4 5.0 66.0
4 15865 Private room Centrum Oost 4 5.0 151.0
5 20168 Private room Centrum Oost 2 4.0 90.0
6 20522 Entire home/apt De Pijp / Rivierenbuurt 2 4.5 92.0
7 22315 Entire home/apt Oud Oost 6 4.0 153.0
8 23651 Private room De Pijp / Rivierenbuurt 3 5.0 122.0
9 25488 Entire home/apt Oostelijk Havengebied / Indische Buurt 6 4.5 229.0
1 27573 Private room Centrum Oost 2 4.5 174.0
Close
Stop
Get more rows

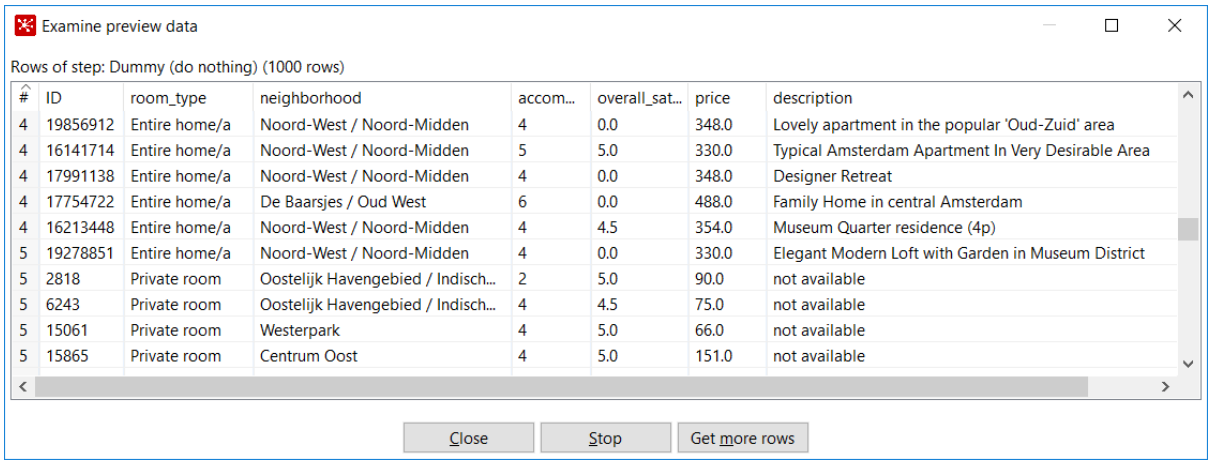
amsterdam 0025_2014-05-24
Select values
Dummy (do nothing)


amsterdam 1476_2017-07-22
Select values 2


amsterdam 0025_2014-05-24
Select values
Add constants
Dummy (do nothing)

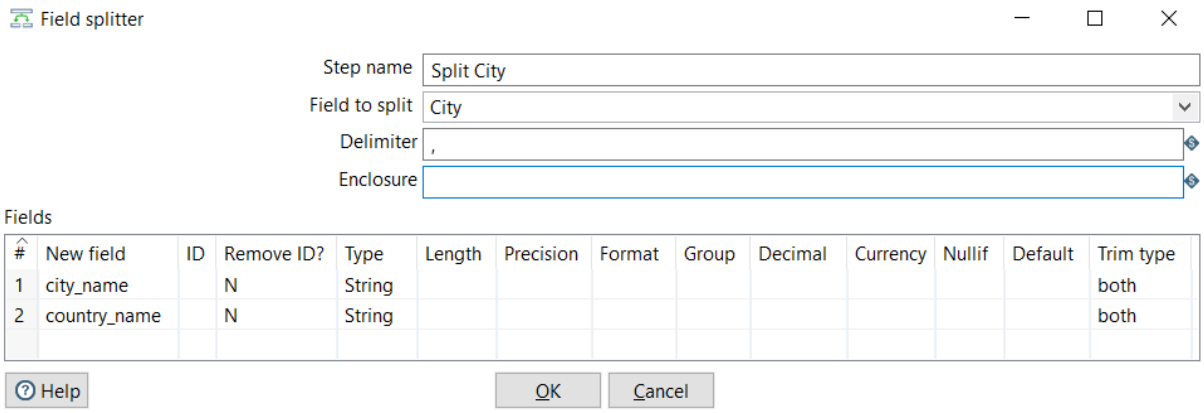

Split hop?
Do you want to split this hop?
Select values --> Dummy (do nothing) (enabled)
Don't ask again
Yes
No

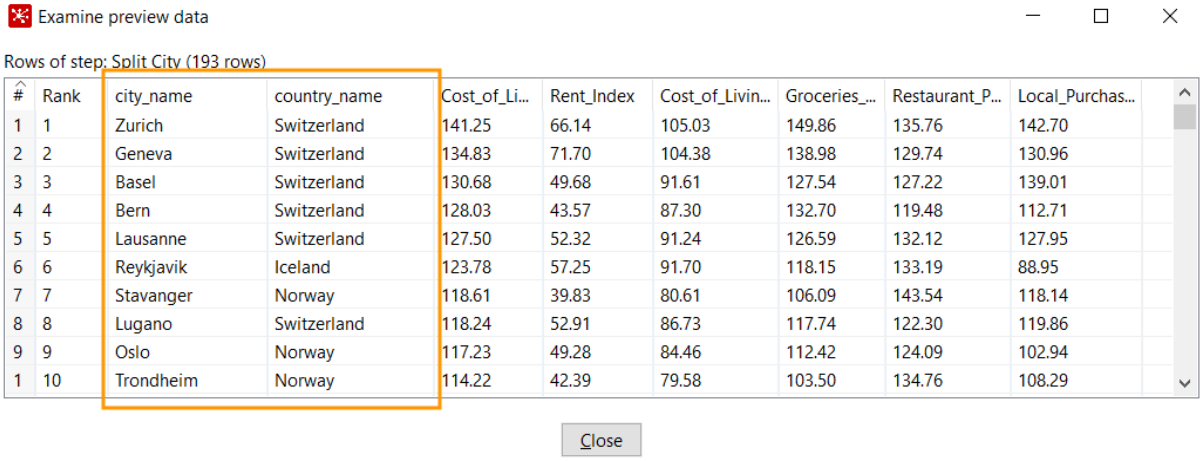


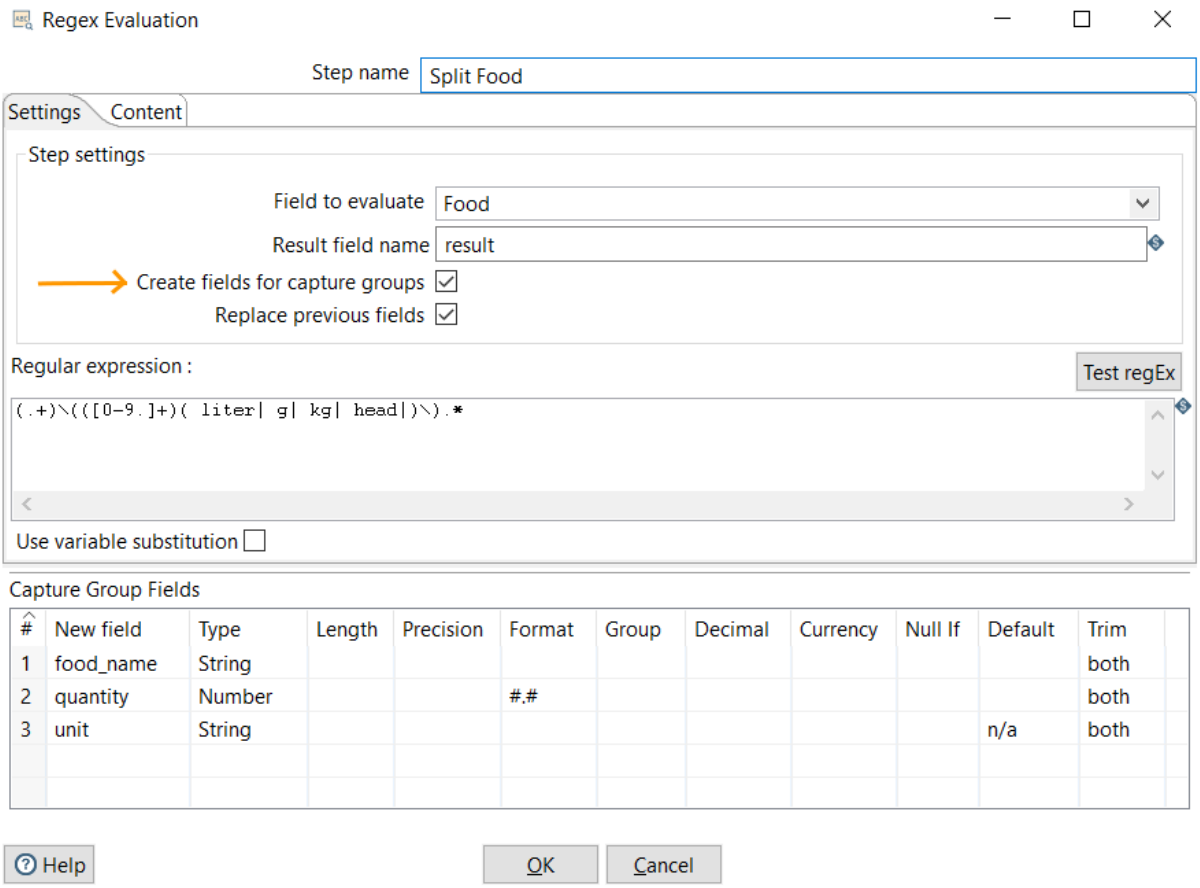


#	ID	room_type	neighborhood	accom...	overall_sat...	price	description
4	19856912	Entire home/a	Noord-West / Noord-Midden	4	0.0	348.0	Lovely apartment in the popular 'Oud-Zuid' area
4	16141714	Entire home/a	Noord-West / Noord-Midden	5	5.0	330.0	Typical Amsterdam Apartment In Very Desirable Area
4	17991138	Entire home/a	Noord-West / Noord-Midden	4	0.0	348.0	Designer Retreat
4	17754722	Entire home/a	De Baarsjes / Oud West	6	0.0	488.0	Family Home in central Amsterdam
4	16213448	Entire home/a	Noord-West / Noord-Midden	4	4.5	354.0	Museum Quarter residence (4p)
5	19278851	Entire home/a	Noord-West / Noord-Midden	4	0.0	330.0	Elegant Modern Loft with Garden in Museum District
5	2818	Private room	Oostelijk Havengebied / Indisch...	2	5.0	90.0	not available
5	6243	Private room	Oostelijk Havengebied / Indisch...	4	4.5	75.0	not available
5	15061	Private room	Westerpark	4	5.0	66.0	not available
5	15865	Private room	Centrum Oost	4	5.0	151.0	not available

Chapter 4: Transforming Data






Regex Evaluation
Step name
Split Food
Settings
Content
Step settings
Field to evaluate
Food
Result field name
result
Create fields for capture groups
Replace previous fields
Regular expression :
Test regEx
(.+)\(([0-9.]+)(liter| g| kg| head|)\).*
Use variable substitution
Capture Group Fields
New field Type Length Precision Format Group Decimal Currency Null If Default Trim
1 food_name String both
2 quantity Number #.# both
3 unit String n/a both
Help
OK
Cancel

Examine preview data
Rows of step: Split Food (14 rows)
Food Price result food_name quantity unit
1 Milk (regular), (0.25 liter) .19 € Y Milk (regular), 0.2 liter
2 Loaf of Fresh White Bread (125.00 g) .24 € Y Loaf of Fresh White Bread 125 g
3 Rice (white), (0.10 kg) .09 € Y Rice (white), 0.1 kg
4 Eggs (regular) (2.40) .33 € Y Eggs (regular) 2.4 n/a
5 Local Cheese (0.10 kg) .89 € Y Local Cheese 0.1 kg
6 Chicken Breasts (Boneless, Skinless), (0.15 .86 € Y Chicken Breasts (Boneless, Skinless), 0.1 kg
7 Beef Round (0.15 kg) (or Equivalent Back ... 1.52 € Y Beef Round 0.1 kg
8 Apples (0.30 kg) .47 € Y Apples 0.3 kg
Close

Calculator

Step name: food_index

Fields:

#	New field	Calculation	Field A	Field B	Field C	Value type	Length	Precision	Remove
1	two	Set field to constant value A	2			Integer			N
2	temp	A + B	Groceries_Index	Restaurant_Price_Index		Number			N
3	food_index	A / B	temp	two		Number			N

Help OK Cancel

Examine preview data

Rows of step: food_index (193 rows)

#	Rank	city_name	country_name	Cost_of_L...	Rent_Index	Cost_of_Li...	Groceries_Index	Restaurant_Price_Index	Local_Purc...	two	temp	food_index
1	1	Zurich	Switzerland	141.25	66.14	105.03	149.86	135.76	142.7	2	285.62	142.81
2	2	Geneva	Switzerland	134.83	71.7	104.38	138.98	129.74	130.96	2	268.72	134.36
3	3	Basel	Switzerland	130.68	49.68	91.61	127.54	127.22	139.01	2	254.76	127.38
4	4	Bern	Switzerland	128.03	43.57	87.3	132.7	119.48	112.71	2	252.18	126.09
5	5	Lausanne	Switzerland	127.5	52.32	91.24	126.59	132.12	127.95	2	258.71	129.355
6	6	Reykjavik	Iceland	123.78	57.25	91.7	118.15	133.19	88.95	2	251.34	125.67
7	7	Stavanger	Norway	118.61	39.83	80.61	106.09	143.54	118.14	2	249.63	124.815
8	8	Lugano	Switzerland	118.24	52.91	86.73	117.74	122.3	119.86	2	240.04	120.02

Close

User Defined Java Expression

Step name: food_index

Fields:

#	New field	Java expression	Value type	Length	Precision	Replace value
1	food_index	(Groceries_Index + Restaurant_Price_Index)/2	Number			

Help OK Cancel

Sort rows

Step name: Sort rows

Sort directory: %%java.io.tmpdir%% Browse...

TMP-file prefix: out

Sort size (rows in memory): 1000000

Free memory threshold (in %):

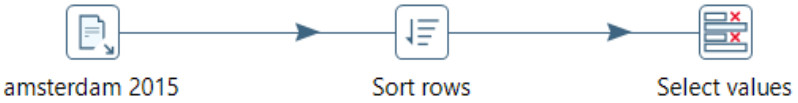
Compress TMP Files? ☐

Only pass unique rows? (verifies keys only) ☐

Fields :

#	Fieldname	Ascending	Case sensitive compare?	Sort based on current locale?	Collator Strength	Presorted?
1	neighborhood	Y	N	N	0	N
2	room_type	Y	N	N	0	N
3	reviews	N	N	N	0	N

Help OK Cancel Get Fields



Select / Rename values

Step name: Select values

Select & Alter | Remove | Meta-data

Fields :

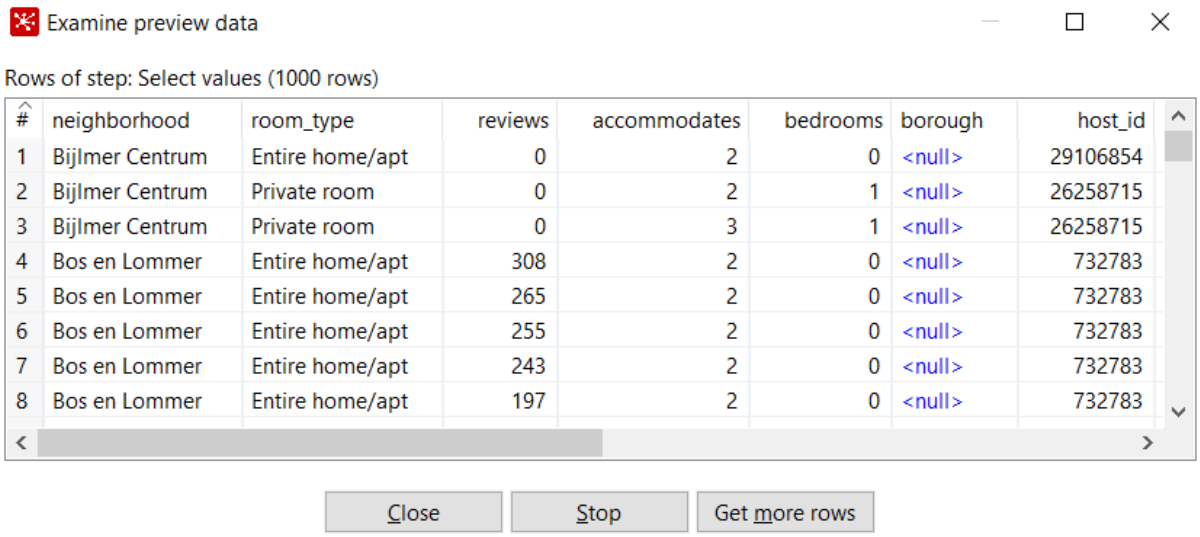
#	Fieldname	Rename to	Length	Precision
1	neighborhood			
2	room_type			
3	reviews			

Get fields to select

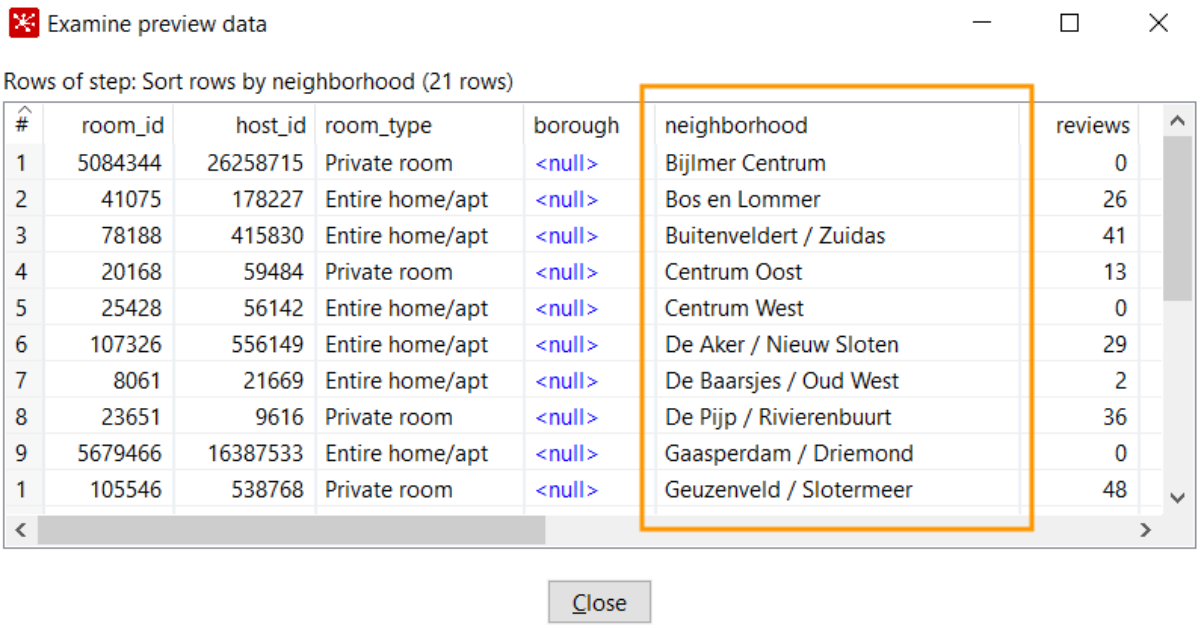
Edit Mapping

Include unspecified fields, ordered by name ☑

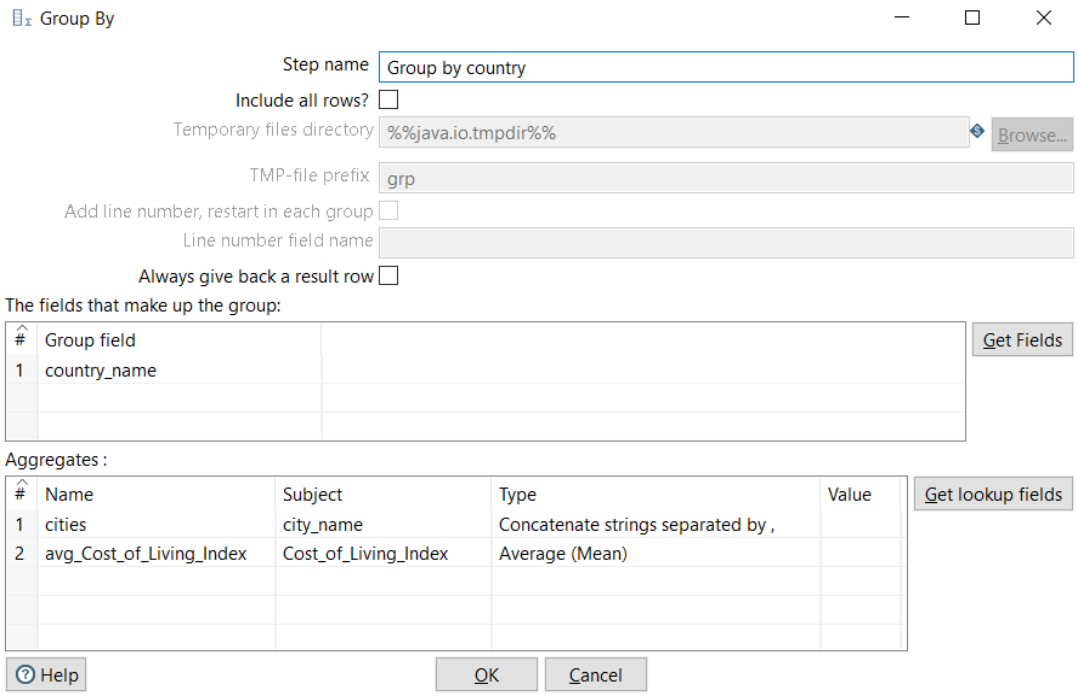
Help OK Cancel

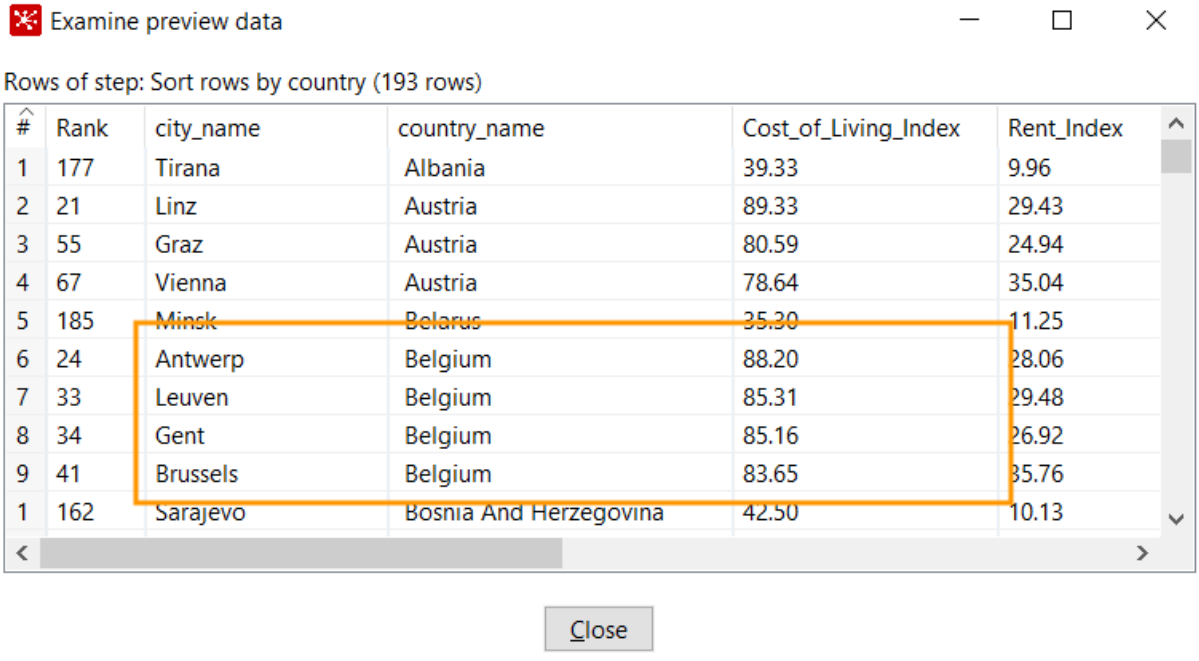


#	neighborhood	room_type	reviews	accommodates	bedrooms	borough	host_id
1	Bijlmer Centrum	Entire home/apt	0	2	0	<null>	29106854
2	Bijlmer Centrum	Private room	0	2	1	<null>	26258715
3	Bijlmer Centrum	Private room	0	3	1	<null>	26258715
4	Bos en Lommer	Entire home/apt	308	2	0	<null>	732783
5	Bos en Lommer	Entire home/apt	265	2	0	<null>	732783
6	Bos en Lommer	Entire home/apt	255	2	0	<null>	732783
7	Bos en Lommer	Entire home/apt	243	2	0	<null>	732783
8	Bos en Lommer	Entire home/apt	197	2	0	<null>	732783



#	room_id	host_id	room_type	borough	neighborhood	reviews
1	5084344	26258715	Private room	<null>	Bijlmer Centrum	0
2	41075	178227	Entire home/apt	<null>	Bos en Lommer	26
3	78188	415830	Entire home/apt	<null>	Buitenveldert / Zuidas	41
4	20168	59484	Private room	<null>	Centrum Oost	13
5	25428	56142	Entire home/apt	<null>	Centrum West	0
6	107326	556149	Entire home/apt	<null>	De Aker / Nieuw Sloten	29
7	8061	21669	Entire home/apt	<null>	De Baarsjes / Oud West	2
8	23651	9616	Private room	<null>	De Pijp / Rivierenbuurt	36
9	5679466	16387533	Entire home/apt	<null>	Gaasperdam / Driemond	0
1	105546	538768	Private room	<null>	Geuzenveld / Slotermeer	48


Group By
Step name
Group by country
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 country_name
Get Fields
Aggregates :
Name Subject Type Value
1 cities city_name Concatenate strings separated by ,
2 avg_Cost_of_Living_Index Cost_of_Living_Index Average (Mean)
Get lookup fields
Help
OK
Cancel


Examine preview data
Rows of step: Sort rows by country (193 rows)
Rank city_name country_name Cost_of_Living_Index Rent_Index
1 177 Tirana Albania 39.33 9.96
2 21 Linz Austria 89.33 29.43
3 55 Graz Austria 80.59 24.94
4 67 Vienna Austria 78.64 35.04
5 185 Minsk Belarus 35.30 11.25
6 24 Antwerp Belgium 88.20 28.06
7 33 Leuven Belgium 85.31 29.48
8 34 Gent Belgium 85.16 26.92
9 41 Brussels Belgium 83.65 35.76
1 162 Sarajevo Bosnia And Herzegovina 42.50 10.13
Close

Examine preview data

Rows of step: Group by country (40 rows)

#	country_name	cities	avg_Cost_of_Living_Index
1	Albania	Tirana	39.33
2	Austria	Linz, Graz, Vienna	82.8533333333
3	Belarus	Minsk	35.3
4	Belgium	Antwerp, Leuven, Gent, Brussels	85.58
5	Bosnia And Herzegovina	Sarajevo, Banja Luka	41.175
6	Bulgaria	Sofia, Varna, Burgas, Plovdiv	41.6175
7	Croatia	Split, Rijeka, Zagreb, Osijek	56.1575
8	Czech Republic	Prague, Brno, Olomouc, Ostrava	48.725
9	Denmark	Copenhagen, Arhus, Aalborg	93.5633333333
1	Estonia	Tallinn, Tartu	58.0
1	Finland	Helsinki, Tampere, Turku	85.7166666667

Close

Group By

Step name: Total

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	

Leave this section empty

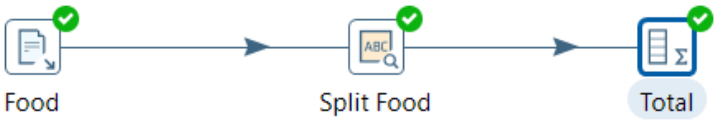
Get Fields

Aggregates :

#	Name	Subject	Type	Value
1	total	Price	Sum	

Get lookup fields

Help OK Cancel



Examine preview ...

Rows of step: Total (1 rows)

#	total
1	6.07

Close

Filter rows

Step name: Filter bedrooms >3

Send 'true' data to step:

Send 'false' data to step:

The condition:

To edit a subcondition, simply click on it

bedrooms > [3]

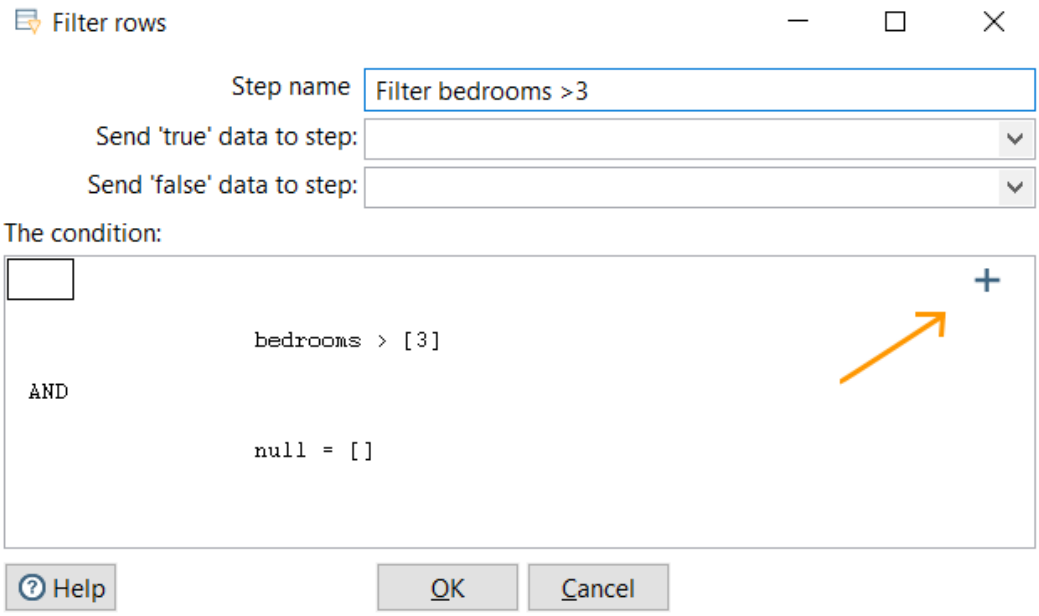
Help OK Cancel

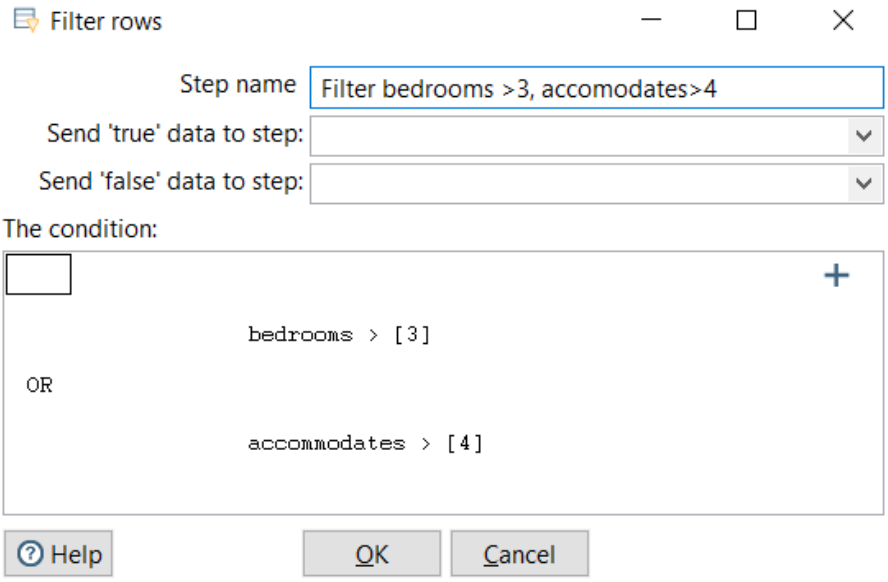
Examine preview data

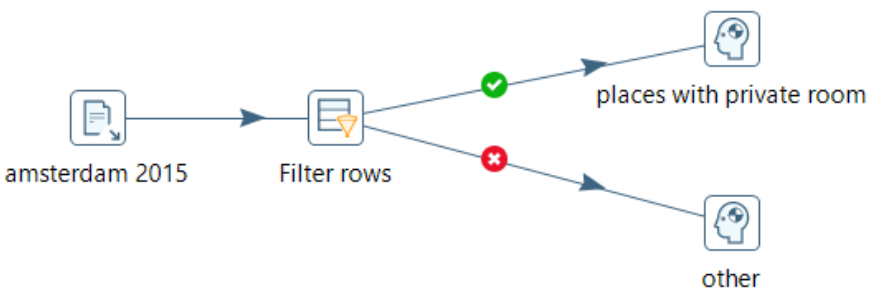
Rows of step: preview (644 rows)

#	room_id	host_id	room_type	borou	neighborhood	reviews	overall_satisfaction	accommodates	bedrooms	price	minsta
1	180424	864444	Entire home/apt	<null>	Centrum West	62	4.5	6	4	321	
2	218994	888292	Entire home/apt	<null>	Noord-West / Noord-Midden	84	5	6	4	271	
3	256670	1350388	Entire home/apt	<null>	De Baarsjes / Oud West	3	4.5	5	4	178	
4	283170	1451657	Entire home/apt	<null>	Noord-West / Noord-Midden	41	4.5	16	5	357	
5	356929	728160	Entire home/apt	<null>	Watergraafsmeer	29	5	4	5	250	
6	438688	2178991	Entire home/apt	<null>	Buitenveldert / Zuidas	2	5	7	4	286	
7	510836	1826139	Entire home/apt	<null>	Centrum West	46	5	15	5	714	
8	513743	2531171	Entire home/apt	<null>	Oostelijk Havengebied / Indi...	12	5	9	5	500	

Close


Filter rows
Step name
Filter bedrooms >3
Send 'true' data to step:
Send 'false' data to step:
The condition:
bedrooms > [3]
AND
null = []
Help
OK
Cancel


Filter rows
Step name
Filter bedrooms >3, accomodates>4
Send 'true' data to step:
Send 'false' data to step:
The condition:
bedrooms > [3]
OR
accommodates > [4]
Help
OK
Cancel


amsterdam 2015
Filter rows
places with private room
other

Examine preview data

Rows of step: places with private room (1000 rows)

#	room_id	host_id	room_type	borough	neighborhood	reviews	c
1	2818	3159	Private room	<null>	Oostelijk Havengebied / Indische Buurt	102	
2	7231	18481	Private room	<null>	Watergraafsmeer	17	
3	15061	59297	Private room	<null>	Westerpark	58	
4	20168	59484	Private room	<null>	Centrum Oost	13	
5	23651	9616	Private room	<null>	De Pijp / Rivierenbuurt	36	
6	28658	123414	Private room	<null>	De Baarsjes / Oud West	207	
7	28871	124245	Private room	<null>	Centrum Oost	2	
8	29051	124245	Private room	<null>	Centrum Oost	85	

Close | Stop | Get more rows

Execution Results

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

#	Stepname	Copyn	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	input/output
1	amsterdam 2015	0	0	25350	25353	0	3	0	0	Finished	0.4s	60,221	-
2	Filter rows	0	25350	25350	0	0	0	0	0	Finished	0.4s	59,368	-
3	other	0	20654	20654	0	0	0	0	0	Finished	0.5s	44,900	-
4	places with private room	0	4696	4696	0	0	0	0	0	Finished	0.5s	10,209	-

Examine preview data

Rows of step: cities in Europe (34 rows)

#	rank	City	Country	Official_popu...	Date	2011_Eurostat_p...	Image	Location	Ref.
1	1	Istanbul[a]	Turkey	15,029,231	31 December 20	<null>	Halic.png	41.013611°N 28.955°E	[4]
2	2	Moscow[b]	Russia	13,197,596	1 January 2017	<null>	Moscow-City2015.jpg	55.75°N 37.616667°E	[5]
3	3	London	United Kingdom	8,787,892	30 June 2016	8,173,941	London from a hot air balloon.jpg	51.507222°N 0.1275°W	[6]
4	4	Saint Petersburg	Russia	5,323,300	1 January 2017	<null>	Saint Isaac's Square SPB (01).jpg	59.95°N 30.3°E	[7][8]
5	5	Berlin	Germany	3,671,000	31 December 20	3,460,725	Siegessaeule Aussicht 10-13 img4 Tierg...	52.516667°N 13.383333°E	[9]
6	6	Madrid	Spain	3,182,981	1 January 2017	3,198,645	Ejército del Aire (25408739923).jpg	40.383333°N 3.716667°W	[10]
7	7	Kiev	Ukraine	2,934,401	1 February 2018	<null>	Lavra KyivJPG	50.45°N 30.523333°E	[11]
8	8	Rome	Italy	2,879,038	19 October 2017	2,873,494	Roma dall'aereo.JPG	41.9°N 12.5°E	[12]
9	9	Paris	France	2,241,346	1 January 2014	2,249,977	Sunset over Paris 5, France August 2013.j	48.8567°N 2.3508°E	[13]
1	10	Bucharest	Romania	2,106,144	1 January 2016	1,903,299	Bucuresti de sus.jpg	44.4325°N 26.103889°E	[14] [15]

Close | Show Log

Examine preview data

Rows of step: cities in Europe (34 rows)

#	rank	City	Country	Official_popu...	Date	2011_Eurostat_p...	Image	Location	Ref.
1	1	Istanbul[a]	Turkey	15,029,231	31 December 20	<null>	Halic.png	41.013611°N 28.955°E	[4]
2	2	Moscow[b]	Russia	13,197,596	1 January 2017	<null>	Moscow-City2015.jpg	55.75°N 37.616667°E	[5]
3	3	London	United Kingdom	8,787,892	30 June 2016	8,173,941	London from a hot air balloon.jpg	51.507222°N 0.1275°W	[6]
4	4	Saint Petersburg	Russia	5,323,300	1 January 2017	<null>	Saint Isaac's Square SPB (01).jpg	59.95°N 30.3°E	[7][8]
5	5	Berlin	Germany	3,671,000	31 December 20	3,460,725	Siegessaeule Aussicht 10-13 img4 Tierg...	52.516667°N 13.383333°E	[9]
6	6	Madrid	Spain	3,182,981	1 January 2017	3,198,645	Ejército del Aire (25408739923).jpg	40.383333°N 3.716667°W	[10]
7	7	Kiev	Ukraine	2,934,401	1 February 2018	<null>	Lavra KyivJPG	50.45°N 30.523333°E	[11]
8	8	Rome	Italy	2,879,038	19 October 2017	2,873,494	Roma dall'aereo.JPG	41.9°N 12.5°E	[12]
9	9	Paris	France	2,241,346	1 January 2014	2,249,977	Sunset over Paris 5, France August 2013.j	48.8567°N 2.3508°E	[13]
1	10	Bucharest	Romania	2,106,144	1 January 2016	1,903,299	Bucuresti de sus.jpg	44.4325°N 26.103889°E	[14] [15]

Close | Show Log

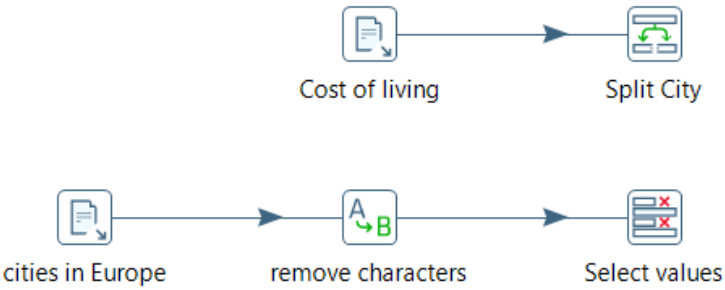
Replace in string

Step name: remove characters

Fields string

#	In stream field	Out stream field	use RegEx	Search	Replace with	Set empty string?	Replace with...	Whole Word	Case sensitive	Is Unicode
1	City		Y	\[.+\]		N		N	N	N

Help | OK | Get fields | Cancel



Stream Value Lookup

Step name: Ik cost of living index

Lookup step: Split City

The key(s) to look up the value(s):

#	Field	LookupField
1	City	city_name
2	Country	country_name

Specify the fields to retrieve :

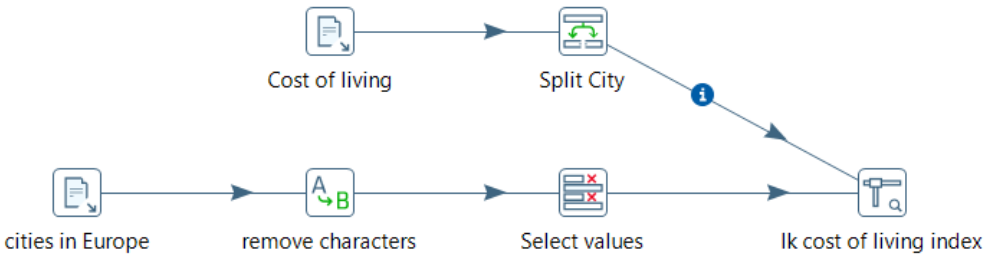
#	Field	New name	Default	Type
1	Cost_of_Living_Index			Number

Preserve memory (costs CPU) ☑

Key and value are exactly one integer ○

Use sorted list (i.s.o. hashtable) ○

Help | OK | Cancel | Get Fields | Get lookup fields



Examine preview data

Rows of step: lk cost of living index (34 rows)

#	rank	City	Country	Official_population	Cost_of_Living_Index
1	1	Istanbul	Turkey	15,029,231	<null>
2	2	Moscow	Russia	13,197,596	51.82
3	3	London	United Kingdom	8,787,892	88.69
4	4	Saint Petersburg	Russia	5,323,300	46.41
5	5	Berlin	Germany	3,671,000	74.32
6	6	Madrid	Spain	3,182,981	67.63
7	7	Kiev	Ukraine	2,934,401	27.52
8	8	Rome	Italy	2,879,038	78.8

Close

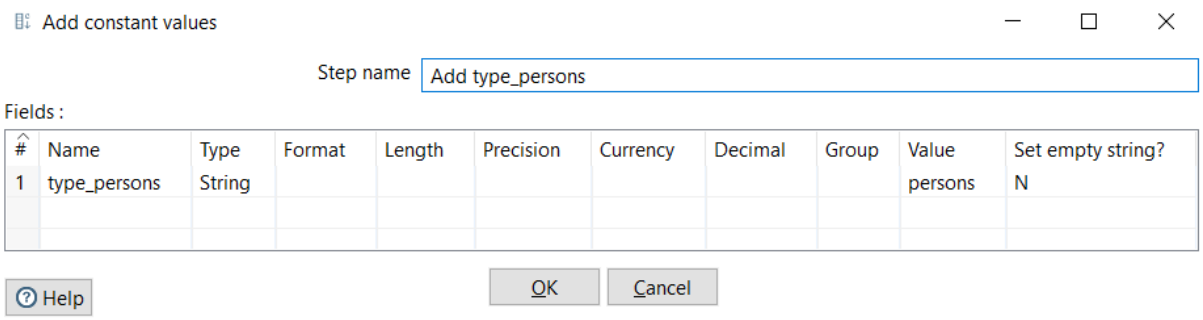
Step fields and their origin

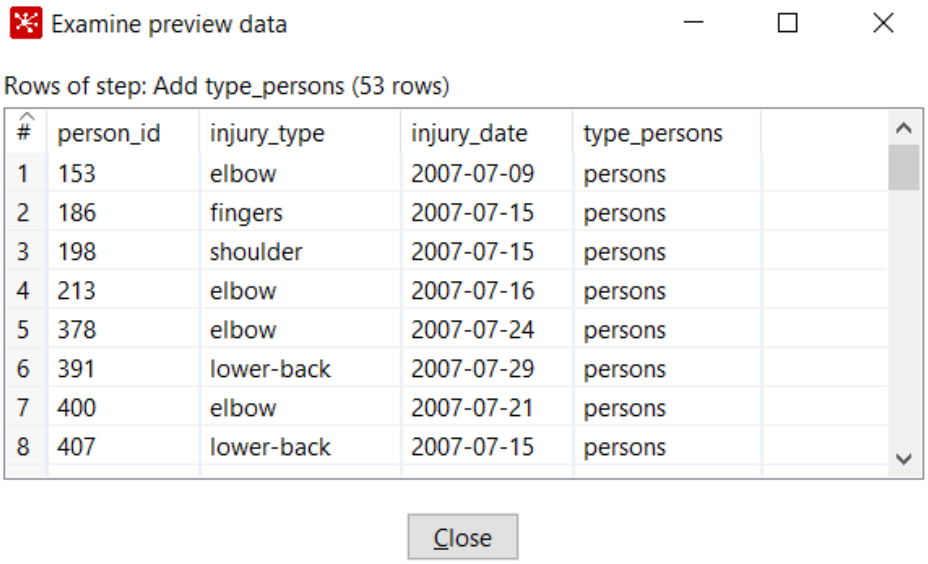
Step name: lk cost of living index

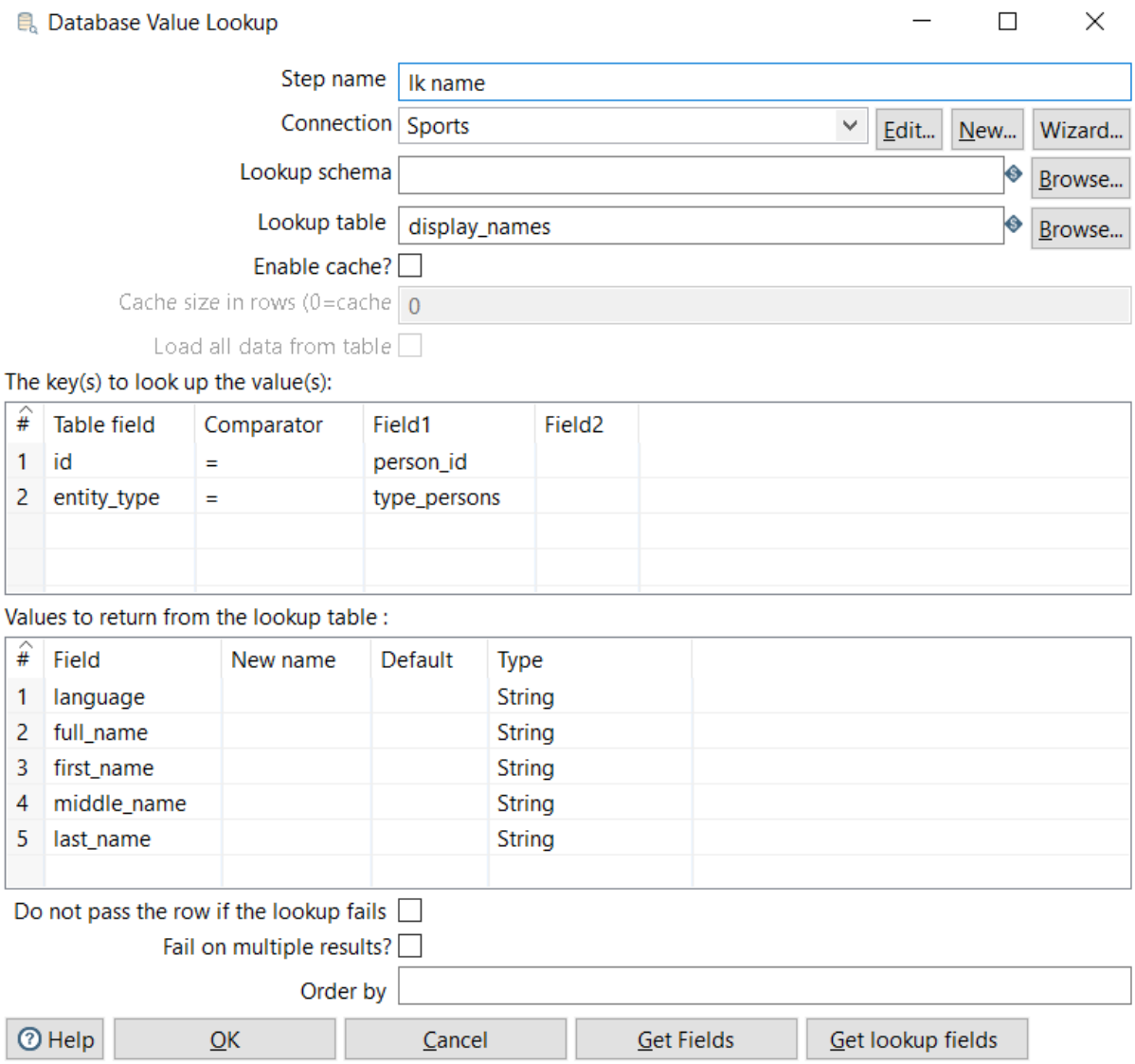
Fields:

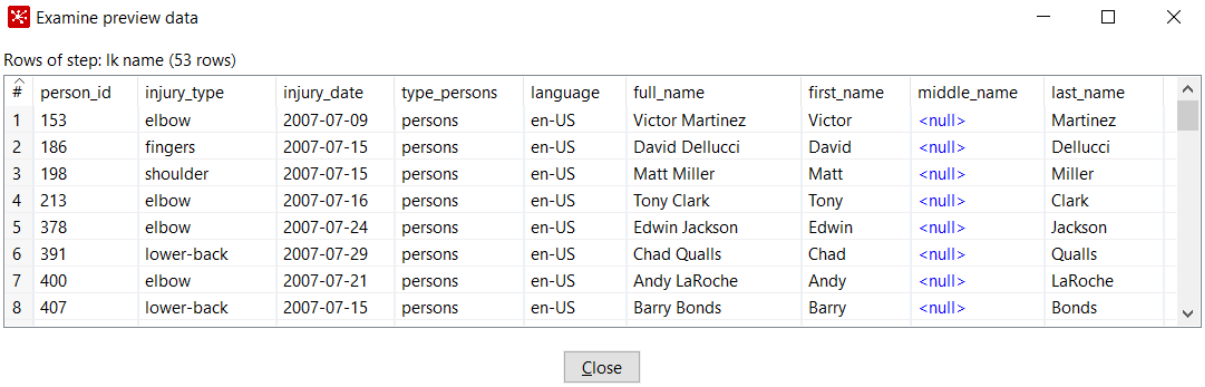
#	Fieldname	Type	Length	Precision	Step origin	Storage	Mask	Currency
1	rank	Integer	-	0	cities in Europe	normal	0	$
2	City	String	-	-	cities in Europe	normal		$
3	Country	String	-	-	cities in Europe	normal		$
4	Official_population	Number	-	-	cities in Europe	normal	#,##0.###	$
5	Cost_of_Living_Index	Number	-	-	lk cost of living index	normal	0.#####	$

Edit origin step Cancel


Add constant values
Step name
Add type_persons
Fields :
Name Type Format Length Precision Currency Decimal Group Value Set empty string?
1 type_persons String persons N
Help
OK
Cancel


Examine preview data
Rows of step: Add type_persons (53 rows)
person_id injury_type injury_date type_persons
1 153 elbow 2007-07-09 persons
2 186 fingers 2007-07-15 persons
3 198 shoulder 2007-07-15 persons
4 213 elbow 2007-07-16 persons
5 378 elbow 2007-07-24 persons
6 391 lower-back 2007-07-29 persons
7 400 elbow 2007-07-21 persons
8 407 lower-back 2007-07-15 persons
Close


Database Value Lookup
Step name
lk name
Connection
Sports
Edit...
New...
Wizard...
Lookup schema
Browse...
Lookup table
display_names
Browse...
Enable cache?
Cache size in rows (0=cache
0
Load all data from table
The key(s) to look up the value(s):
Table field Comparator Field1 Field2
1 id = person_id
2 entity_type = type_persons
Values to return from the lookup table :
Field New name Default Type
1 language String
2 full_name String
3 first_name String
4 middle_name String
5 last_name String
Do not pass the row if the lookup fails
Fail on multiple results?
Order by
Help
OK
Cancel
Get Fields
Get lookup fields


Examine preview data
Rows of step: lk name (53 rows)
person_id injury_type injury_date type_persons language full_name first_name middle_name last_name
1 153 elbow 2007-07-09 persons en-US Victor Martinez Victor <null> Martinez
2 186 fingers 2007-07-15 persons en-US David Dellucci David <null> Dellucci
3 198 shoulder 2007-07-15 persons en-US Matt Miller Matt <null> Miller
4 213 elbow 2007-07-16 persons en-US Tony Clark Tony <null> Clark
5 378 elbow 2007-07-24 persons en-US Edwin Jackson Edwin <null> Jackson
6 391 lower-back 2007-07-29 persons en-US Chad Qualls Chad <null> Qualls
7 400 elbow 2007-07-21 persons en-US Andy LaRoche Andy <null> LaRoche
8 407 lower-back 2007-07-15 persons en-US Barry Bonds Barry <null> Bonds
Close

Chapter 5: Loading Data

Text file output

Step name: listing

File | Content | Fields

#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim Type	Null
1	room_id	Integer	0						both	
2	room_type	String							both	
3	neighborhood	String							both	
4	price	Number	$0.00						both	
5	minstay	Integer	0						both	
6	last_modified	Date	yyyy-MM-dd						both	

Get Fields | Minimal width

Help | OK | Cancel

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/output
1	amsterdam 2015	0	0	25350	25353	0	3	0	0	Finished	1.3s	19,265	-
2	Java Filter	0	25350	1961	0	0	0	0	0	Finished	1.3s	19,219	-
3	listing	0	1961	1961	0	1962	0	0	0	Finished	1.3s	1,471	-

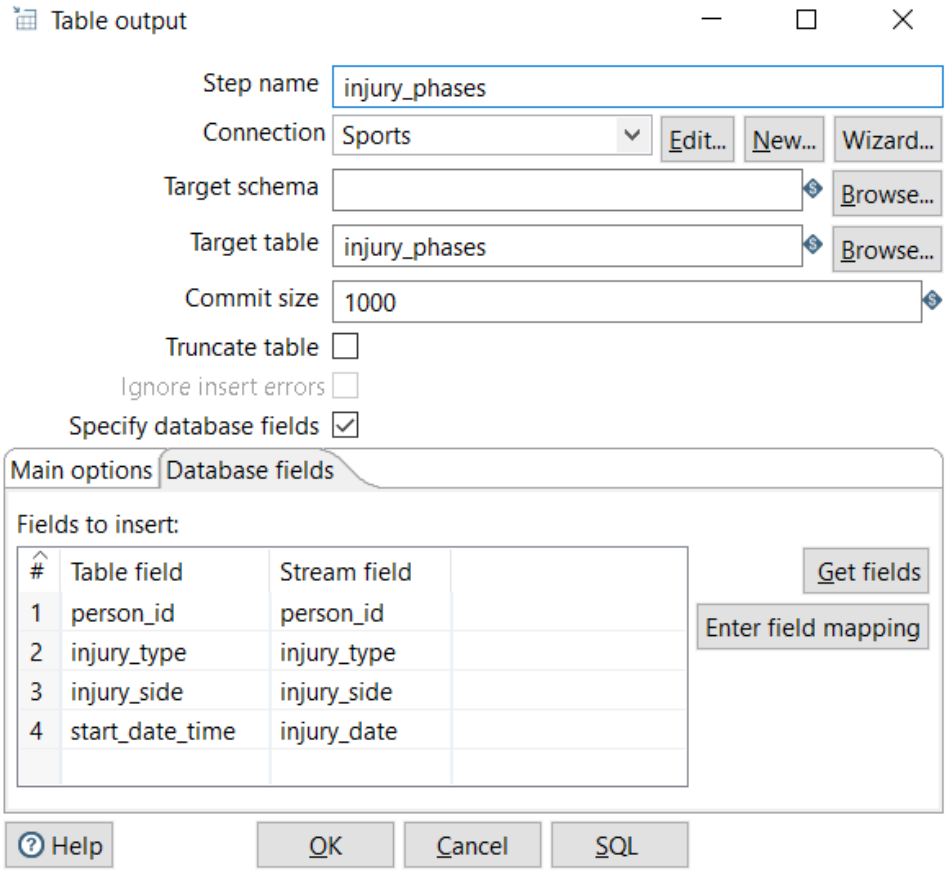
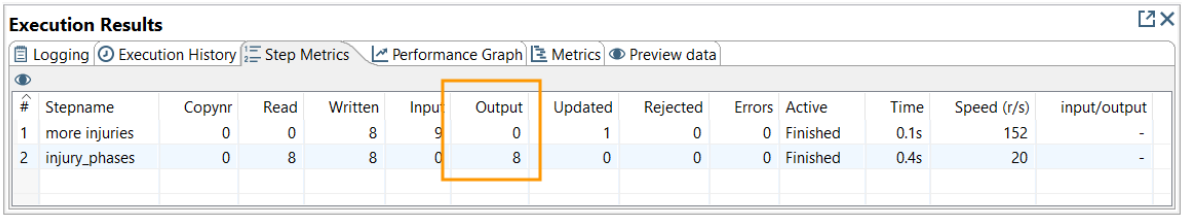


Table output
Step name
injury_phases
Connection
Sports
Edit...
New...
Wizard...
Target schema
Browse...
Target table
injury_phases
Browse...
Commit size
1000
Truncate table
Ignore insert errors
Specify database fields
Main options
Database fields
Fields to insert:
Table field Stream field
1 person_id person_id
2 injury_type injury_type
3 injury_side injury_side
4 start_date_time injury_date
Get fields
Enter field mapping
Help
OK
Cancel
SQL


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/output
1 more injuries 0 0 8 9 0 1 0 0 Finished 0.1s 152 -
2 injury_phases 0 8 8 0 8 0 0 0 Finished 0.4s 20 -

Update

Step name: Update injury details

Connection: Sports — Edit... New... Wizard...

Target schema: Browse...

Target table: injury_phases — Browse...

Commit size: 100

Use batch updates? ☐

Skip lookup ☐

Ignore lookup failure? ☐ Flag field (key found)

The key(s) to look up the value(s):

#	Table field	Comparator	Stream field1	Stream field2
1	person_id	=	person_id	
2	start_date_time	=	injury_date	

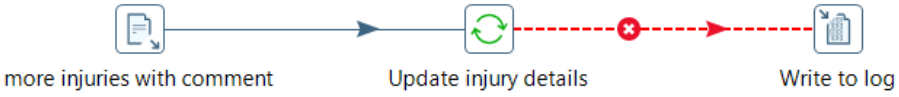
Get fields

Update fields:

#	Table field	Stream field
1	injury_comment	comment

Get update fields

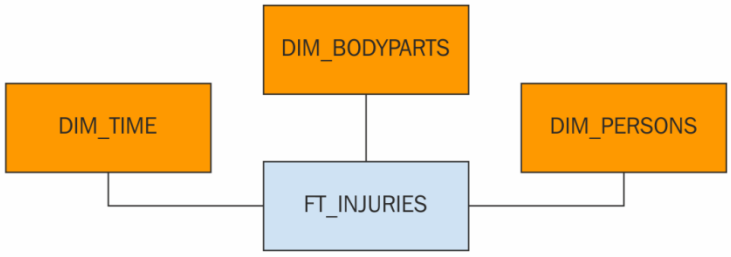
Help OK Cancel SQL

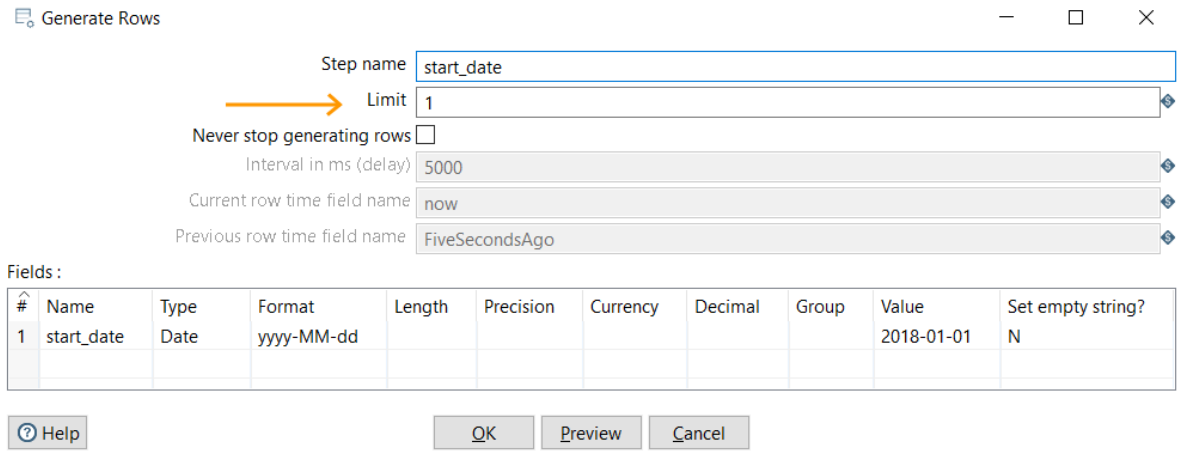


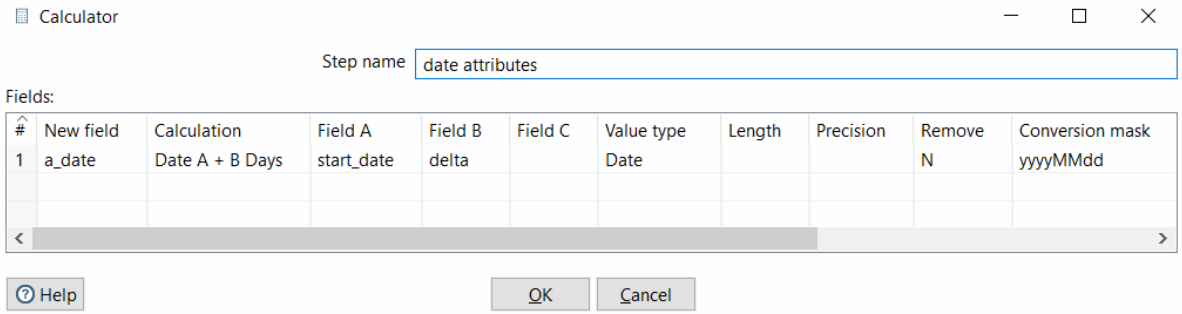
Execution Results

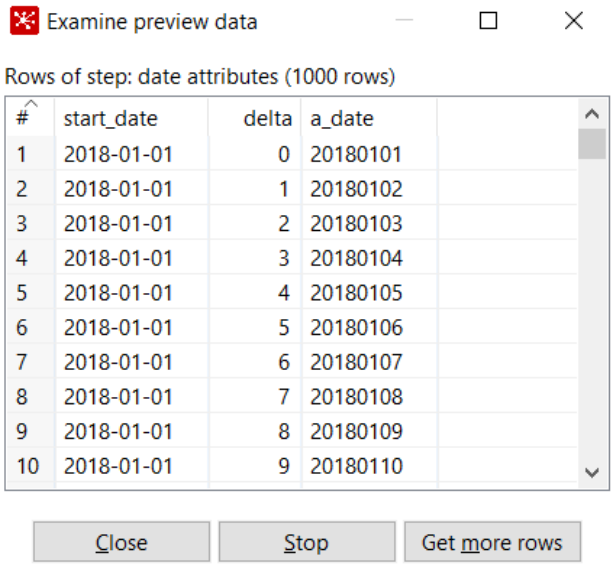
Logging | Execution History | Step Metrics | Performance Graph | Metrics | Preview data

#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	input/output
1	more injuries with comment	0	0	4	5	0	1	0	0	Finished	0.0s	454	-
2	Update injury details	0	4	3	4	0	3	1	0	Finished	0.3s	14	-
3	rejected lines	0	1	1	0	0	0	0	0	Finished	0.3s	4	-




Generate Rows
Step name
start_date
Limit
1
Never stop generating rows
Interval in ms (delay)
5000
Current row time field name
now
Previous row time field name
FiveSecondsAgo
Fields :
Name Type Format Length Precision Currency Decimal Group Value Set empty string?
1 start_date Date yyyy-MM-dd 2018-01-01 N
Help
OK
Preview
Cancel


Calculator
Step name
date attributes
Fields:
New field Calculation Field A Field B Field C Value type Length Precision Remove Conversion mask
1 a_date Date A + B Days start_date delta Date N yyyyMMdd
Help
OK
Cancel


Examine preview data
Rows of step: date attributes (1000 rows)
start_date delta a_date
1 2018-01-01 0 20180101
2 2018-01-01 1 20180102
3 2018-01-01 2 20180103
4 2018-01-01 3 20180104
5 2018-01-01 4 20180105
6 2018-01-01 5 20180106
7 2018-01-01 6 20180107
8 2018-01-01 7 20180108
9 2018-01-01 8 20180109
10 2018-01-01 9 20180110
Close
Stop
Get more rows

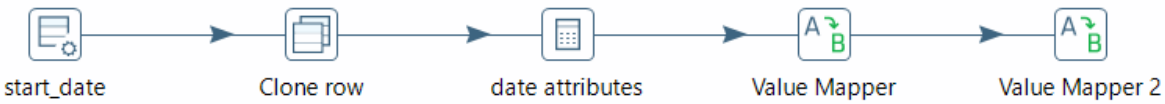
Calculator

Step name: date attributes

Fields:

#	New field	Calculation	Field A	Field B	Field C	Value type	Length	Precision	Remove	Conversion mask
1	a_date	Date A + B Days	start_date	delta		Date			N	yyyyMMdd
2	year	Year of date A	a_date			Integer			N	
3	month	Month of date A	a_date			Integer			N	
4	day	Day of month of date A	a_date			Integer			N	
5	week_day	Day of week of date A	a_date			Integer			N	

Help OK Cancel



Value Mapper

Step name : week_day_desc

Fieldname to use : week_day

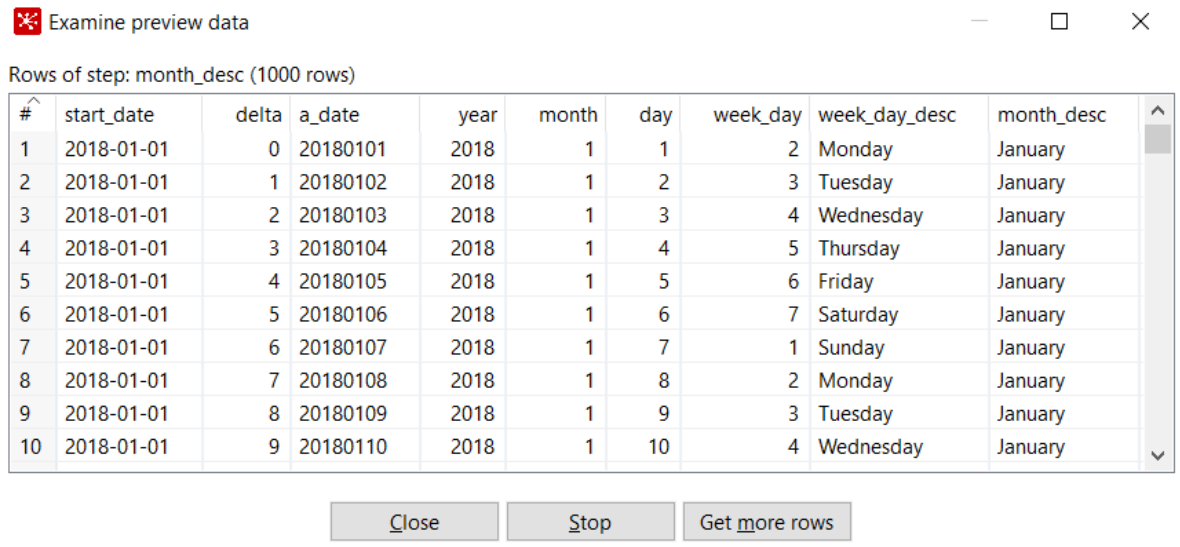
Target field name : week_day_desc

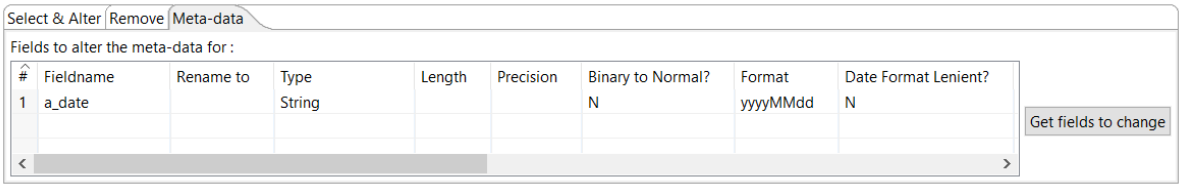
Default upon non-matching :

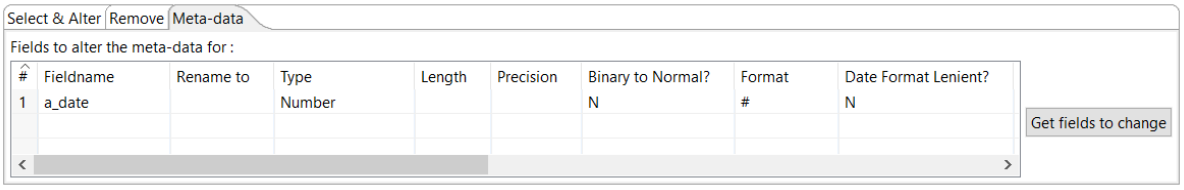
Field values:

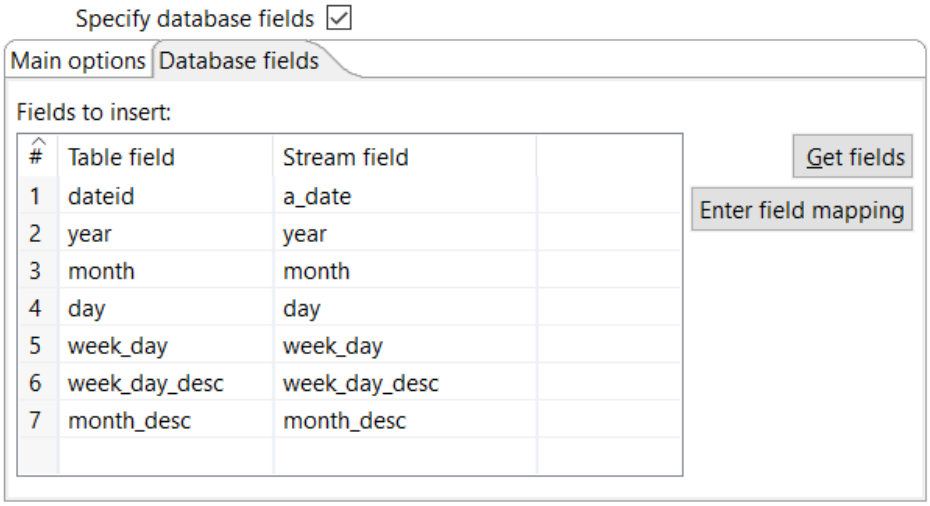
#	Source value	Target value
1	1	Sunday
2	2	Monday
3	3	Tuesday
4	4	Wednesday
5	5	Thursday
6	6	Friday
7	7	Saturday

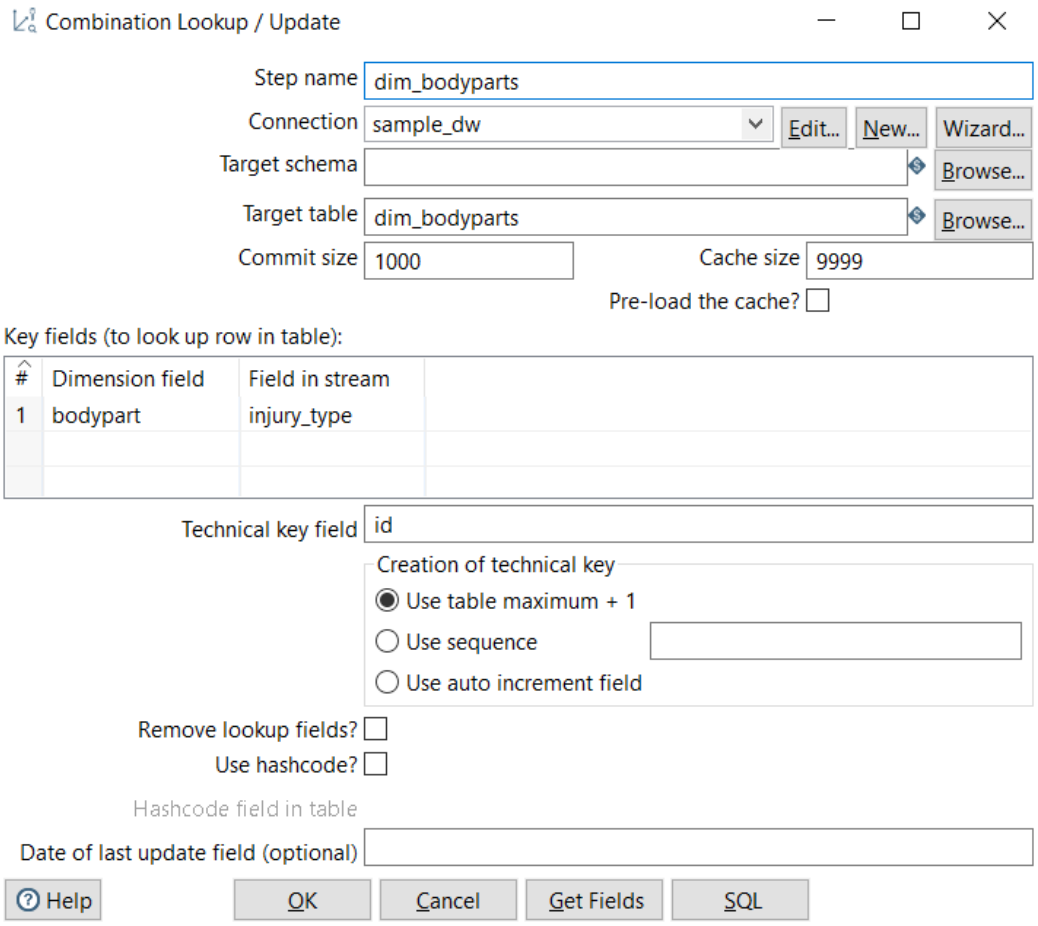
Help OK Cancel

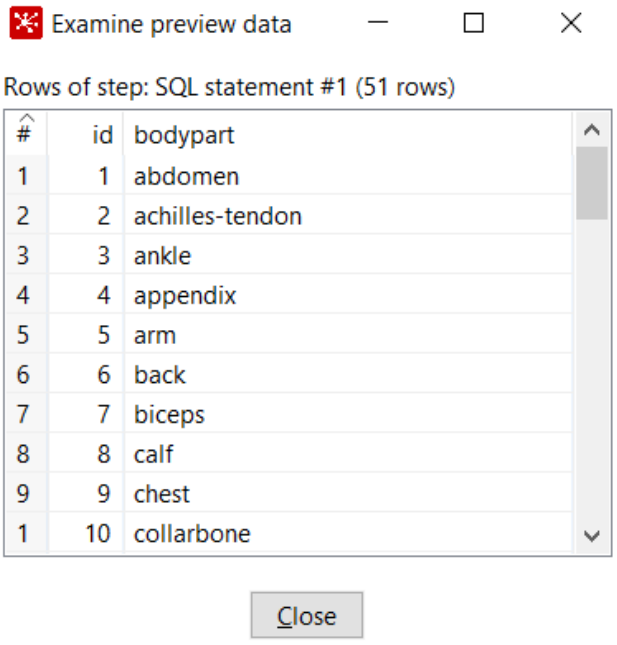

Examine preview data
Rows of step: month_desc (1000 rows)
start_date delta a_date year month day week_day week_day_desc month_desc
1 2018-01-01 0 20180101 2018 1 1 2 Monday January
2 2018-01-01 1 20180102 2018 1 2 3 Tuesday January
3 2018-01-01 2 20180103 2018 1 3 4 Wednesday January
4 2018-01-01 3 20180104 2018 1 4 5 Thursday January
5 2018-01-01 4 20180105 2018 1 5 6 Friday January
6 2018-01-01 5 20180106 2018 1 6 7 Saturday January
7 2018-01-01 6 20180107 2018 1 7 1 Sunday January
8 2018-01-01 7 20180108 2018 1 8 2 Monday January
9 2018-01-01 8 20180109 2018 1 9 3 Tuesday January
10 2018-01-01 9 20180110 2018 1 10 4 Wednesday January
Close
Stop
Get more rows


Select & Alter
Remove
Meta-data
Fields to alter the meta-data for :
Fieldname Rename to Type Length Precision Binary to Normal? Format Date Format Lenient?
1 a_date String N yyyyMMdd N
Get fields to change


Select & Alter
Remove
Meta-data
Fields to alter the meta-data for :
Fieldname Rename to Type Length Precision Binary to Normal? Format Date Format Lenient?
1 a_date Number N # N
Get fields to change


Specify database fields
Main options
Database fields
Fields to insert:
Table field Stream field
1 dateid a_date
2 year year
3 month month
4 day day
5 week_day week_day
6 week_day_desc week_day_desc
7 month_desc month_desc
Get fields
Enter field mapping


Combination Lookup / Update
Step name
dim_bodyparts
Connection
sample_dw
Edit...
New...
Wizard...
Target schema
Browse...
Target table
dim_bodyparts
Browse...
Commit size
1000
Cache size
9999
Pre-load the cache?
Key fields (to look up row in table):
Dimension field Field in stream
1 bodypart injury_type
Technical key field
id
Creation of technical key
Use table maximum + 1
Use sequence
Use auto increment field
Remove lookup fields?
Use hashcode?
Hashcode field in table
Date of last update field (optional)
Help
OK
Cancel
Get Fields
SQL


Examine preview data
Rows of step: SQL statement #1 (51 rows)
id bodypart
1 1 abdomen
2 2 achilles-tendon
3 3 ankle
4 4 appendix
5 5 arm
6 6 back
7 7 biceps
8 8 calf
9 9 chest
1 10 collarbone
Close

Keys | Fields

Key fields (to look up row in dimension):

#	Dimension field	Field in stream
1	entity_id	person_id

Keys | Fields

Lookup/Update fields

#	Dimension field	Stream field to compare with	Type of dimension update
1	first_name	first_name	Update
2	last_name	last_name	Update

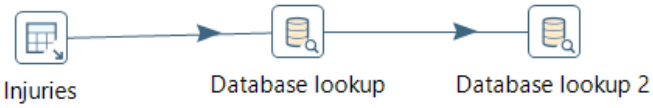
Examine preview data

Rows of step: dim_persons (100 rows)

#	id	first_name	last_name	entity_id	version	date_from	date_to
1	0	N/A	N/A	0	1	<null>	<null>
2	1	Jeff	Salazar	323	1	1900/01/01 00:00:0...	2199/12/31 23:59:59.999...
3	2	Robby	Hammock	326	1	1900/01/01 00:00:0...	2199/12/31 23:59:59.999...
4	3	Alberto	Callaspo	328	1	1900/01/01 00:00:0...	2199/12/31 23:59:59.999...
5	4	Brian	Barden	330	1	1900/01/01 00:00:0...	2199/12/31 23:59:59.999...
6	5	Donnie	Sadler	331	1	1900/01/01 00:00:0...	2199/12/31 23:59:59.999...
7	6	Randy	Choate	334	1	1900/01/01 00:00:0...	2199/12/31 23:59:59.999...
8	7	Mike	Schultz	335	1	1900/01/01 00:00:0...	2199/12/31 23:59:59.999...
9	8	Jailen	Peguero	336	1	1900/01/01 00:00:0...	2199/12/31 23:59:59.999...
1	9	Enrique	Gonzalez	333	1	1900/01/01 00:00:0...	2199/12/31 23:59:59.999...
1	10	Dana	Eveland	332	1	1900/01/01 00:00:0...	2199/12/31 23:59:59.999...

Close

Version field	version			
Stream Datefield				
Date range start field	date_from		Min. year	1900
Use an alternative start date?	<Select Option>			
Table date range end	date_to		Max. year	2199



The key(s) to look up the value(s):

#	Table field	Comparator	Field1	Field2
1	entity_id	=	person_id	

Values to return from the lookup table :

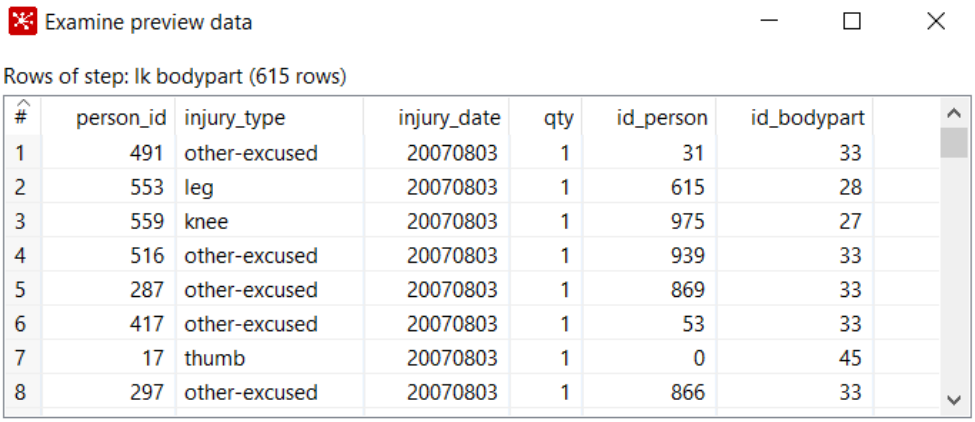
#	Field	New name	Default	Type
1	id	id_person	0	Integer

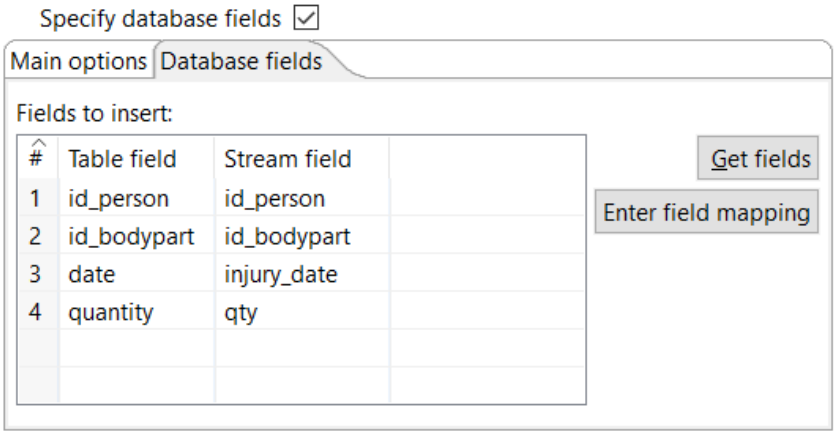
The key(s) to look up the value(s):

#	Table field	Comparator	Field1	Field2
1	bodypart	=	injury_type	

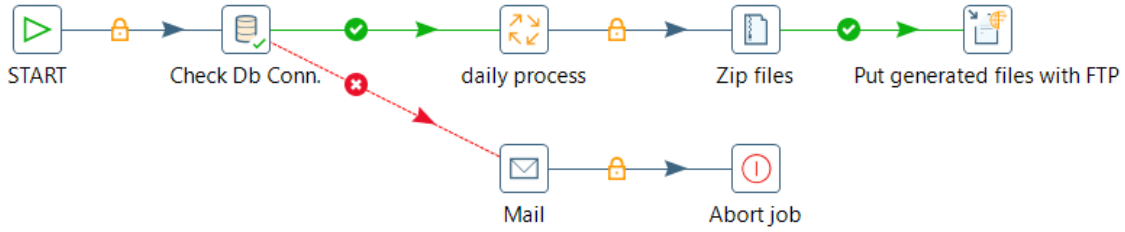
Values to return from the lookup table :

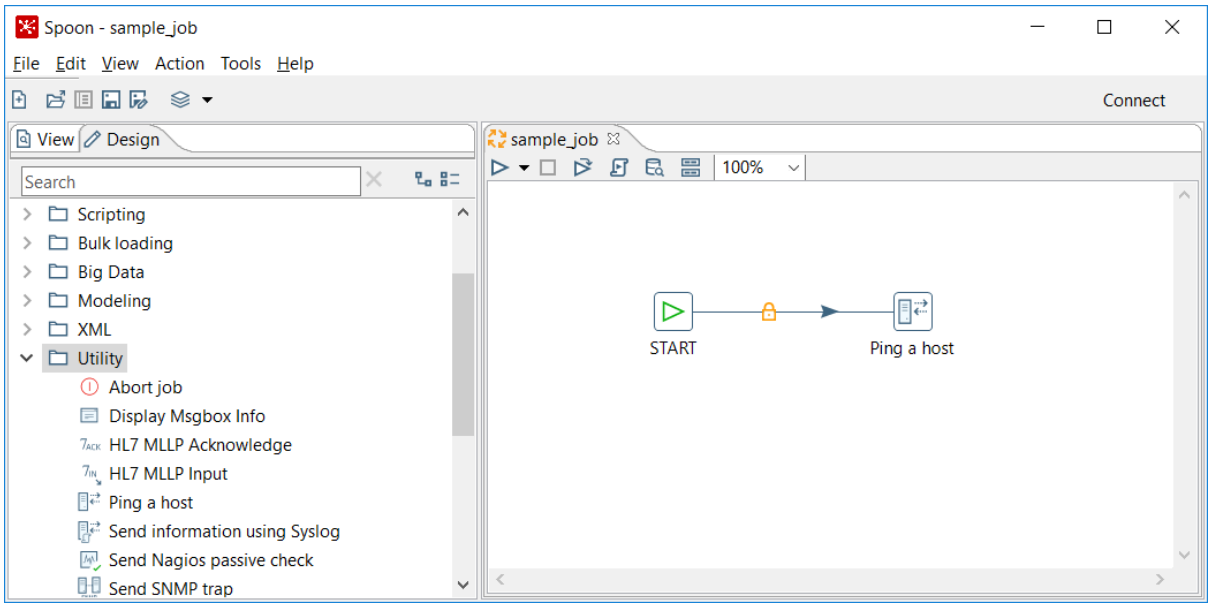
#	Field	New name	Default	Type
1	id	id_bodypart	0	Integer

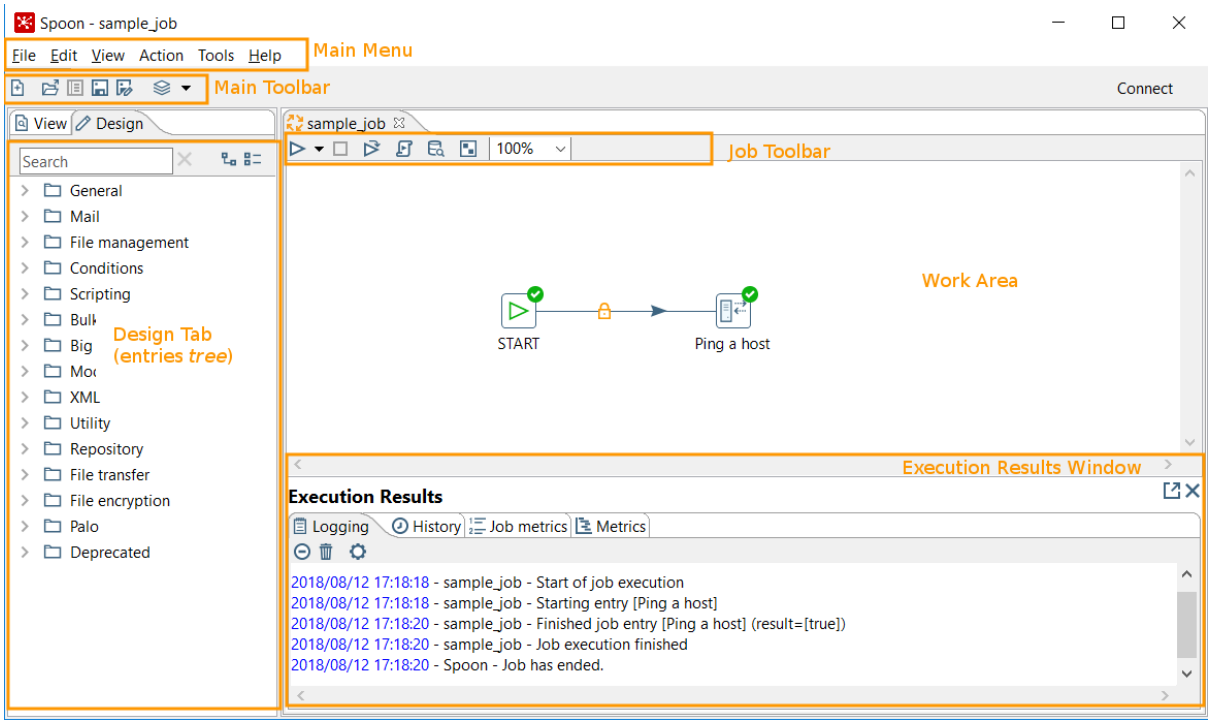

Examine preview data
Rows of step: lk bodypart (615 rows)
person_id injury_type injury_date qty id_person id_bodypart
1 491 other-excused 20070803 1 31 33
2 553 leg 20070803 1 615 28
3 559 knee 20070803 1 975 27
4 516 other-excused 20070803 1 939 33
5 287 other-excused 20070803 1 869 33
6 417 other-excused 20070803 1 53 33
7 17 thumb 20070803 1 0 45
8 297 other-excused 20070803 1 866 33
Close


Specify database fields
Main options
Database fields
Fields to insert:
Table field Stream field
1 id_person id_person
2 id_bodypart id_bodypart
3 date injury_date
4 quantity qty
Get fields
Enter field mapping

Chapter 6: Orchestrating Your Work



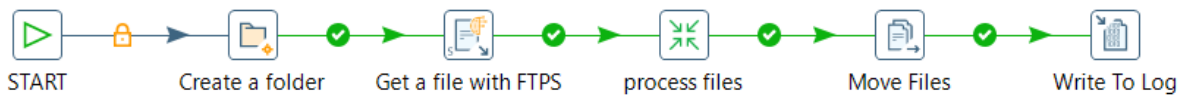


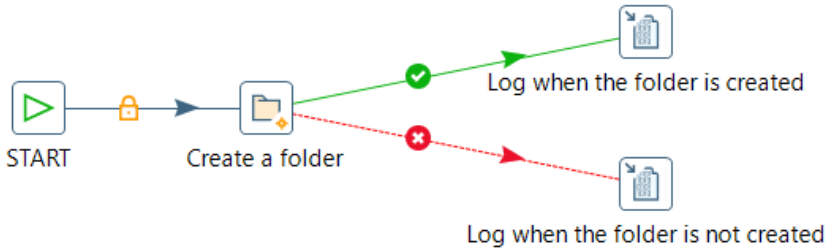


Execution Results

Logging | History | Job metrics | Metrics

Job / Job Entry	Comment	Result	Reason	Filename	Nr	Log date
sample_job						
Job: sample_job	Start of job execution		start			2018/08/12 17:04:44
START	Start of job execution		start			2018/08/12 17:04:44
START	Job execution finished	Success			0	2018/08/12 17:04:44
Ping a host	Start of job execution		(2) Followed unconditional link			2018/08/12 17:04:44 (1)
Ping a host	Job execution finished	(4) Failure			0	2018/08/12 17:04:44 (3)
Job: sample_job	Job execution finished	Failure	finished		0	2018/08/12 17:04:44





Run Options

Run configuration:

Pentaho local

Options

- [ ] Expand remote job
- [x] Clear log before running
- [ ] Enable safe mode
- [ ] Gather performance metrics

Log level: Basic

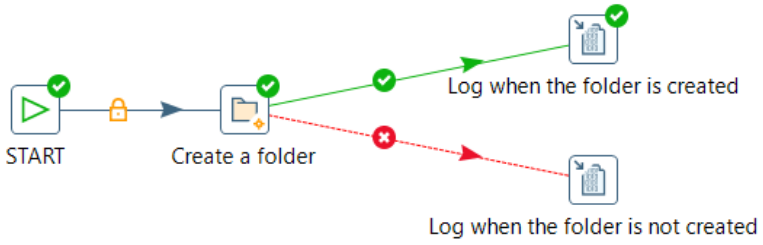
Start job at:

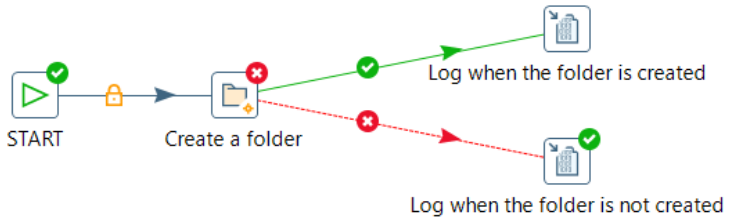
Parameters | Variables

Variable	Value
NEW_FOLDER	c:/pdi_logs

- [x] Always show dialog on run

Help | Run | Cancel





Execution Results

Logging | History | Job metrics | Metrics

Job / Job Entry	Comment	Result	Reason	Filename
sample_job_run_upon_conditions				
Job: sample_job_run_upon_conditions	Start of job execution		start	
START	Start of job execution		start	
START	Job execution finished	Success		
Create a folder	Start of job execution		Followed unconditional link	
Create a folder	Job execution finished	Failure		
Log when the folder is not created	Start of job execution		Followed link after failure	
Log when the folder is not created	Job execution finished	Success		
Job: sample_job_run_upon_conditions	Job execution finished	Success	finished	

Job mail details

Name of mail job entry: Sample email

Addresses | Server | EMail Message | Attached Files

SMTP Server

SMTP Server: smtp.gmail.com

Port: 465

Authentication

Use authentication? ☑

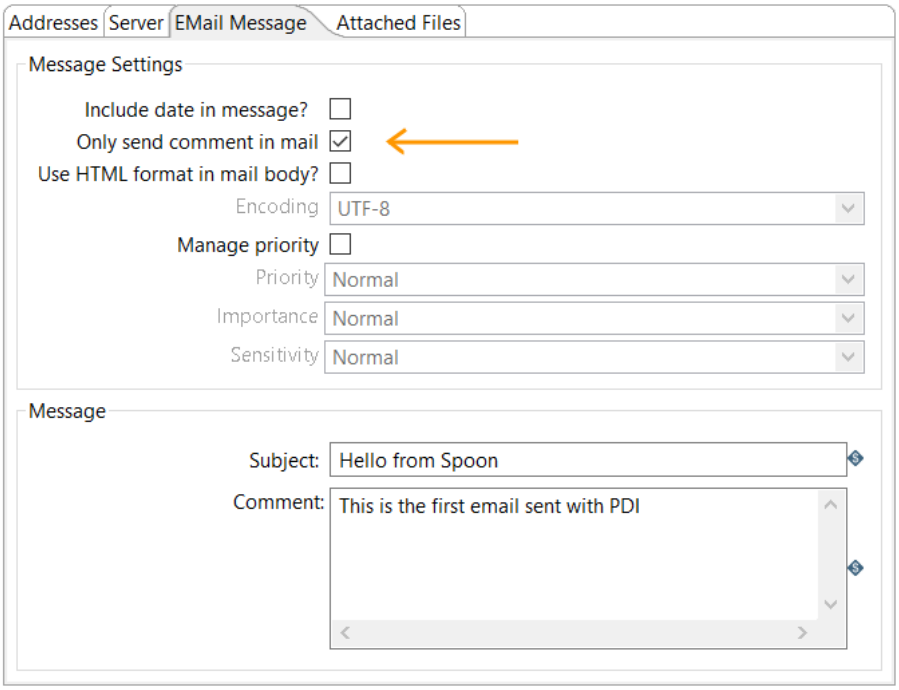
Authentication user: ....@gmail.com

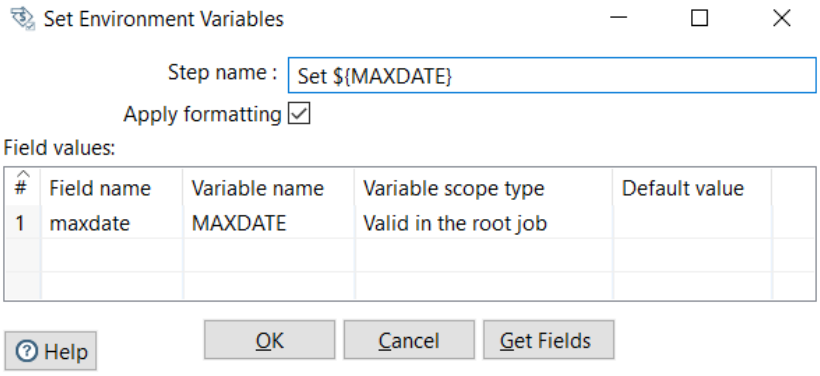
Authentication password: ●●●●●●●

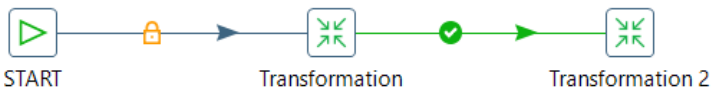
Use secure authentication? ☑

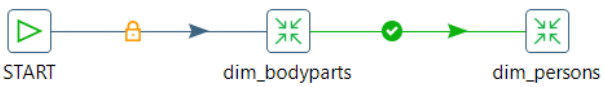
Secure connection type: SSL

Help | OK | Cancel

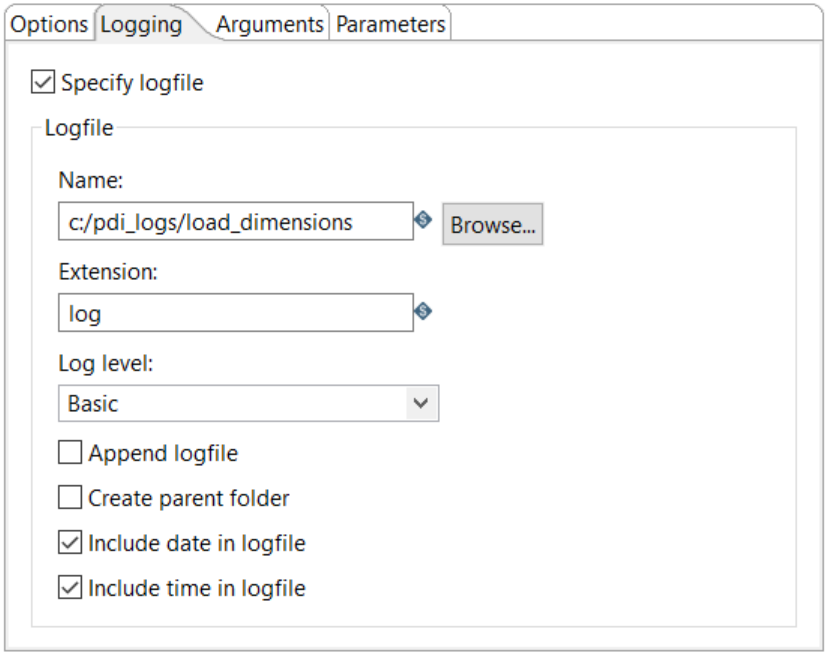

Addresses
Server
EMail Message
Attached Files
Message Settings
Include date in message?
Only send comment in mail
Use HTML format in mail body?
Encoding
UTF-8
Manage priority
Priority
Normal
Importance
Normal
Sensitivity
Normal
Message
Subject:
Hello from Spoon
Comment:
This is the first email sent with PDI

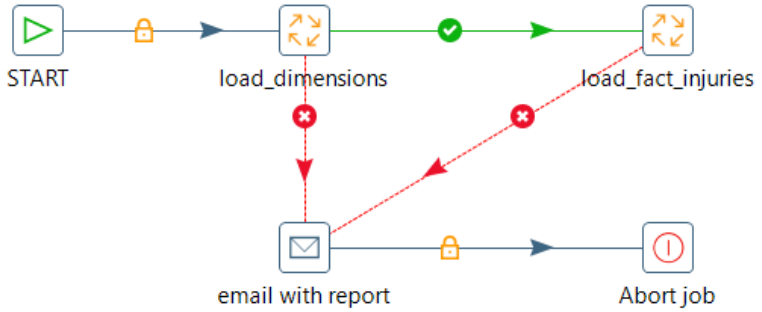

Set Environment Variables
Step name :
Set ${MAXDATE}
Apply formatting
Field values:
Field name Variable name Variable scope type Default value
1 maxdate MAXDATE Valid in the root job
Help
OK
Cancel
Get Fields


START
Transformation
Transformation 2


START
dim_bodyparts
dim_persons

START
load_dimensions
load_fact_injuries


Options
Logging
Arguments
Parameters
Specify logfile
Logfile
Name:
c:/pdi_logs/load_dimensions
Browse...
Extension:
log
Log level:
Basic
Append logfile
Create parent folder
Include date in logfile
Include time in logfile


START
load_dimensions
load_fact_injuries
email with report
Abort job