

# Exercise 1 – Group 50

José Cardoso

Matheus Trindade

# Transformation

Activities SWT

Sun 22:12

Spoon - Project1

File Edit View Action Tools Help

View Design output

Input

- HL7 input
- JSON input
- YAML input

Output

- Automatic documentation output
- JSON output
- LDAP output
- Microsoft Access output
- Microsoft Excel output
- Pentaho reporting output
- Properties output
- RSS output
- S3 file output
- SQL file output
- Table output
- Text file output**
- XML output

Streaming

Transform

- Concat fields
- Row denormaliser

Utility

- Delay row

Flow

- Blocking step
- Detect empty stream
- ETL metadata injection

Scripting

Transformation 1 Transformation 2 lab05 ordertotals dim\_customer dim\_product dim\_time fact\_order load\_dw Project1

100%

Dataset\_A

Select values

Filter rows

Sort rows

Group by 2

Join rows

Select values 3

Text file output

Dataset\_B

Select values 2

Filter rows 2

Sort rows 2

Group by

Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Preview data

2023/10/15 22:10:14 - Select values 2.0 - linenr 150000  
2023/10/15 22:10:14 - Filter rows 2.0 - linenr 150000  
2023/10/15 22:10:14 - Dataset\_B.0 - Finished processing (I=151617, O=0, R=0, W=151616, U=0, E=0)  
2023/10/15 22:10:14 - Select values 2.0 - Finished processing (I=0, O=0, R=151616, W=151616, U=0, E=0)  
2023/10/15 22:10:14 - Filter rows 2.0 - Finished processing (I=0, O=0, R=151616, W=4469, U=0, E=0)  
2023/10/15 22:10:14 - Sort rows 2.0 - Finished processing (I=0, O=0, R=4469, W=4469, U=0, E=0)  
2023/10/15 22:10:14 - Group by.0 - Finished processing (I=0, O=0, R=4469, W=278, U=0, E=0)  
2023/10/15 22:10:14 - Join rows .0 - Finished processing (I=0, O=0, R=556, W=278, U=0, E=0)  
2023/10/15 22:10:14 - Select values 3.0 - Finished processing (I=0, O=0, R=278, W=278, U=0, E=0)  
2023/10/15 22:10:14 - Text file output.0 - Finished processing (I=0, O=279, R=278, W=278, U=0, E=0)  
2023/10/15 22:10:14 - Spoon - The transformation has finished!!

# Configuration of CSV File Input – Dataset A

Spoon - Project1 Tue 21:45

Activities SWT

File Edit View Action Tools Help

View Design database loo

Input Output Streaming Transform Utility Flow Scripting Pentaho Server Lookup Joins Data Warehouse Validation Statistics Big Data Cryptography Job Mapping Bulk loading Inline Experimental Deprecated History

fact\_order load\_dw Project1 Exercise3 dim\_location /home/aid/Documents/ dim\_voltage fact\_consumption Job\_Exercise6

CSV file input

Step name: Dataset\_A  
Filename: \${Internal.Entry.Current.Directory}/21-contadores-de-energia.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: UTF-8

Name	Type	Format	Length	Precision	Currency	Decimal
1 Year	Integer	#	15	0	\$	.
2 Month	Integer	#	15	0	\$	.
3 Date	String		7		\$	.
4 District	String		16		\$	.
5 Municipality	String		27		\$	.
6 Parish	String		89		\$	.
7 incl. Smart Meter	String		3		\$	.
8 Number of CPE's	Integer	#	15	0	\$	.
9 DistrictCode	String	#	15	0	\$	.
10 DistrictMunicipalityCode	String	#	15	0	\$	.
11 DistrictMunicipalityParishCode	String	#	15	0	\$	.
12 Contrato Ativo	String		3		\$	.

Help OK Get Fields Preview Cancel

```
graph LR; Dataset_A[Dataset_A] --> SelectA[Select values]; SelectA --> FilterA[Filter rows]; FilterA --> SortA[Sort rows]; Dataset_B[Dataset_B] --> SelectB[Select values 2]; SelectB --> FilterB[Filter rows 2]; FilterB --> SortB[Sort rows 2]
```

# Preview of CSV File Input – Dataset A

Activities SWT Tue 21:48

Examine preview data

	Year	Month	Date	District	Municipality	Parish	incl. Smart Meter	Number of CPE's	DistrictCode	DistrictMunicipalityCode	DistrictMunicipalityParishCode	Contrato Ativo
1	2022	1	2022-01	Aveiro	Arouca	Santa Eulália	SIM	265	01	0104	010416	SIM
2	2022	1	2022-01	Aveiro	Espinho	União das freguesias de Anta e Guetim	NÃO	1837	01	0107	010706	SIM
3	2022	1	2022-01	Aveiro	Oliveira do Bairro	Oliveira do Bairro	NÃO	255	01	0114	011404	SIM
4	2022	1	2022-01	Aveiro	Ovar	União das freguesias de Ovar, São João,	SIM	12886	01	0115	011509	SIM
5	2022	1	2022-01	Beja	Mértola	Santana de Cambas	NÃO	801	02	0209	020905	SIM
6	2022	1	2022-01	Beja	Moura	União das freguesias de Safara e Santo António	SIM	363	02	0210	021010	SIM
7	2022	1	2022-01	Beja	Ourique	União das freguesias de Garvão e Santa Maria	SIM	156	02	0212	021207	SIM
8	2022	1	2022-01	Braga	Amares	Barreiros	SIM	110	03	0301	030102	SIM
9	2022	1	2022-01	Braga	Amares	Fiscal	SIM	141	03	0301	030111	SIM
10	2022	1	2022-01	Braga	Amares	União das freguesias de Vilela, Seramil e Freixo	SIM	71	03	0301	030129	SIM
11	2022	1	2022-01	Braga	Braga	Priscos	NÃO	365	03	0303	030336	SIM
12	2022	1	2022-01	Braga	Braga	União das freguesias de Nogueiró e Tenedo	SIM	1191	03	0303	030378	SIM
13	2022	1	2022-01	Braga	Fafe	São Gens	SIM	219	03	0307	030728	SIM
14	2022	1	2022-01	Braga	Guimarães	Aldão	NÃO	19	03	0308	030801	SIM
15	2022	1	2022-01	Braga	Guimarães	São Torcato	SIM	1162	03	0308	030865	SIM
16	2022	1	2022-01	Braga	Vieira do Minho	Salamonde	SIM	96	03	0311	031115	SIM
17	2022	1	2022-01	Braga	Vila Nova de Famalicão	Bairro	SIM	648	03	0312	031204	SIM
18	2022	1	2022-01	Braga	Vila Nova de Famalicão	Brufe	NÃO	55	03	0312	031206	SIM
19	2022	1	2022-01	Braga	Vila Nova de Famalicão	Castelões	SIM	311	03	0312	031210	SIM
20	2022	1	2022-01	Braga	Vila Nova de Famalicão	Joane	NÃO	2447	03	0312	031219	SIM
21	2022	1	2022-01	Braga	Vila Verde	Dossãos	SIM	66	03	0313	031313	SIM
22	2022	1	2022-01	Bragança	Bragança	Gimonde	NÃO	170	04	0402	040216	SIM
23	2022	1	2022-01	Bragança	Bragança	União das freguesias de Castrelos e Carrazeda de Ansiães	SIM	147	04	0402	040251	SIM
24	2022	1	2022-01	Bragança	Carrazeda de Ansiães	Pinhal do Norte	SIM	223	04	0403	040313	SIM
25	2022	1	2022-01	Bragança	Macedo de Cavaleiros	Corujas	NÃO	120	04	0405	040511	SIM
26	2022	1	2022-01	Bragança	Macedo de Cavaleiros	Lamalonga	SIM	50	04	0405	040517	SIM
27	2022	1	2022-01	Bragança	Macedo de Cavaleiros	União das freguesias de Espadanedo, Ec	SIM	110	04	0405	040542	SIM
28	2022	1	2022-01	Bragança	Mirandela	Cobro	NÃO	70	04	0407	040713	SIM
29	2022	1	2022-01	Bragança	Mirandela	Múrias	NÃO	76	04	0407	040722	SIM
30	2022	1	2022-01	Bragança	Vila Flor	Freixiel	NÃO	424	04	0410	041005	SIM
31	2022	1	2022-01	Bragança	Vila Flor	Santa Comba de Vilariça	NÃO	285	04	0410	041012	SIM
32	2022	1	2022-01	Castelo Branco	Castelo Branco	Lardosa	SIM	299	05	0502	050211	SIM
33	2022	1	2022-01	Castelo Branco	Fundão	Alcaria	SIM	638	05	0504	050402	SIM

Close

Stop

Get more rows

# Configuration of Select Values – Dataset A

Spoon - Project1 Tue 21:46

Activities SWT

File Edit View Action Tools Help

View Design database loo

Input Output Streaming Transform Utility Flow Scripting Pentaho Server Lookup Joins Data Warehouse Validation Statistics Big Data Cryptography Job Mapping Bulk loading Inline Experimental Deprecated History

fact\_order load\_dw Project1 Exercise3 dim\_location /home/aid/Documents/ dim\_voltage fact\_consumption Job\_Exercise6

Connect

Dataset\_A → Select values → Filter rows → Sort rows → Group by

Dataset\_B → Select values 2 → Filter rows 2 → Sort rows 2 → Group by

Select values Step name: Select values

Select & Alter Remove Meta-data

Fields:

Fieldname	Rename to	Length	Precision
1 Year	Year_A		
2 Month	Month_A		
3 Municipality	Municipality_A		
4 Number of CPE's	Number of CPE's_A		
5 Contrato Ativo	Contrato Ativo_A		

Get fields to select Edit Mapping

Include unspecified fields, ordered |

OK Cancel Help

```
graph LR; subgraph DA [Dataset_A]; A1[Input] --> A2[Select values]; A2 --> A3[Filter rows]; A3 --> A4[Sort rows]; A4 --> A5[Group by]; end; subgraph DB [Dataset_B]; B1[Input] --> B2[Select values 2]; B2 --> B3[Filter rows 2]; B3 --> B4[Sort rows 2]; B4 --> B5[Group by]; end;
```

```
graph TD; subgraph DV [Select values]; DV1[Step name: Select values]; DV2[Select & Alter Remove Meta-data]; DV3[Fields:]; DV4[Get fields to select]; DV5[Edit Mapping]; DV6[Include unspecified fields, ordered | ]; DV7[OK]; DV8[Cancel]; DV9[Help]; end;
```

Fieldname	Rename to	Length	Precision
1 Year	Year_A		
2 Month	Month_A		
3 Municipality	Municipality_A		
4 Number of CPE's	Number of CPE's_A		
5 Contrato Ativo	Contrato Ativo_A		

# Preview of Select Values – Dataset A

Sun 22:14  
Spoon - Project1

File Edit View Action Tools Help

Transformation 1 Transformation 2 lab05 ordertotals dim\_customer dim\_product dim\_time fact\_order load\_dw Project1

View Design output

Examine preview data

Rows of step: Select values (100 rows)

	Year_A	Month_A	Municipality_A	Number of CPE's_A	Contrato Ativo_A
1	2022	1	Arouca	265	SIM
2	2022	1	Espinho	1837	SIM
3	2022	1	Oliveira do Bairro	255	SIM
4	2022	1	Ovar	12886	SIM
5	2022	1	Mértola	801	SIM
6	2022	1	Moura	363	SIM
7	2022	1	Ourique	156	SIM
8	2022	1	Amares	110	SIM
9	2022	1	Amares	141	SIM
10	2022	1	Amares	71	SIM
11	2022	1	Braga	365	SIM
12	2022	1	Braga	1191	SIM
13	2022	1	Fafe	219	SIM
14	2022	1	Guimarães	19	SIM
15	2022	1	Guimarães	1162	SIM
16	2022	1	Vieira do Minho	96	SIM
17	2022	1	Vila Nova de Famalicão	648	SIM
18	2022	1	Vila Nova de Famalicão	55	SIM
19	2022	1	Vila Nova de Famalicão	311	SIM
20	2022	1	Vila Nova de Famalicão	2447	SIM
21	2022	1	Vila Verde	66	SIM
22	2022	1	Bragança	170	SIM
23	2022	1	Bragança	147	SIM
24	2022	1	Carrazeda de Ansiães	223	SIM

Close Stop Get more rows

Activities SWT

ETL metadata injection Scripting

# Configuration of Filter Rows – Dataset A

Sun 22:56  
Spoon - Project1

Activities    SWT    Connect

File Edit View Action Tools Help

View Design

output

Transformation 1 Transformation 2 lab05 ordertotals dim\_customer dim\_product dim\_time fact\_order load\_dw Project1

100%

Dataset\_A → Select values → Filter rows → Sort rows → Group by 2 → Join rows → Select values 3 → Text file output

Filter rows

Step name: Filter rows

Send 'true' data to step: Sort rows

Send 'false' data to step:

The condition:

Year\_A = [2023]

AND

Month\_A = [6]

OK Cancel

Metrics Preview data

it	Updated	Rejected	Errors	Active	Time	Speed (r/s)	input/output
0	0	0	0	Finished	1.3s	118,451	
0	0	0	0	Finished	1.3s	117,623	
0	0	0	0	Finished	1.1s	114,774	
0	0	0	0	Finished	1.1s	110,278	
0	0	0	0	Finished	1.2s	109,139	
0	0	0	0	Finished	1.3s	116,897	
0	0	0	0	Finished	1.3s	3,404	
0	0	0	0	Finished	1.3s	3,352	

5 Filter rows    0    126710    5759    0    0  
6 Filter rows 2    0    151616    4469    0    0  
7 Sort rows 2    0    4469    4469    0    0  
8 Group by    0    4469    278    0    0

# Preview of Filter Rows – Dataset A

Sun 22:15  
Spoon - Project1

File Edit View Action Tools Help

Transformation 1 Transformation 2 lab05 ordertotals dim\_customer dim\_product dim\_time fact\_order load\_dw Project1

View Design output

Input Output Text XML Stream Transformations Copy Rows Utilities Flows

Examine preview data

Rows of step: Filter rows (100 rows)

	Year_A	Month_A	Municipality_A	Number of CPE's_A	Contrato Ativo_A
1	2023	6	Castelo de Paiva	67	SIM
2	2023	6	Santa Maria da Feira	458	SIM
3	2023	6	Mealhada	445	SIM
4	2023	6	Vale de Cambra	1332	SIM
5	2023	6	Beja	444	SIM
6	2023	6	Beja	759	SIM
7	2023	6	Beja	133	SIM
8	2023	6	Castro Verde	3018	SIM
9	2023	6	Amares	120	SIM
10	2023	6	Amares	89	SIM
11	2023	6	Barcelos	327	SIM
12	2023	6	Barcelos	668	SIM
13	2023	6	Braga	311	SIM
14	2023	6	Braga	225	SIM
15	2023	6	Guimarães	628	SIM
16	2023	6	Póvoa de Lanhoso	128	SIM
17	2023	6	Terras de Bouro	17	SIM
18	2023	6	Terras de Bouro	915	SIM
19	2023	6	Terras de Bouro	154	SIM
20	2023	6	Vieira do Minho	185	SIM
21	2023	6	Vila Nova de Famalicão	228	SIM
22	2023	6	Vila Verde	210	SIM
23	2023	6	Vizela	105	SIM
24	2023	6	Alfândega da Fé	160	SIM

Close Stop Get more rows

ETL metadata injection Scripting

# Configuration of Sort Rows – Dataset A

Sun 22:56  
Spoon - Project1

Activities    SWT    Connect

File Edit View Action Tools Help

View Design Transformation 1 Transformation 2 lab05 ordertotals dim\_customer dim\_product dim\_time fact\_order load\_dw Project1

100%

output

Input

- HL7 input
- JSON input
- YAML input

Output

- Automatic document
- JSON output
- LDAP output
- Microsoft Access output
- Microsoft Excel output
- Pentaho reporting
- Properties output
- RSS output
- S3 file output
- SQL file output
- Table output
- Text file output**
- XML output

Transform

- Concat fields
- Row denormaliser

Utility

- Delay row

Flow

- Blocking step
- Detect empty stream
- ETL metadata injection
- Scripting

Sort rows

Step name: Sort rows

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

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	Presort
Municipality_A	Y	N	N	0	N

OK Cancel Get Fields

Transformation 1 Transformation 2 lab05 ordertotals dim\_customer dim\_product dim\_time fact\_order load\_dw Project1

Active Finished

100%

Errors

Select values 3 Text file output

w data

Rejected	Errors	Active	Time	Speed (r/s)	input/output
0	0	Finished	1.3s	118,451	
0	0	Finished	1.3s	117,623	
0	0	Finished	1.1s	114,774	
0	0	Finished	1.1s	110,278	
0	0	Finished	1.2s	109,139	
0	0	Finished	1.3s	116,897	
0	0	Finished	1.3s	3,404	
0	0	Finished	1.3s	3,352	

# Preview of Sort Rows– Dataset A

Sun 22:15

Spoon - Project1

Activities SWT

File Edit View Action Tools Help

View Design Transformation 1 Transformation 2 lab05 ordertotals dim\_customer dim\_product dim\_time fact\_order load\_dw Project1

100%

Examine preview data

Rows of step: Sort rows (100 rows)

	Year_A	Month_A	Municipality_A	Number of CPE's_A	Contrato Ativo_A
1	2023	6	Abrantes	167	SIM
2	2023	6	Abrantes	714	SIM
3	2023	6	Abrantes	725	SIM
4	2023	6	Abrantes	243	SIM
5	2023	6	Abrantes	307	SIM
6	2023	6	Abrantes	751	SIM
7	2023	6	Abrantes	504	SIM
8	2023	6	Abrantes	263	SIM
9	2023	6	Abrantes	1437	SIM
10	2023	6	Abrantes	380	SIM
11	2023	6	Abrantes	310	SIM
12	2023	6	Abrantes	9840	SIM
13	2023	6	Abrantes	220	SIM
14	2023	6	Abrantes	260	SIM
15	2023	6	Abrantes	1391	SIM
16	2023	6	Abrantes	903	SIM
17	2023	6	Abrantes	660	SIM
18	2023	6	Abrantes	743	SIM
19	2023	6	Abrantes	348	SIM
20	2023	6	Abrantes	471	SIM
21	2023	6	Abrantes	934	SIM
22	2023	6	Abrantes	517	SIM
23	2023	6	Abrantes	378	SIM
24	2023	6	Abrantes	1598	SIM

Close Stop Get more rows

2023/10/15 22:15:23 - SORTROWS 2.0 - Finished processing (i=0, O=0, R=4402, W=4402, U=0, E=0)  
2023/10/15 22:15:23 - Group by=0 - Finished processing (i=0, O=0, R=4469, W=278, U=0, E=0)

ETL metadata injection Scripting

# Configuration of Group by 2– Dataset A

Sun 22:56  
Spoon - Project1

Activities SWT

File Edit View Action Tools Help

View Design Transformation

output

Input

- HL7 input
- JSON input
- YAML input

Output

- Automatic documentation output
- JSON output
- LDAP output
- Microsoft Access output
- Microsoft Excel output
- Pentaho reporting output
- Properties output
- RSS output
- S3 file output
- SQL File output
- Table output
- Text file output
- XML output

Streaming

Transform

- Concat fields
- Row denormaliser

Utility

- Delay row

Flow

- Blocking step
- Detect empty stream
- ETL metadata injection

Scripting

Group by

Step name: Group by 2

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: Municipality\_A

Get Fields

Aggregates:

Name	Subject	Type
1 Total CPE's	Number of CPE's_A	Sum

Get lookup fields

Execution Results

Logging

Stepname

Stepname	Time	Speed (r/s)	input/output
1 Dataset_B	1.3s	118,451	
2 Select values	1.3s	117,623	
3 Dataset_A	1.1s	114,774	
4 Select values	1.1s	110,278	
5 Filter rows	1.2s	109,139	
6 Filter rows 2	1.3s	116,897	
7 Sort rows 2	1.3s	3,404	
8 Group by	1.3s	3,352	

OK Cancel

# Preview of Group by 2– Dataset A

Activities SWT ▾ Sun 22:15 Spoon - Project1

File Edit View Action Tools Help

View Design Transformation 1 Transformation 2 lab05 ordertotals dim\_customer dim\_product dim\_time Fact\_order load\_dw Project1

100% Errors

Examine preview data

Rows of step: Group by 2 (100 rows)

	Municipality_A	Total CPE's
1	Abrantes	25180
2	Aguiar da Beira	4933
3	Alandroal	4133
4	Albergaria-a-Velha	14241
5	Albufeira	54157
6	Alcanena	8265
7	Alcobaça	39245
8	Alcochete	11819
9	Alcoutim	3703
10	Alcácer do Sal	8480
11	Alenquer	24858
12	Alfândega da Fé	4167
13	Alijó	9146
14	Aljezur	6523
15	Aljustrel	6058
16	Almada	116447
17	Almeida	7014
18	Almeirim	14117
19	Almodôvar	5693
20	Alpiarça	4630
21	Alter do Chão	2892
22	Alvaiázere	5897
23	Alvito	1725
24	Amadora	98733

Close Stop Get more rows

2023/10/15 22:15:47 - SELECT values.0 - Finished processing (I=0, O=0, R=126710, W=126710, U=0, E=0)  
2023/10/15 22:15:47 - Filter rows.0 - Finished processing (I=0, O=0, R=126710, W=5759, U=0, E=0)

ETL metadata injection Scripting

# Configuration of CSV File Input– Dataset B

Spoon - Project1

Tue 21:46

Activities SWT

File Edit View Action Tools Help

View Design

database loo

Input Output Streaming Transform Utility Flow Scripting Pentaho Server Lookup Joins Data Warehouse Validation Statistics Big Data Cryptography Job Mapping Bulk loading Inline Experimental Deprecated History

fact\_order load\_dw Project1 Exercise3 dim\_location /home/aid/Documents/ dim\_voltage fact\_consumption Job\_Exercise6

Connect

Dataset\_A → Select values → Filter rows → Sort rows → Group by

Dataset\_B → Select values 2 → Filter rows 2 → Sort rows 2 → Group by

Select values

Step name Select values

Select & Alter Remove Meta-data

Fields:

Fieldname	Rename to	Length	Precision
1 Year	Year_A		
2 Month	Month_A		
3 Municipality	Municipality_A		
4 Number of CPE's	Number of CPE's_A		
5 Contrato Ativo	Contrato Ativo_A		

Get fields to select Edit Mapping

Include unspecified fields, ordered |

OK Cancel Help

# Preview of CSV File Input– Dataset B

Activities Tue 21:49

Examine preview data

Rows of step: Dataset\_B (100 rows)

	Year	Month	Date	District	Municipality	parish	Voltage level	Active Energy (kWh)	DistrictCode	DistrictMunicipalityCode	DistrictMunicipalityParishCode
1	2020	11	2020-11	GUARDA	Guarda	ARRIFANA	Muito Alta, Alta e Média Tensões	57149.7	09	0907	090706
2	2020	11	2020-11	PORTO	Vila do Conde	UF MALTA E CANIDELO	Baixa Tensão	303181.9	13	1316	131633
3	2020	11	2020-11	SETUBAL	Alcácer do Sal	UF ALCACER SAL E SANTA SUSANA	Muito Alta, Alta e Média Tensões	2994085.1	15	1501	150107
4	2020	12	2020-12	AVEIRO	Anadia	UF ARCOS E MOGOFORES	Baixa Tensão	1599645.3	01	0103	010317
5	2020	12	2020-12	BRAGA	Barcelos	FRAGOSO	Muito Alta, Alta e Média Tensões	18873.2	03	0302	030235
6	2020	12	2020-12	LISBOA	Sintra	UF SINTRA	Baixa Tensão	9924017.6	11	1111	111128
7	2020	12	2020-12	PORCALEGRE	Alter do Chão	CUNHEIRA	Muito Alta, Alta e Média Tensões	17042.9	12	1201	120104
8	2020	12	2020-12	VIANA DO CASTELO	Valença	BOIVAO	Baixa Tensão	22110.5	16	1608	160802
9	2021	1	2021-01	AVEIRO	Aveiro	OLIVEIRINHA	Baixa Tensão	859496.3	01	0105	010508
10	2021	1	2021-01	BRAGANCA	Torre de Moncorvo	UF ADEGANHA E CARDANHA	Baixa Tensão	88215.5	04	0409	040918
11	2021	1	2021-01	FARO	Albufeira	PADERNE	Baixa Tensão	1025357.6	08	0801	080103
12	2021	2	2021-02	AVEIRO	Arouca	SAO MIGUEL DO MATO	Baixa Tensão	74644.6	01	0104	010417
13	2021	2	2021-02	GUARDA	Aguiar da Beira	UF SOUTO AGUIAR BEIRA VALVERDE	Baixa Tensão	93260.4	09	0901	090116
14	2021	2	2021-02	LISBOA	Torres Vedras	UF DOIS PORTOS E RUNA	Muito Alta, Alta e Média Tensões	297381.6	11	1113	111324
15	2021	2	2021-02	PORTO	Amarante	SAVADOR DO MONTE	Baixa Tensão	118804.7	13	1301	130129
16	2021	2	2021-02	PORTO	Penafiel	CROCA	Muito Alta, Alta e Média Tensões	43728.2	13	1311	131108
17	2021	3	2021-03	COIMBRA	Góis	GOIS	Baixa Tensão	419797.5	06	0606	060604
18	2021	3	2021-03	LISBOA	Alenquer	OLHALVO	Baixa Tensão	274569.1	11	1101	110108
19	2021	3	2021-03	SANTAREM	Torres Novas	UF BROQUEIRA P IGREJA ALCOROCHE	Baixa Tensão	426617.5	14	1419	141918
20	2021	3	2021-03	VILA REAL	Alijó	VILA CHA	Muito Alta, Alta e Média Tensões	125764.7	17	1701	170116
21	2021	4	2021-04	AVEIRO	Ilhavo	ILHAZO (SAO SALVADOR)	Muito Alta, Alta e Média Tensões	3817408.5	01	0110	011008
22	2021	4	2021-04	BEJA	Aljustrel	ERVIDEL	Baixa Tensão	135291.7	02	0201	020102
23	2021	4	2021-04	BEJA	Almodôvar	SAO BARNABE	Baixa Tensão	76417.8	02	0202	020206
24	2021	4	2021-04	BEJA	Beja	BALEIZAO	Muito Alta, Alta e Média Tensões	972687.5	02	0205	020502
25	2021	4	2021-04	BRAGA	Braga	UF NOGUEIRA FRAIAO E LAMACAES	Baixa Tensão	2210890	03	0303	030377
26	2021	4	2021-04	PORTO	Vila do Conde	ARVORE	Muito Alta, Alta e Média Tensões	349935.5	13	1316	131602
27	2021	4	2021-04	SETUBAL	Santiago do Cacém	ABELA	Baixa Tensão	189779.1	15	1509	150901
28	2021	4	2021-04	VIANA DO CASTELO	Ponte da Barca	UF P BARCA MUIA P V MAGALHAES	Muito Alta, Alta e Média Tensões	54819.8	16	1606	160628
29	2021	5	2021-05	BRAGA	Vila Nova de Famalicão	RIBEIRAO	Baixa Tensão	1558100.3	03	0312	031235
30	2021	5	2021-05	GUARDA	Guarda	FAIA	Muito Alta, Alta e Média Tensões	8121.8	09	0907	090716
31	2021	5	2021-05	LISBOA	Oeiras	UF CARNAXIDE E QUEIJAS	Baixa Tensão	5793445.5	11	1110	111013
32	2021	5	2021-05	PORCALEGRE	Ponte de Sor	GALVEIAS	Muito Alta, Alta e Média Tensões	36218.7	12	1213	121301
33	2021	5	2021-05	PORTO	Felgueiras	UF UNHAO E LORDELO	Baixa Tensão	116918.1	13	1303	130338

Close

Stop

Get more rows

# Configuration of Select Values 2 – Dataset B

Activities    SWT    Sun 22:57    Spoon - Project1

File Edit View Action Tools Help    Connect

View Design    output

Transformation 1 Transformation 2 lab05 ordertotals dim\_customer dim\_product dim\_time Fact\_order load\_dw Project1

100%

Input

- HL7 input
- JSON input
- YAML input

Output

- Automatic documentation output
- JSON output
- LDAP output
- Microsoft Access output
- Microsoft Excel output
- Pentaho reporting output
- Properties output
- RSS output
- S3 file output
- SQL file output
- Table output
- Text file output
- XML output
- Streaming

Transform

- Concat fields
- Row denormaliser

Utility

- Delay row

Flow

- Blocking step
- Detect empty stream
- ETL metadata injection
- Scripting

Transformation 1

Dataset\_A

Dataset\_B

Select values 2

Filter rows 2

Sort rows 2

Execution Results

Logging Execution History Step Metrics Performance

Select values

Step name: Select values 2

Select & Alter Remove Meta-data

Fields:

Fieldname	Rename to	Length	Get Fields to select
1 Year	Year_B		
2 Month	Month_B		
3 Municipality	Municipality_B		
4 Active Energy (kWh)	Active Energy(kWh)_B		

OK Cancel

Help

Stepname Copyn Read Written In

1 Dataset_B	0	0	151616	151616
2 Select values 2	0	151616	151616	0
3 Dataset_A	0	0	126710	126711
4 Select values	0	126710	126710	0
5 Filter rows	0	126710	5759	0
6 Filter rows 2	0	151616	4469	0
7 Sort rows 2	0	4469	4469	0
8 Group by	0	4469	278	0

input/output

# Preview of Select Values 2– Dataset B

Sun 22:18  
Spoon - Project1

Activities SWT

File Edit View Action Tools Help

View Design Transformation 1 Transformation 2 lab05 ordertotals dim\_customer dim\_product dim\_time fact\_order load\_dw Project1

100%

Examine preview data

Rows of step: Select values 2 (100 rows)

	Year_B	Month_B	Municipality_B	Active Energy(kWh)_B
1	2020	11	Guarda	57149.7
2	2020	11	Vila do Conde	303181.9
3	2020	11	Alcácer do Sal	2994085.1
4	2020	12	Anadia	1599645.3
5	2020	12	Barcelos	18873.2
6	2020	12	Sintra	9924017.6
7	2020	12	Alter do Chão	17042.9
8	2020	12	Valença	22110.5
9	2021	1	Aveiro	859496.3
10	2021	1	Torre de Moncorvo	88215.5
11	2021	1	Albufeira	1025357.6
12	2021	2	Arouca	74644.6
13	2021	2	Aguiar da Beira	93260.4
14	2021	2	Torres Vedras	297381.6
15	2021	2	Amarante	118804.7
16	2021	2	Penafiel	43728.2
17	2021	3	Góis	419797.5
18	2021	3	Alenquer	274569.1
19	2021	3	Torres Novas	426617.5
20	2021	3	Alijo	125764.7
21	2021	4	Ílhavo	3817408.5
22	2021	4	Aljustrel	135291.7
23	2021	4	Almodôvar	76417.8
24	2021	4	Beja	972687.5

Close Stop Get more rows

ETL metadata injection Scripting

# Configuration of Filter Rows 2– Dataset B

Activities    SWT    Sun 22:57    Spoon - Project1

File Edit View Action Tools Help

View Design

output

Transformation 1 Transformation 2 lab05 ordertotals dim\_customer dim\_product dim\_time Fact\_order load\_dw Project1

Connect

Input

- HL7 input
- JSON input
- YAML input

Output

- Automatic documentation output
- JSON output
- LDAP output
- Microsoft Access output
- Microsoft Excel output
- Pentaho reporting output
- Properties output
- RSS output
- S3 file output
- SQL file output
- Table output
- Text file output
- XML output

Streaming

Transform

- Concat fields
- Row denormaliser

Utility

- Delay row

Flow

- Blocking step
- Detect empty stream
- ETL metadata injection
- Scripting

Dataset\_A → Select rows → Group by 2 → Join rows → Select values 3 → Text file output

Filter rows 2

Step name: Filter rows 2

Send 'true' data to step:

Send 'false' data to step:

The condition:

Year\_B = [2023]

AND

Month\_B = [6]

Metrics Preview data

it	Updated	Rejected	Errors	Active	Time	Speed (r/s)	input/output
0	0	0	0	0	1.3s	118,451	
0	0	0	0	0	1.3s	117,623	
0	0	0	0	0	1.1s	114,774	
0	0	0	0	0	1.1s	110,278	
0	0	0	0	0	1.2s	109,139	
0	0	0	0	0	1.3s	116,897	
0	0	0	0	0	1.3s	3,404	
0	0	0	0	0	1.3s	3,352	

5	Filter rows	0	126710	5759	0	0	0	0
6	Filter rows 2	0	151616	4469	0	0	0	0
7	Sort rows 2	0	4469	4469	0	0	0	0
8	Group by	0	4469	278	0	0	0	0

# Preview of Filter Rows 2 – Dataset B

Activities SWT ▾ Sun 22:18 Spoon - Project1

File Edit View Action Tools Help

View Design Transformation 1 Transformation 2 lab05 ordertotals dim\_customer dim\_product dim\_time fact\_order load\_dw Project1

100%

Examine preview data

Rows of step: Filter rows 2 (100 rows)

	Year_B	Month_B	Municipality_B	Active Energy(kwh)_B
1	2023	6	Barcelos	376114.4
2	2023	6	Cabeceiras de Basto	333985.3
3	2023	6	Bragança	1089143.2
4	2023	6	Macedo de Cavaleiros	40268.9
5	2023	6	Alcobaça	919690.3
6	2023	6	Penafiel	360064.8
7	2023	6	Santo Tirso	188428.5
8	2023	6	Santarém	255940.5
9	2023	6	Arcos de Valdevez	66253.7
10	2023	6	Melgaço	150966.8
11	2023	6	Cuba	634448
12	2023	6	Portimão	723411.5
13	2023	6	Baião	106574.3
14	2023	6	Santarém	762128.2
15	2023	6	Viseu	214605.3
16	2023	6	Amares	50399.4
17	2023	6	Macedo de Cavaleiros	59809.3
18	2023	6	Vila Real de Santo António	1474251.3
19	2023	6	Alcobaça	1973409.4
20	2023	6	Lisboa	2486999.3
21	2023	6	Maia	1658333.8
22	2023	6	Abrantes	38562.8
23	2023	6	Ponte de Lima	43192.4
24	2023	6	Chaves	84624.3

Blocks Delete ETL metadata injection Scripting

# Configuration of Sort Rows 2 – Dataset B

Sun 22:57  
Spoon - Project1

Activities    SWT    Connect

File Edit View Action Tools Help

View Design Transformation 1 Transformation 2 lab05 ordertotals dim\_customer dim\_product dim\_time fact\_order load\_dw Project1

100%

output

Input

- HL7 input
- JSON input
- YAML input

Output

- Automatic document
- JSON output
- LDAP output
- Microsoft Access output
- Microsoft Excel output
- Pentaho reporting
- Properties output
- RSS output
- S3 file output
- SQL file output
- Table output
- Text file output**
- XML output
- Streaming

Transform

- Concat fields
- Row denormaliser

Utility

- Delay row

Flow

- Blocking step
- Detect empty stream
- ETL metadata injection
- Scripting

Help

Sort rows

Step name: Sort rows 2

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

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	Presc
1 Municipality_B	Y	N	N	0	N

OK Cancel Get Fields

Transformation 1

Transformation 2

lab05

ordertotals

dim\_customer

dim\_product

dim\_time

fact\_order

load\_dw

Project1

Read 4469

1169

Select values 3 Text file output

Rejected Errors Active Time Speed (r/s)

0	0	Finished	1.3s	118,451
0	0	Finished	1.3s	117,623
0	0	Finished	1.1s	114,774
0	0	Finished	1.1s	110,278
0	0	Finished	1.2s	109,139
0	0	Finished	1.3s	116,897
0	0	Finished	1.3s	3,404
0	0	Finished	1.3s	3,352

input/output

# Preview of Sort Rows 2– Dataset B

Sun 22:19  
Spoon - Project1

File Edit View Action Tools Help

View Design Transformation 1 Transformation 2 lab05 ordertotals dim\_customer dim\_product dim\_time fact\_order load\_dw Project1

100%

Examine preview data

Rows of step: Sort rows 2 (100 rows)

	Year_B	Month_B	Municipality_B	Active Energy(kwh)_B
1	2023	6	Abrantes	38562.8
2	2023	6	Abrantes	82350.5
3	2023	6	Abrantes	200941.5
4	2023	6	Abrantes	177322.7
5	2023	6	Abrantes	640829.1
6	2023	6	Abrantes	159217.5
7	2023	6	Abrantes	151997.8
8	2023	6	Abrantes	463526.4
9	2023	6	Abrantes	286989.4
10	2023	6	Abrantes	85255.8
11	2023	6	Abrantes	259332.8
12	2023	6	Abrantes	147393.7
13	2023	6	Abrantes	609789.6
14	2023	6	Abrantes	666703.8
15	2023	6	Abrantes	170560.6
16	2023	6	Abrantes	534904.6
17	2023	6	Abrantes	2429622.7
18	2023	6	Abrantes	257602.1
19	2023	6	Abrantes	564403.1
20	2023	6	Abrantes	80626.3
21	2023	6	Abrantes	2600628
22	2023	6	Aguilar da Beira	40657.7
23	2023	6	Aguilar da Beira	29085.4
24	2023	6	Aguilar da Beira	102457.9

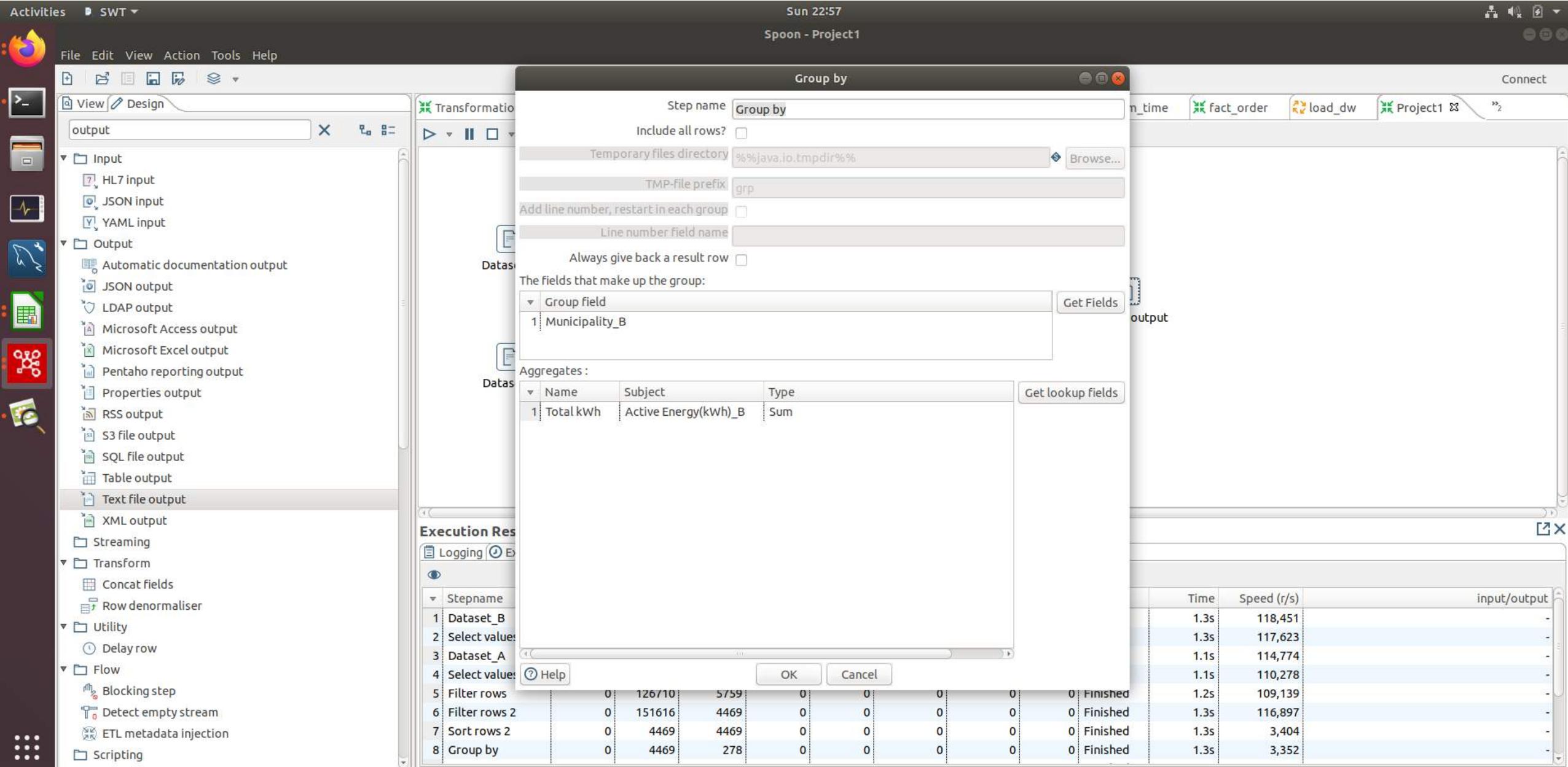
Close Stop Get more rows

2023/10/15 22:19:14 - Select values 2.0 - Finished processing (I=0, O=0, R=151010, W=151010, U=0, E=0)  
2023/10/15 22:19:14 - Filter rows 2.0 - Finished processing (I=0, O=0, R=151616, W=4469, U=0, E=0)

Activities SWT

ETL metadata injection Scripting

# Configuration of Group by – Dataset B



# Preview of Group by – Dataset B

Activities Sun 22:19  
Spoon - Project1

File Edit View Action Tools Help

View Design Transformation 1 Transformation 2 lab05 ordertotals dim\_customer dim\_product dim\_time fact\_order load\_dw Project1

100%

Examine preview data

Rows of step: Group by (100 rows)

Municipality_B	Total kwh
Abrantes	10608560.943
Aguiar da Beira	1139801.933
Alandroal	1852296.697999999
Albergaria-a-Velha	14949082.90199999
Albufeira	33869453.228
Alcanena	5636737.78799999
Alcobaça	19789133.29400003
Alcochete	6882790.095
Alcoutim	613384.366000002
Alcácer do Sal	6356300.414
Alenquer	18907304.88300005
Alfândega da Fé	763195.206
Alijó	2054365.586000006
Aljezur	1885274.2
Aljustrel	15164099.99800002
Almada	39432583.813
Almeida	1223638.615
Almeirim	6700353.45800001
Almodôvar	1465194.055000002
Alpiarça	4004363.866
Alter do Chão	856578.936
Alvaiázere	1252542.006
Alvito	1238709.864
Amadora	38839552.647

Close Stop Get more rows

2023/10/15 22:19:30 - Filter rows 2.0 - Finished processing (I=0, O=0, R=151010, W=1469, U=0, E=0)  
2023/10/15 22:19:30 - Sort rows 2.0 - Finished processing (I=0, O=0, R=4469, W=4469, U=0, E=0)

ETL metadata injection Scripting

# Configuration of Join Rows

Sun 22:57  
Spoon - Project1

Activities    SWT

File Edit View Action Tools Help

View Design

Transformation 1 Transformation 2 lab05 ordertotals dim\_customer dim\_product dim\_time Fact\_order load\_dw Project1

100%

Output

Input

- HL7 input
- JSON input
- YAML input

Output

- Automatic documentation output
- JSON output
- LDAP output
- Microsoft Access output
- Microsoft Excel output
- Pentaho reporting output
- Properties output
- RSS output
- S3 file output
- SQL file output
- Table output
- Text file output
- XML output

Streaming

Transform

- Concat fields
- Row denormaliser

Utility

- Delay row

Flow

- Blocking step
- Detect empty stream
- ETL metadata injection

Scripting

Join rows (cartesian product)

Step name: Join rows

Temp directory: %java.io.tmpdir%

TMP-file prefix: out

Max. cache size (in rows): 500

Main step to read from:

The condition:

Municipality\_A = Municipality\_B

Execution

Logging OK Cancel

Preview data

Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	input/output
1 Dataset_B	0	0	151616	151617	0	0	0	0	Finished	1.3s	118,451	-
2 Select values 2	0	151616	151616	0	0	0	0	0	Finished	1.3s	117,623	-
3 Dataset_A	0	0	126710	126711	0	0	0	0	Finished	1.1s	114,774	-
4 Select values	0	126710	126710	0	0	0	0	0	Finished	1.1s	110,278	-
5 Filter rows	0	126710	5759	0	0	0	0	0	Finished	1.2s	109,139	-
6 Filter rows 2	0	151616	4469	0	0	0	0	0	Finished	1.3s	116,897	-
7 Sort rows 2	0	4469	4469	0	0	0	0	0	Finished	1.3s	3,404	-
8 Group by	0	4469	278	0	0	0	0	0	Finished	1.3s	3,352	-

# Preview of Join Rows

Sun 22:19  
Spoon - Project1

Activities SWT

File Edit View Action Tools Help

View Design Transformation 1 Transformation 2 lab05 ordertotals dim\_customer dim\_product dim\_time fact\_order load\_dw Project1

output

Input Output

Rows of step: Join rows (100 rows)

	Municipality_A	Total CPE's	Municipality_B	Total kWh
1	Abrantes	25180	Abrantes	10608560.943
2	Aguiar da Beira	4933	Aguiar da Beira	1139801.933
3	Alandroal	4133	Alandroal	1852296.697999999
4	Albergaria-a-Velha	14241	Albergaria-a-Velha	14949082.90199999
5	Albufeira	54157	Albufeira	33869453.228
6	Alcanena	8265	Alcanena	5636737.78799999
7	Alcobaça	39245	Alcobaça	19789133.29400003
8	Alcochete	11819	Alcochete	6882790.095
9	Alcoutim	3703	Alcoutim	613384.366000002
10	Alcácer do Sal	8480	Alcácer do Sal	6356300.414
11	Alenquer	24858	Alenquer	18907304.88300005
12	Alfândega da Fé	4167	Alfândega da Fé	763195.206
13	Alijó	9146	Alijó	2054365.586000006
14	Aljezur	6523	Aljezur	1885274.2
15	Aljustrel	6058	Aljustrel	15164099.99800002
16	Almada	116447	Almada	39432583.813
17	Almeida	7014	Almeida	1223638.615
18	Almeirim	14117	Almeirim	6700353.45800001
19	Almodôvar	5693	Almodôvar	1465194.055000002
20	Alpiarça	4630	Alpiarça	4004363.866
21	Alter do Chão	2892	Alter do Chão	856578.936
22	Alvaiázere	5897	Alvaiázere	1252542.006
23	Alvito	1725	Alvito	1238709.864
24	Amadora	98733	Amadora	38839552.647

Examine preview data

Close Stop Get more rows

2023/10/15 22:19:40 - 500 Rows 2.0 - Finished processing (i=0, O=0, R=4469, W=4469, U=0, L=0)  
2023/10/15 22:19:46 - Group by\_0 - Finished processing (i=0, O=0, R=4469, W=278, U=0, E=0)

ETL metadata injection Scripting

# Configuration of Select Values 3

Activities SWT Sun 22:57 Spoon - Project1

File Edit View Action Tools Help Connect

View Design output

Input  
HL7 input JSON input YAML input

Output  
Automatic documentation output JSON output LDAP output Microsoft Access output Microsoft Excel output Pentaho reporting output Properties output RSS output S3 file output SQL file output Table output Text file output XML output Streaming

Transform  
Concat fields Row denormaliser

Utility  
Delay row

Flow  
Blocking step Detect empty stream ETL metadata injection Scripting

Transformation 1 Transformation 2 lab05 ordertotals dim\_customer dim\_product dim\_time fact\_order load\_dw Project1

100%

Dataset\_A → Select values → Filter rows → Sort rows

Dataset\_B → Select values 2 → Filter rows 2 → Sort rows 2

Select values Step name: Select values 3

Select & Alter Remove Meta-data

Fields:

Fieldname	Rename to	Length	Precision	Get fields to select
1 Municipality_A	Municipality			Edit Mapping
2 Total CPE's	CPEs			
3 Total kwh	kwh			

Execution Results

Stepname	Copynr	Read	Written	In	Out	Help	OK	Cancel
1 Dataset_B	0	0	151616	151616	0	0	0	0
2 Select values 2	0	151616	151616	0	0	0	0	0
3 Dataset_A	0	0	126710	126711	0	0	0	0
4 Select values	0	126710	126710	0	0	0	0	0
5 Filter rows	0	126710	5759	0	0	0	0	0
6 Filter rows 2	0	151616	4469	0	0	0	0	0
7 Sort rows 2	0	4469	4469	0	0	0	0	0
8 Group by	0	4469	278	0	0	0	0	0

Logging Execution History Step Metrics Performance

Include unspecified fields, ordered |

input/output

# Preview of Select Values 3

Sun 22:20  
Spoon - Project1

File Edit View Action Tools Help

View Design Transformation 1 Transformation 2 lab05 ordertotals dim\_customer dim\_product dim\_time fact\_order load\_dw Project1

100%

Examine preview data

Rows of step: Select values 3 (100 rows)

	Municipality	CPEs	kWh
1	Abrantes	25180	10608560.943
2	Aguiar da Beira	4933	1139801.933
3	Alandroal	4133	1852296.697999999
4	Albergaria-a-Velha	14241	14949082.90199999
5	Albufeira	54157	33869453.228
6	Alcanena	8265	5636737.78799999
7	Alcobaça	39245	19789133.29400003
8	Alcochete	11819	6882790.095
9	Alcoutim	3703	613384.366000002
10	Alcácer do Sal	8480	6356300.414
11	Alenquer	24858	18907304.88300005
12	Alfândega da Fé	4167	763195.206
13	Alijó	9146	2054365.586000006
14	Aljezur	6523	1885274.2
15	Aljustrel	6058	15164099.99800002
16	Almada	116447	39432583.813
17	Almeida	7014	1223638.615
18	Almeirim	14117	6700353.45800001
19	Almodôvar	5693	1465194.055000002
20	Alpiarça	4630	4004363.866
21	Alter do Chão	2892	856578.936
22	Alvaiázere	5897	1252542.006
23	Alvito	1725	1238709.864
24	Amadora	98733	38839552.647

Close Stop Get more rows

2023/10/15 22:20:08 - 5000 ROWS 2.0 - FINISHED processing (I=0, O=0, R=1103, W=1103, U=0, E=0)  
2023/10/15 22:20:08 - Group by 0 - Finished processing (I=0, O=0, R=4469, W=278, U=0, E=0)

Activities SWT

ETL metadata injection Scripting

# Configuration of Text File Output

Sun 22:57  
Spoon - Project1

File Edit View Action Tools Help

Text file output

Step name Text file output

File Content Fields

Filename \${Internal.Entry.Current.Directory}/Project1\_Exercise1

Pass output to servlet

Create Parent Folder

Do not create file at start

Accept file name from field?

File name field

Extension csv

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

OK Cancel

dim\_product dim\_time fact\_order load\_dw Project1

Output 102  
Updated 0  
Rejected 0  
Errors 0  
Active Paused  
Time 37mn 14s  
Speed (r/s) 0  
input/output 100/0

Select values 3 Text file output

Dataset\_B Stepname Copynr Read Written Input Output Updated Rejected Errors Active Time Speed (r/s) input/output

1	Dataset_B	0	0	151616	151617	0	0	0	0	Finished	1.3s	118,451	-
2	Select values 2	0	151616	151616	0	0	0	0	0	Finished	1.3s	117,623	-
3	Dataset_A	0	0	126710	126711	0	0	0	0	Finished	1.1s	114,774	-
4	Select values	0	126710	126710	0	0	0	0	0	Finished	1.1s	110,278	-
5	Filter rows	0	126710	5759	0	0	0	0	0	Finished	1.2s	109,139	-
6	Filter rows 2	0	151616	4469	0	0	0	0	0	Finished	1.3s	116,897	-
7	Sort rows 2	0	4469	4469	0	0	0	0	0	Finished	1.3s	3,404	-
8	Group by	0	4469	278	0	0	0	0	0	Finished	1.3s	3,352	-

# Preview of Text File Output

Activities SWT

Tue 21:49

Rows of step: Text file output (100 rows)

	Municipality	CPEs	kwh
1	Abrantes	25180	10608560.943
2	Aguiar da Beira	4933	1139801.933
3	Alandroal	4133	1852296.697999999
4	Albergaria-a-Velha	14241	14949082.90199999
5	Albufeira	54157	33869453.228
6	Alcanena	8265	5636737.78799999
7	Alcobaça	39245	19789133.294000003
8	Alcochete	11819	6882790.095
9	Alcoutim	3703	613384.3660000002
10	Alcácer do Sal	8480	6356300.414
11	Alenquer	24858	18907304.883000005
12	Alfândega da Fé	4167	763195.206
13	Alijó	9146	2054365.5860000006
14	Aljezur	6523	1885274.2
15	Aljustrel	6058	15164099.998000002
16	Almada	116447	39432583.813
17	Almeida	7014	1223638.615
18	Almeirim	14117	6700353.458000001
19	Almodôvar	5693	1465194.0550000002
20	Alpiarça	4630	4004363.866
21	Alter do Chão	2892	856578.936
22	Alvaiázere	5897	1252542.006
23	Alvito	1725	1238709.864
24	Amadora	98733	38839552.647
25	Amarante	31797	9327311.992000004
26	Amares	11329	3541293.486
27	Anadia	17295	9934817.32099999
28	Ansião	9048	2615723.724000004
29	Arcos de Valdevez	19112	7771312.0249999985
30	Arganil	11140	3587683.684
31	Armamar	5147	1476715.985999998
32	Arouca	12670	4100205.77899999
33	Arraiolos	4874	2089171.2529999996
34	Arronches	2319	799418.363

Close Stop Get more rows

# Output File

Activities LibreOffice Calc ▾ Sun 22:11

Name

File Edit View Insert Format Styles Sheet Data Tools Window Help

A1 Municipality

	A	B	C	D	E	F	G	H	I	J
1	Municipality	CPEs	kWh							
2	Abrantes	25180	10608560.943							
3	Aguilar da Beira	4933	1139801.933							
4	Alandroal	4133	1852296.698							
5	Albergaria-a-Velha	14241	14949082.902							
6	Albufeira	54157	33869453.228							
7	Alcanena	8265	5636737.788							
8	Alcobaça	39245	19789133.294							
9	Alcochete	11819	6882790.095							
10	Alcoutim	3703	613384.366							
11	Alcácer do Sal	8480	6356300.414							
12	Alenquer	24858	18907304.883							
13	Alfândega da Fé	4167	763195.206							
14	Alijó	9146	2054365.586							
15	Aljezur	6523	1885274.2							
16	Aljustrel	6058	15164099.998							
17	Almada	116447	39432583.813							
18	Almeida	7014	1223638.615							
19	Almerim	14117	6700353.458							
20	Almodôvar	5693	1465194.055							
21	Alpiarça	4630	4004363.866							
22	Alter do Chão	2892	856578.936							
23	Alvalázere	5897	1252542.006							
24	Alvito	1725	1238709.864							
25	Amadora	98733	38839552.647							
26	Amarante	31797	9327311.992							
27	Amares	11329	3541293.486							
28	Anadia	17295	9934817.321							
29	Ansião	9048	2615723.724							
30	Arcos de Valdevez	19112	7771312.025							
31	Arganil	11140	3587683.684							
32	Armamar	5147	1476715.986							
33	Arouca	12670	4100205.779							
34	Arraiolos	4874	2089171.253							
35	Arronches	2319	799418.363							
36	Arruda dos Vinhos	7678	2990286.882							
37	Aveiro	51443	6004287.198							
38	Avis	3182	4497951.333							
39	Azambuja	12414	12789492.241							

Project1\_Exercise1

Sheet 1 of 1 Default English (USA) Average: ; Sum: 0 100%

Size Modified

15.0 MB 19:14

11.0 MB 19:14

35.3 kB 22:10

8.4 kB 22:10

"Project1\_Exercise1.csv" selected (8.4 kB)

# Exercise 2 – Group 50

José Cardoso

Matheus Trindade

# Analysis Job

Activities org-datacleaner-Main ▾

Fri 17:05

Municipality consumption | Analysis Job | DataCleaner

New Open Save Save As... More ▾ Execute | ↗

Click here to run job

Municipality consumption

information\_schema

Ex2

Project1\_Exercise1.csv

default\_table

Library

Transform

Improve

Analyze

Date and time

Machine learning

Visualization

- Density plot
- Scatter plot
- Stacked area plot

Boolean analyzer

Character set distribution

Completeness analyzer

Fill pattern

Mark rows

Number analyzer

Pattern finder

Reference data matcher

Referential integrity

String analyzer

Unique key check

Value distribution

Value matcher

Write

Project1\_Exercise1.csv

Convert to number

Scatter plot

Ready to execute

Click the 'Execute' button in the upper-right corner when you're ready to run the job.

Search component library... X

Job is correctly configured

Community edition - Up to date

# Configuration Window - CSV

Activities

Fri 17:06

Municipality consumption | Analysis Job | DataCleaner

New Open Save Save As... More Execute

Click here to run job

Municipality consumption information\_schema Ex2 Project1\_Exercise1.csv default\_table Library Transform Improve Analyze Date and time Machine learning Visualization Density plot Scatter plot Stacked area plot Boolean analyzer Character set distribution Completeness analyzer Fill pattern Mark rows Number analyzer Pattern finder Reference data matcher Referential integrity String analyzer Unique key check Value distribution Value matcher Write

hide tree

Project1\_Exercise1.csv | DataCleaner

Project1\_Exercise1.csv

Ex2.Project1\_Exercise1.csv

Name	Type
Municipality	String
CPEs	String
kWh	String

Close

Search component library... X

Job is correctly configured

Community edition - Up to date

The screenshot shows the DataCleaner application interface. At the top, there's a toolbar with 'Activities', the current time 'Fri 17:06', and tabs for 'Municipality consumption', 'Analysis Job', and 'DataCleaner'. Below the toolbar is a menu bar with 'New', 'Open', 'Save', 'Save As...', 'More', 'Execute', and other options. On the left, a sidebar lists various tools and components under categories like 'Library', 'Transform', 'Improve', and 'Analyze'. A specific project 'Ex2' is selected, containing files 'Project1\_Exercise1.csv' and 'default\_table'. A large central workspace displays a data flow diagram: 'Project1\_Exercise1.csv' is connected to a 'Convert to number' component, which then feeds into a 'Scatter plot' component. To the right of the workspace, a callout bubble says 'Click here to run job'. In the bottom right corner, there's a status message 'Community edition - Up to date'. The main workspace also contains a configuration window for 'Project1\_Exercise1.csv' with a table mapping column names to types: 'Municipality', 'CPEs', and 'kWh' all mapped to 'String'. A 'Close' button is at the bottom of this window. A search bar for the component library is at the bottom left, and a status message 'Job is correctly configured' is at the bottom center.

# Configuration Window – Convert to number

Activities

Fri 17:06

Municipality consumption | Analysis Job | DataCleaner

New Open Save Save As... More Execute

Municipality consumption information\_schema Ex2 Project1\_Exercise1.csv default\_table Library Transform Improve Analyze Date and time Machine learning Visualization Density plot Scatter plot Stacked area plot Boolean analyzer Character set distribution Completeness analyzer Fill pattern Mark rows Number analyzer Pattern finder Reference data matcher Referential integrity String analyzer Unique key check Value distribution Value matcher Write

Search component library... X

Job is correctly configured

Convert to number | DataCleaner

# Convert to number

Documentation Rename

Input columns

Input: Select all Select none Municipality CPEs kWh

Required properties

Decimal separator: Minus sign: Thousand separator:

Optional properties (1)

Output columns

Name	Type
Municipality (as number)	Number
CPEs (as number)	Number
kWh (as number)	Number

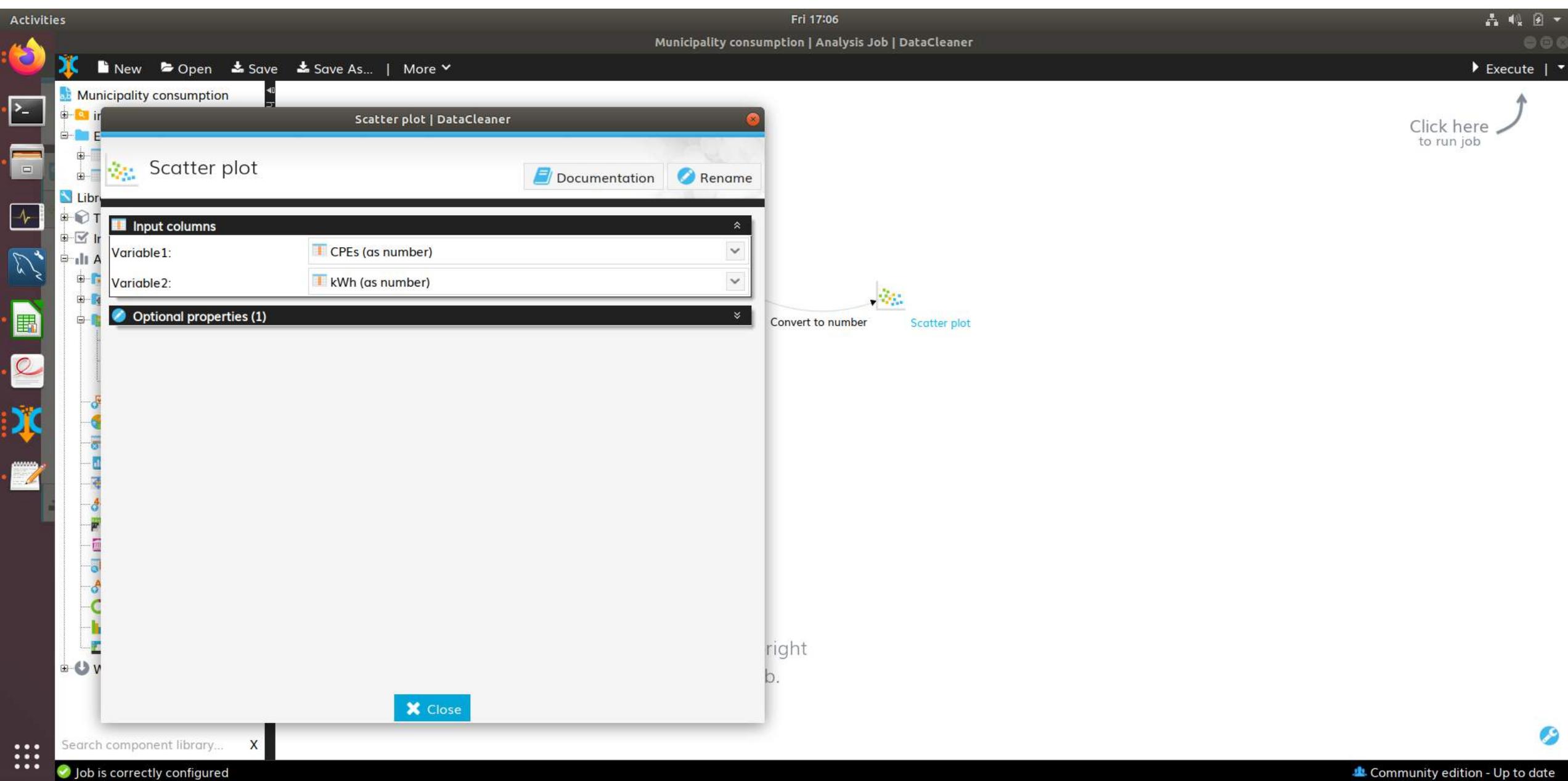
Write data Preview data | Close

Ready to execute

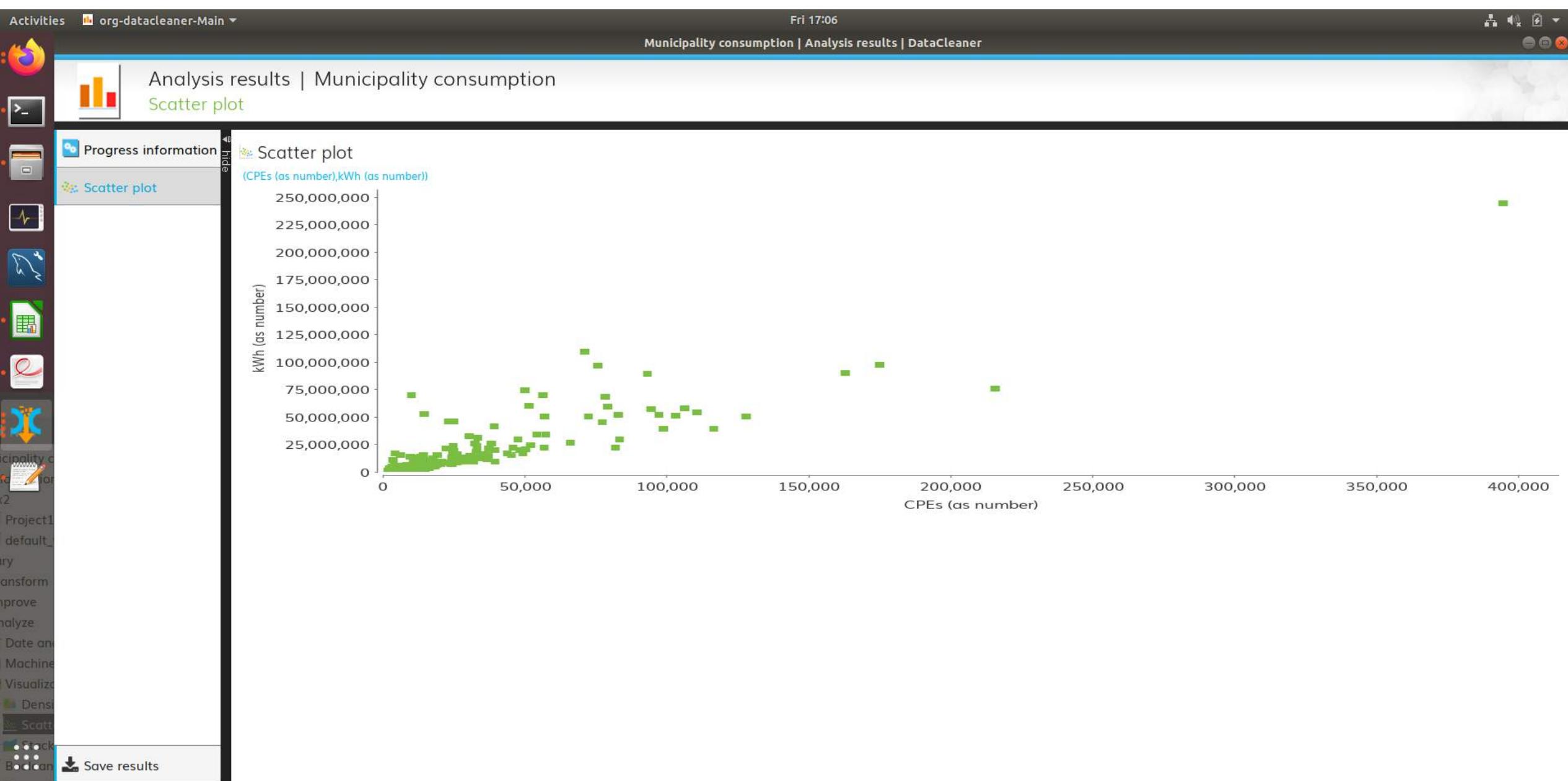
Click the 'Execute' button in the upper-right corner when you're ready to run the job.

The screenshot shows the Talend Data Integration interface. On the left, there's a sidebar with various activity icons and a tree view of a project named 'Municipality consumption'. The main workspace displays a workflow with a 'Project1\_Exercise1.csv' source connected to a 'Convert to number' component. A large play button icon is overlaid on the bottom-left of the workspace. To the right, a detailed configuration window for the 'Convert to number' job is open, showing sections for input columns (with 'Municipality', 'CPEs', and 'kWh' selected), required properties (decimal, minus, and thousand separators), optional properties (one entry), and output columns (three columns named 'Municipality (as number)', 'CPEs (as number)', and 'kWh (as number)' all set to 'Number'). At the bottom of the configuration window are 'Write data' and 'Preview data' buttons, along with a 'Close' button. The top status bar indicates it's Friday at 17:06, and the title bar says 'Municipality consumption | Analysis Job | DataCleaner'.

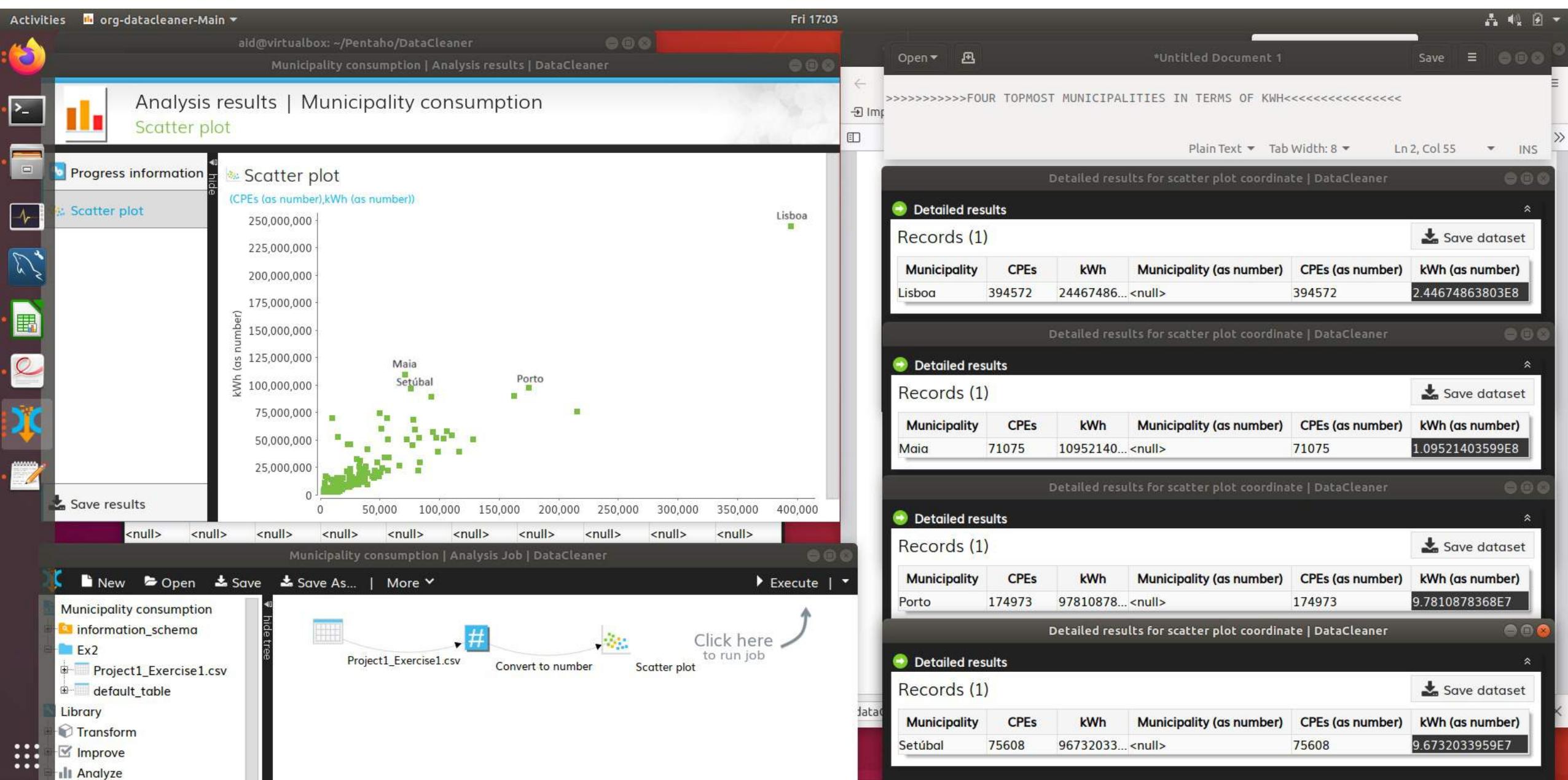
## • Configuration Window – Scatter Plot



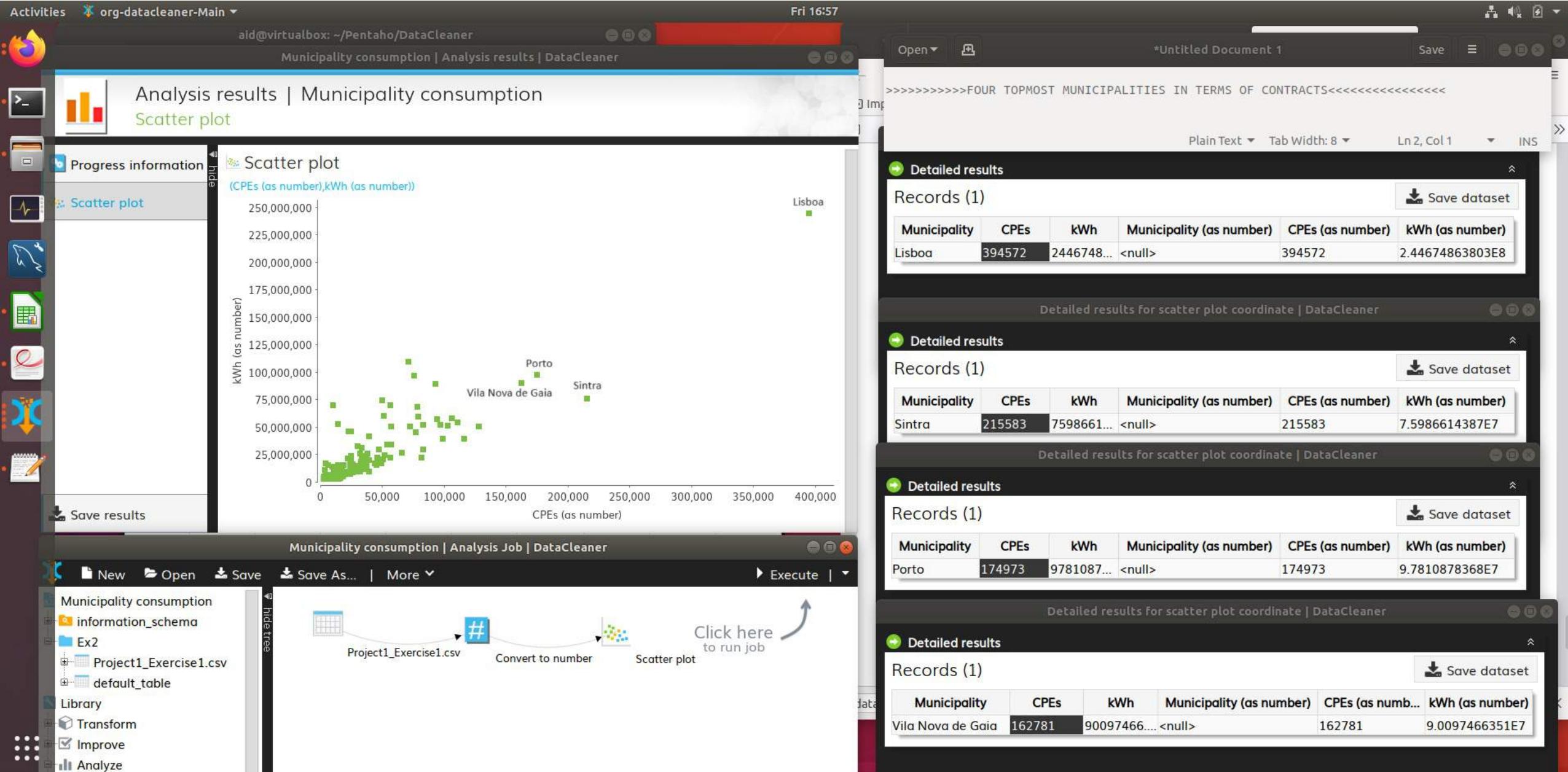
# • Plot– Analysis Results



- Top 4 municipalities - KWH



- Top 4 municipalities - CPE



# Exercise 3 – Group 50

José Cardoso

Matheus Trindade

# Transformation

Activities SWT

Sat 22:18  
Spoon - EX3

File Edit View Action Tools Help

View Design Search

Transformations

- EX3
  - Run configurations
  - Database connections
  - Steps
  - Hops
  - Partition schemas
  - Slave server
  - Kettle cluster schemas
  - VFS Connections

Connect

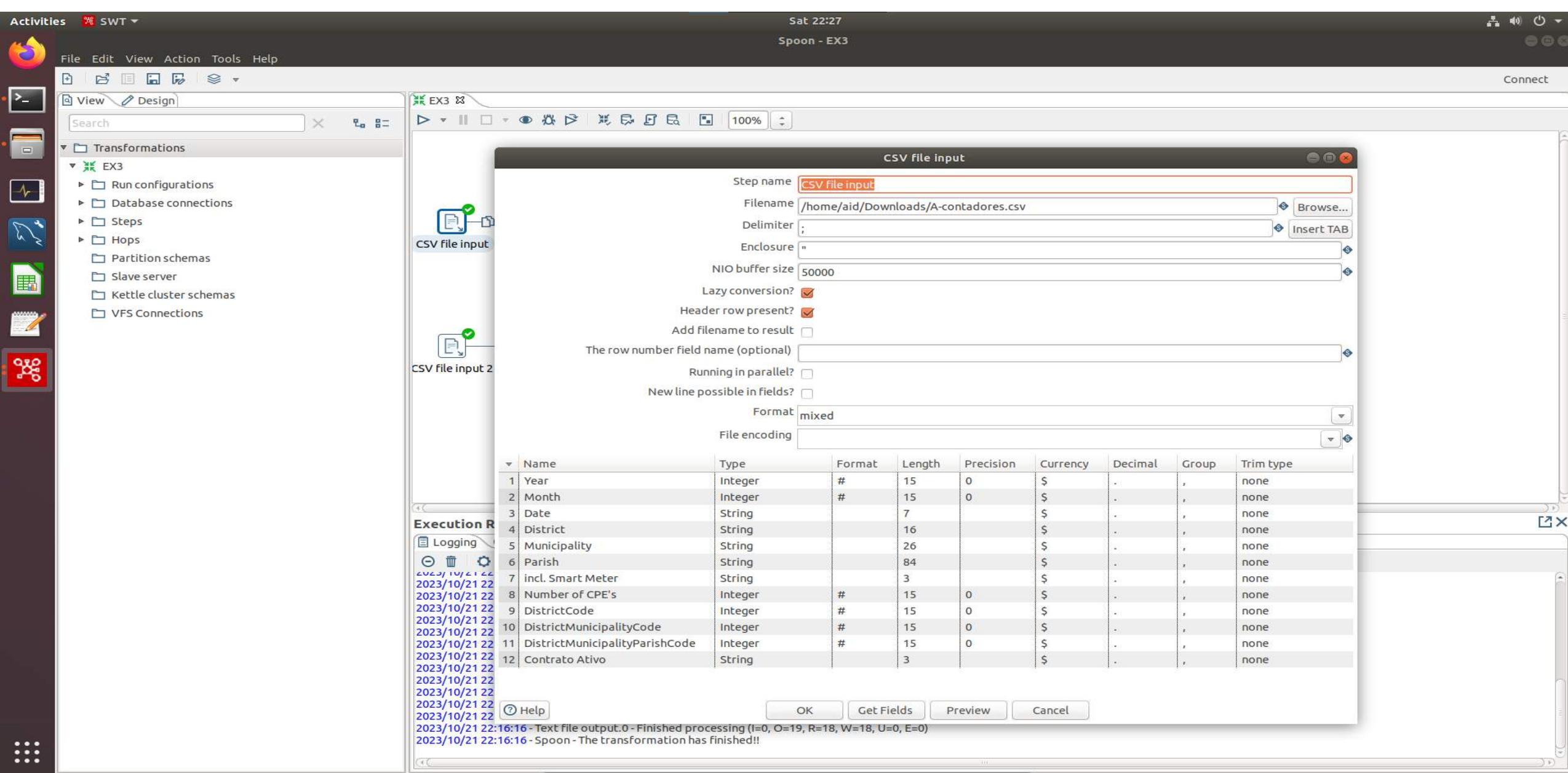
The diagram illustrates a kettle transformation named EX3. It features two parallel data flows. The top flow starts with a "CSV file input" step, followed by "Select values", "Sort rows", and "Group by" steps. The output of "Group by" is connected to a "Join rows" step, which then feeds into a "Calculator" step. Subsequent steps include "Sort rows 3" and "Group by 3", leading to a final "Text file output". The bottom flow follows a similar sequence: "CSV file input 2", "Select values 2", "Sort rows 2", and "Group by 2". The output of "Group by 2" is also connected to the "Join rows" step.

Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Preview data

2023/10/21 22:16:14 - SELECT values 2.0 - Finished processing (I=0, O=0, R=147107, W=147107, U=0, E=0)  
2023/10/21 22:16:15 - Group by 0 - Linenr 50000  
2023/10/21 22:16:15 - Group by 0 - Linenr 100000  
2023/10/21 22:16:15 - Sort rows 0 - Finished processing (I=0, O=0, R=120943, W=120943, U=0, E=0)  
2023/10/21 22:16:15 - Group by 0 - Finished processing (I=0, O=0, R=120943, W=18, U=0, E=0)  
2023/10/21 22:16:15 - Group by 2.0 - Linenr 5000  
2023/10/21 22:16:16 - Group by 2.0 - Linenr 100000  
2023/10/21 22:16:16 - Sort rows 2.0 - Finished processing (I=0, O=0, R=147107, W=147107, U=0, E=0)  
2023/10/21 22:16:16 - Group by 2.0 - Finished processing (I=0, O=0, R=147107, W=18, U=0, E=0)  
2023/10/21 22:16:16 - Join rows 0 - Finished processing (I=0, O=0, R=36, W=324, U=0, E=0)  
2023/10/21 22:16:16 - Calculator 0 - Finished processing (I=0, O=0, R=324, W=324, U=0, E=0)  
2023/10/21 22:16:16 - Sort rows 3.0 - Finished processing (I=0, O=0, R=324, W=324, U=0, E=0)  
2023/10/21 22:16:16 - Group by 3.0 - Finished processing (I=0, O=0, R=324, W=18, U=0, E=0)  
2023/10/21 22:16:16 - Text file output 0 - Finished processing (I=0, O=19, R=18, W=18, U=0, E=0)  
2023/10/21 22:16:16 - Spoon - The transformation has finished!!

# Configuration of CSV File Input – Dataset A



# Preview of CSV File Input – Dataset A

Activities SWT

Sun 00:10  
Spoon - EX3

File Edit View Action Tools Help

View Design

Welcome! EX3

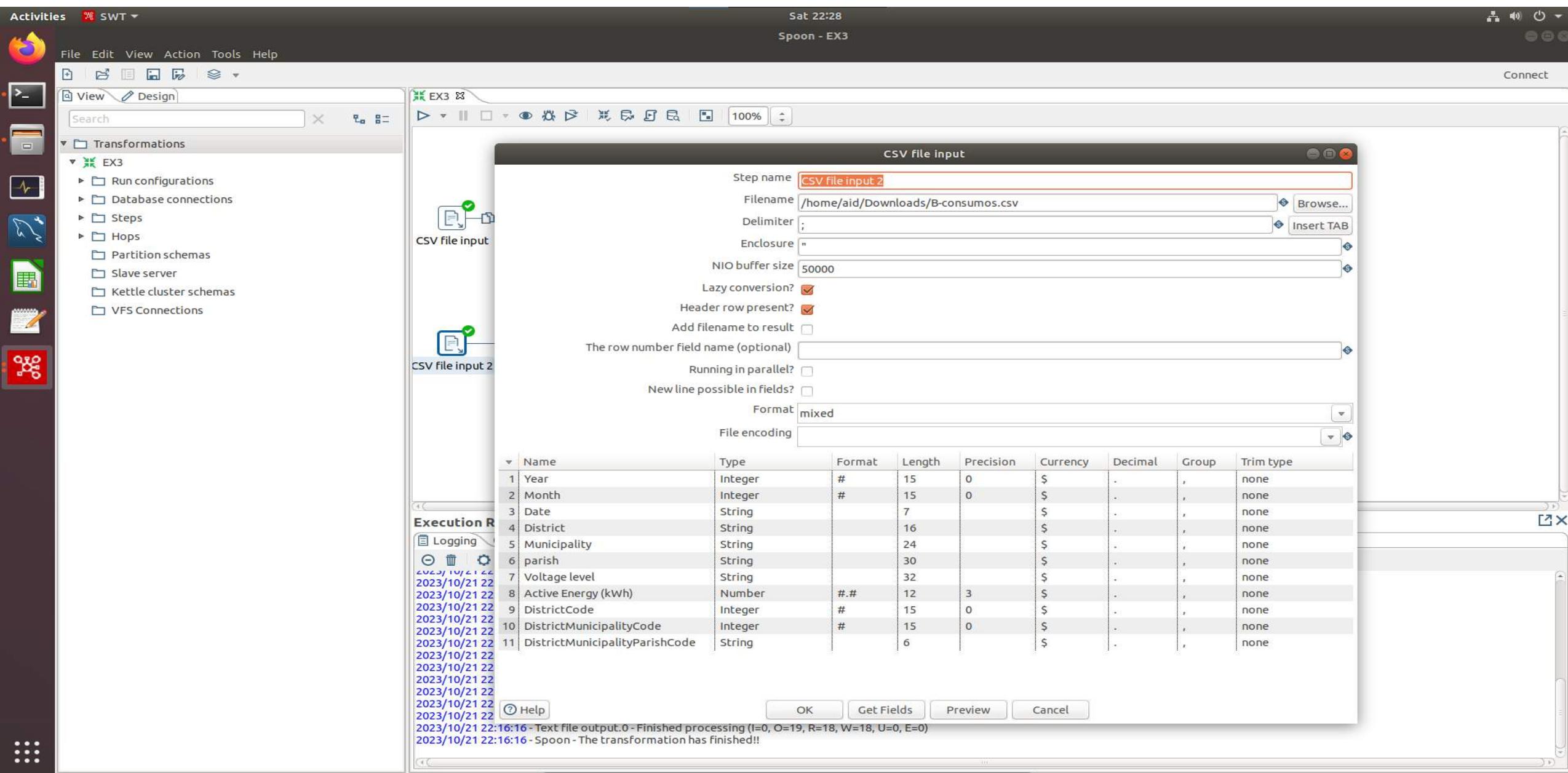
Connect

Examine preview data

Rows of step: CSV file input (1000 rows)

	Year	Month	Date	District	Municipality	Parish	incl. Smart Meter	Number of CPE's	DistrictCode	District
1	2022	7	2022-07	Beja	Cuba	Cuba	NÃO	163	2	Alentejo
2	2022	9	2022-09	Évora	Portel	Santana	NÃO	221	7	Alentejo
3	2022	9	2022-09	Bragança	Mogadouro	Castro Vicente	SIM	42	4	Chaves
4	2022	5	2022-05	Braga	Fafe	União de freguesias de Monte e Queimadela	NÃO	545	3	Chaves
5	2022	5	2022-05	Leiria	Caldas da Rainha	A dos Francos	NÃO	612	10	Leiria
6	2022	6	2022-06	Aveiro	Espinho	Silvalde	NÃO	453	1	Aveiro
7	2022	6	2022-06	Braga	Guimarães	Mesão Frio	NÃO	134	3	Chaves
8	2023	3	2023-03	Braga	Braga	Pedralva	NÃO	333	3	Chaves
9	2023	3	2023-03	Bragança	Vila Flor	União das freguesias de Valtorno e Mourão	SIM	93	4	Chaves
10	2023	7	2023-07	Guarda	Seia	Alvoco da Serra	NÃO	323	9	Guarda
11	2023	9	2023-09	Viseu	Tondela	Dardavaz	SIM	188	18	Viseu
12	2022	7	2022-07	Porto	Lousada	União das freguesias de Nespereira e Casais	SIM	486	13	Porto
13	2022	7	2022-07	Santarém	Ourém	Nossa Senhora da Piedade	SIM	4951	14	Santarém
14	2022	2	2022-02	Guarda	Celorico da Beira	Vale de Azares	SIM	101	9	Guarda
15	2022	1	2022-01	Vila Real	Montalegre	Cervos	SIM	62	17	Vila Real
16	2023	2	2023-02	Bragança	Vinhais	União das freguesias de Moimenta e Montouto	NÃO	215	4	Chaves
17	2022	4	2022-04	Castelo Branco	Castelo Branco	Salgueiro do Campo	NÃO	566	5	Castelo Branco
18	2022	11	2022-11	Guarda	Guarda	Vela	SIM	238	9	Guarda
19	2023	6	2023-06	Guarda	Celorico da Beira	Vale de Azares	NÃO	163	9	Guarda
20	2023	5	2023-05	Évora	Borba	Orada	NÃO	315	7	Évora
21	2022	9	2022-09	Viseu	Moimenta da Beira	União das freguesias de Peva e Segões	SIM	185	18	Viseu
22	2022	9	2022-09	Évora	Alandroal	Santiago Maior	SIM	340	7	Évora
23	2022	5	2022-05	Viseu	Nelas	Canas de Senhorim	SIM	1066	18	Viseu
24	2022	5	2022-05	Vila Real	Montalegre	União das freguesias de Viade de Baixo e Fervidelas	SIM	110	17	Vila Real
25	2022	5	2022-05	Guarda	Gouveia	Arcozelo	SIM	386	9	Guarda
26	2022	12	2022-12	Viana do Castelo	Ponte da Barca	União das freguesias de Ponte da Barca, Vila Nova de Muíá e Paço Vedro de Magalhães	NÃO	542	16	Viana do Castelo
27	2022	12	2022-12	Porto	Felgueiras	União das freguesias de Matriz e Lixa e Caramos	SIM	575	13	Porto

# Configuration of CSV File Input – Dataset B



# Preview of CSV File Input – Dataset B

Activities SWT

Sun 00:11 Spoon - EX3

File Edit View Action Tools Help

View Design Connect

Search

Transformations

Rows of step: CSV file input 2 (1000 rows)

	Year	Month	Date	District	Municipality	parish	Voltage level	Active Energy (kWh)	DistrictCode	DistrictMunicipalityCode	DistrictMu
1	2021	3	2021-03	PORTO	Felgueiras	REGILDE	Baixa Tensão	181866.9	13	1303	130318
2	2021	3	2021-03	PORTO	Paredes	SOBROSA	Muito Alta, Alta e Média Tensões	273375	13	1310	131021
3	2021	3	2021-03	SANTAREM	Golegã	POMBALINHO	Baixa Tensão	79022.3	14	1412	141203
4	2021	3	2021-03	SETUBAL	Sesimbra	SESIMBRA (SANTIAGO)	Muito Alta, Alta e Média Tensões	263582.7	15	1511	151102
5	2021	3	2021-03	VISEU	Viseu	FRAGOSELA	Baixa Tensão	508651.7	18	1823	182315
6	2021	4	2021-04	COIMBRA	Coimbra	UF S MARTINHO ARVORE LAMAROSA	Baixa Tensão	342794	6	603	060337
7	2021	4	2021-04	EVORA	Vila Viçosa	BENCATEL	Baixa Tensão	200918.4	7	714	071401
8	2021	4	2021-04	GUARDA	Gouveia	VILA NOVA DE TAZEM	Muito Alta, Alta e Média Tensões	6132.1	9	906	090621
9	2021	4	2021-04	GUARDA	Manteigas	MANTEIGAS (SANTA MARIA)	Baixa Tensão	303825	9	908	090802
10	2021	4	2021-04	GUARDA	Vila Nova de Foz Côa	VILA NOVA DE FOZ COA	Muito Alta, Alta e Média Tensões	237087.9	9	914	091419
11	2021	4	2021-04	LISBOA	Sintra	COLARES	Baixa Tensão	2110346.3	11	1111	111105
12	2021	4	2021-04	PORTO	Penafiel	CABECA SANTA	Muito Alta, Alta e Média Tensões	370909.4	13	1311	131104
13	2021	4	2021-04	VIANA DO CASTELO	Arcos de Valdevez	SISTELO	Baixa Tensão	26991.6	16	1601	160145
14	2021	4	2021-04	VISEU	Mortágua	PALA	Baixa Tensão	84457.9	18	1808	180807
15	2021	5	2021-05	BRAGA	Barcelos	UF BARCELOS V BOA V FRESCAINHA	Muito Alta, Alta e Média Tensões	2794312.3	3	302	030293
16	2021	5	2021-05	CASTELO BRANCO	Sertã	CABECUDO	Baixa Tensão	120004.3	5	509	050901
17	2021	5	2021-05	COIMBRA	Arganil	CELAVISA	Baixa Tensão	18459.6	6	601	060105
18	2021	5	2021-05	LISBOA	Alenquer	UF RIBAFRIA PEREIRO PALHACANA	Baixa Tensão	199535.6	11	1101	110121
19	2021	5	2021-05	LISBOA	Azambuja	ALCOENTRE	Baixa Tensão	409862.9	11	1103	110301
20	2021	5	2021-05	SANTAREM	Tomar	SAO PEDRO DE TOMAR	Baixa Tensão	432834.4	14	1418	141813
21	2021	5	2021-05	VIANA DO CASTELO	Caminha	VILE	Baixa Tensão	37537.5	16	1602	160220
22	2021	5	2021-05	VILA REAL	Montalegre	UF PARADELA CONTIM E FIAES	Muito Alta, Alta e Média Tensões	15211.1	17	1706	170639
23	2021	5	2021-05	VILA REAL	Sabrosa	PARADA DE PINHAO	Baixa Tensão	44249.6	17	1710	171005
24	2021	5	2021-05	VISEU	Penalva do Castelo	INSUA	Muito Alta, Alta e Média Tensões	95597.8	18	1811	181105
25	2021	6	2021-06	BEJA	Serpa	UF V NOVA S BENTO E VALE VARGO	Muito Alta, Alta e Média Tensões	86142.7	2	213	021309
26	2021	6	2021-06	LISBOA	Amadora	ENCOSTA DO SOL	Baixa Tensão	3330354.3	11	1115	111514
27	2021	6	2021-06	PORTO	Santo Tirso	ACRELA	Muito Alta, Alta e Média Tensões	57220.3	13	1314	131401

Examine preview data

Close Stop Get more rows

# Configuration of Select Values – Dataset A

Sat 22:28  
Spoon - EX3

Activities SWT

File Edit View Action Tools Help

View Design

Search

Transformations

- EX3
  - Run configurations
  - Database connections
  - Steps
  - Hops
  - Partition schemas
  - Slave server
  - Kettle cluster schemas
  - VFS Connections

Connect

CSV file input → Select values → Sort rows → Group by

CSV file input 2 → Select values

Select values

Step name: Select values

Select & Alter Remove Meta-data

Fields:

Fieldname	Rename to	Length	Precision
1 District	District_A		

Get fields to select Edit Mapping

Include unspecified fields, ordered by name

OK Cancel

Execution Results

Logging Help

2023/10/21 22:16:14 - Select values 2.0 - Finished processing (I=0, O=0, R=147107, W=147107, U=0, E=0)  
2023/10/21 22:16:15 - Group by 0 - Linenr 50000  
2023/10/21 22:16:15 - Group by 0 - Linenr 100000  
2023/10/21 22:16:15 - Sort rows 0 - Finished processing (I=0, O=0, R=120943, W=120943, U=0, E=0)  
2023/10/21 22:16:15 - Group by 0 - Finished processing (I=0, O=0, R=120943, W=18, U=0, E=0)  
2023/10/21 22:16:15 - Group by 2 0 - Linenr 50000  
2023/10/21 22:16:16 - Group by 2 0 - Linenr 100000  
2023/10/21 22:16:16 - Sort rows 2 0 - Finished processing (I=0, O=0, R=147107, W=147107, U=0, E=0)  
2023/10/21 22:16:16 - Group by 2 0 - Finished processing (I=0, O=0, R=147107, W=18, U=0, E=0)  
2023/10/21 22:16:16 - Join rows 0 - Finished processing (I=0, O=0, R=36, W=324, U=0, E=0)  
2023/10/21 22:16:16 - Calculator 0 - Finished processing (I=0, O=0, R=324, W=324, U=0, E=0)  
2023/10/21 22:16:16 - Sort rows 3 0 - Finished processing (I=0, O=0, R=324, W=324, U=0, E=0)  
2023/10/21 22:16:16 - Group by 3 0 - Finished processing (I=0, O=0, R=324, W=18, U=0, E=0)  
2023/10/21 22:16:16 - Text file output 0 - Finished processing (I=0, O=19, R=18, W=18, U=0, E=0)  
2023/10/21 22:16:16 - Spoon - The transformation has finished!!

# Preview of Select Values – Dataset A

Activities SWT

Sun 00:12

Spoon - EX3 (changed)

File Edit View Action Tools Help

View Design

Welcome! EX3

Errors 0

100%

Examine preview data

Rows of step: Select values (1000 rows)

District\_A

Row	District
1	Beja
2	Évora
3	Bragança
4	Braga
5	Leiria
6	Aveiro
7	Braga
8	Braga
9	Bragança
10	Guarda
11	Viseu
12	Porto
13	Santarém
14	Guarda
15	Vila Real
16	Bragança
17	Castelo Branco
18	Guarda
19	Guarda
20	Évora
21	Viseu
22	Évora
23	Viseu
24	Vila Real
25	Guarda
26	Viana do Castelo
27	Porto

Close Stop Get more rows

Connect

# Configuration of Select Values – Dataset B

Activities SWT

Sat 22:28 Spoon - EX3

File Edit View Action Tools Help Connect

View Design Search Transformations EX3 Run configurations Database connections Steps Hops Partition schemas Slave server Kettle cluster schemas VFS Connections

EX3

CSV file input Select values Sort rows Group by

CSV file input 2 Select values 2 Step name Select values 2

Select & Alter Remove Meta-data

Fields :

Fieldname	Rename to	Length	Precision
1 District	District_B		

Get fields to select Edit Mapping

Include unspecified fields, ordered by name

OK Cancel Help

Execution Results

Logging

2023/10/21 22:16:14 - Select values 2.0 - Finished processing (I=0, O=0, R=147107, W=147107, U=0, E=0)  
2023/10/21 22:16:15 - Group by.0 - Linenr 50000  
2023/10/21 22:16:15 - Group by.0 - Linenr 100000  
2023/10/21 22:16:15 - Sort rows.0 - Finished processing (I=0, O=0, R=120943, W=120943, U=0, E=0)  
2023/10/21 22:16:15 - Group by.0 - Finished processing (I=0, O=0, R=120943, W=18, U=0, E=0)  
2023/10/21 22:16:15 - Group by 2.0 - Linenr 50000  
2023/10/21 22:16:16 - Group by 2.0 - Linenr 100000  
2023/10/21 22:16:16 - Sort rows 2.0 - Finished processing (I=0, O=0, R=147107, W=147107, U=0, E=0)  
2023/10/21 22:16:16 - Group by 2.0 - Finished processing (I=0, O=0, R=147107, W=18, U=0, E=0)  
2023/10/21 22:16:16 - Join rows.0 - Finished processing (I=0, O=0, R=36, W=324, U=0, E=0)  
2023/10/21 22:16:16 - Calculator.0 - Finished processing (I=0, O=0, R=324, W=324, U=0, E=0)  
2023/10/21 22:16:16 - Sort rows 3.0 - Finished processing (I=0, O=0, R=324, W=324, U=0, E=0)  
2023/10/21 22:16:16 - Group by 3.0 - Finished processing (I=0, O=0, R=324, W=18, U=0, E=0)  
2023/10/21 22:16:16 - Text file output.0 - Finished processing (I=0, O=19, R=18, W=18, U=0, E=0)  
2023/10/21 22:16:16 - Spoon - The transformation has finished!!

# Preview of Select Values – Dataset B

Activities    SWT    Sun 00:12    Spoon - EX3 (changed)

File Edit View Action Tools Help

View Design

Search

Trax Selections

Rows of step: Select values 2 (1000 rows)

District\_B

1 PORTO  
2 PORTO  
3 SANTAREM  
4 SETUBAL  
5 VISEU  
6 COIMBRA  
7 EVORA  
8 GUARDA  
9 GUARDA  
10 GUARDA  
11 LISBOA  
12 PORTO  
13 VIANA DO CASTELO  
14 VISEU  
15 BRAGA  
16 CASTELO BRANCO  
17 COIMBRA  
18 LISBOA  
19 LISBOA  
20 SANTAREM  
21 VIANA DO CASTELO  
22 VILA REAL  
23 VILA REAL  
24 VISEU  
25 BEJA  
26 LISBOA  
27 PORTO

Examine preview data

Close Stop Get more rows

Connect

The screenshot shows a Java application window titled "Spoon - EX3 (changed)". The main area is a preview window titled "Examine preview data" which displays a list of 1000 rows from a dataset. The columns are labeled "Row" and "Value". The first few rows show values such as "1 PORTO", "2 PORTO", "3 SANTAREM", etc. The preview window has a toolbar at the top with icons for zooming, sorting, and filtering. Below the preview are three buttons: "Close", "Stop", and "Get more rows". The left sidebar contains a tree view of the dataset structure, showing a single step named "Rows of step: Select values 2 (1000 rows)" under a folder named "District\_B". The Spoon interface has a dark theme with orange and white highlights. The top bar also shows the date and time as "Sun 00:12".

# Configuration of Sort Rows – Dataset A

Activities SWT

Sat 22:28 Spoon - EX3

File Edit View Action Tools Help

View Design Search Transformations

EX3

CSV file input → Select values → Sort rows → Group by

CSV file input 2 → Select value

**Sort rows**

Step name: Sort rows

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

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 District_A	Y	N	N	0	N

Execution Results

Logging Execution

2023/10/21 22:16:14 - Select  
2023/10/21 22:16:15 - Group  
2023/10/21 22:16:15 - Group  
2023/10/21 22:16:15 - Sort r  
2023/10/21 22:16:15 - Group  
2023/10/21 22:16:16 - Group  
2023/10/21 22:16:16 - Sort r  
2023/10/21 22:16:16 - Group  
2023/10/21 22:16:16 - Join r  
2023/10/21 22:16:16 - Calculu  
2023/10/21 22:16:16 - Sort rows 3.0 - Finished processing (l=0, O=0, R=324, W=324, U=0, E=0)  
2023/10/21 22:16:16 - Group by 3.0 - Finished processing (l=0, O=0, R=324, W=18, U=0, E=0)  
2023/10/21 22:16:16 - Text file output.0 - Finished processing (l=0, O=19, R=18, W=18, U=0, E=0)  
2023/10/21 22:16:16 - Spoon - The transformation has finished!!

OK Cancel Get Fields

# Preview of Sort Rows – Dataset A

Activities SWT

Sun 00:19  
Spoon - EX3 (changed)

File Edit View Action Tools Help

View Design

Search

Transformations

Rows of step: Sort rows (1000 rows)

District\_A

1 Aveiro  
2 Aveiro  
3 Aveiro  
4 Aveiro  
5 Aveiro  
6 Aveiro  
7 Aveiro  
8 Aveiro  
9 Aveiro  
10 Aveiro  
11 Aveiro  
12 Aveiro  
13 Aveiro  
14 Aveiro  
15 Aveiro  
16 Aveiro  
17 Aveiro  
18 Aveiro  
19 Aveiro  
20 Aveiro  
21 Aveiro  
22 Aveiro  
23 Aveiro  
24 Aveiro  
25 Aveiro  
26 Aveiro  
27 Aveiro

Examine preview data

Close Stop Get more rows

2023/10/22 00:19:31 - Sort rows.0 - Linear 100000  
2023/10/22 00:19:31 - CSV file input.0 - Finished processing (I=120944, O=0, R=0, W=120943, U=0, E=0)  
2023/10/22 00:19:31 - CSV file input 2.0 - Finished processing (I=147108, O=0, R=0, W=147107, U=0, E=0)  
2023/10/22 00:19:31 - Select values.0 - Finished processing (I=0, O=0, R=120943, W=120943, U=0, E=0)  
2023/10/22 00:19:31 - Select values 2.0 - Finished processing (I=0, O=0, R=147107, W=147107, U=0, E=0)

# Configuration of Sort Rows – Dataset B

Sat 22:29  
Spoon - EX3

Activities SWT

File Edit View Action Tools Help

View Design Search Transformations EX3 Run configurations Database connections Steps Hops Partition schemas Slave server Kettle cluster schemas VFS Connections

CSV file input Select values Sort rows Group by

CSV file input 2 Select value

Sort rows

Step name Sort rows 2

Sort directory %%java.io.tmpdir%%

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 District_B	Y	N	N	0	N

Execution Results

Logging Execution

2023/10/21 22:16:14 - Select 2023/10/21 22:16:15 - Group 2023/10/21 22:16:15 - Group 2023/10/21 22:16:15 - Sort r 2023/10/21 22:16:15 - Group 2023/10/21 22:16:15 - Group 2023/10/21 22:16:16 - Group 2023/10/21 22:16:16 - Sort r 2023/10/21 22:16:16 - Group 2023/10/21 22:16:16 - Join r 2023/10/21 22:16:16 - Calculu 2023/10/21 22:16:16 - Sort rows 3.0 - Finished processing (l=0, O=0, R=324, W=324, U=0, E=0) 2023/10/21 22:16:16 - Group by 3.0 - Finished processing (l=0, O=0, R=324, W=18, U=0, E=0) 2023/10/21 22:16:16 - Text file output.0 - Finished processing (l=0, O=19, R=18, W=18, U=0, E=0) 2023/10/21 22:16:16 - Spoon - The transformation has finished!!

OK Cancel Get Fields

```
2023/10/21 22:16:14 - Select
2023/10/21 22:16:15 - Group
2023/10/21 22:16:15 - Group
2023/10/21 22:16:15 - Sort r
2023/10/21 22:16:15 - Group
2023/10/21 22:16:15 - Group
2023/10/21 22:16:16 - Group
2023/10/21 22:16:16 - Sort r
2023/10/21 22:16:16 - Group
2023/10/21 22:16:16 - Join r
2023/10/21 22:16:16 - Calculu
2023/10/21 22:16:16 - Sort rows 3.0 - Finished processing (l=0, O=0, R=324, W=324, U=0, E=0)
2023/10/21 22:16:16 - Group by 3.0 - Finished processing (l=0, O=0, R=324, W=18, U=0, E=0)
2023/10/21 22:16:16 - Text file output.0 - Finished processing (l=0, O=19, R=18, W=18, U=0, E=0)
2023/10/21 22:16:16 - Spoon - The transformation has finished!!
```

# Preview of Sort Rows – Dataset B

Activities SWT

Sun 00:17  
Spoon - EX3 (changed)

File Edit View Action Tools Help

View Design

Search

Transformations

Rows of step: Sort rows 2 (1000 rows)

District\_B

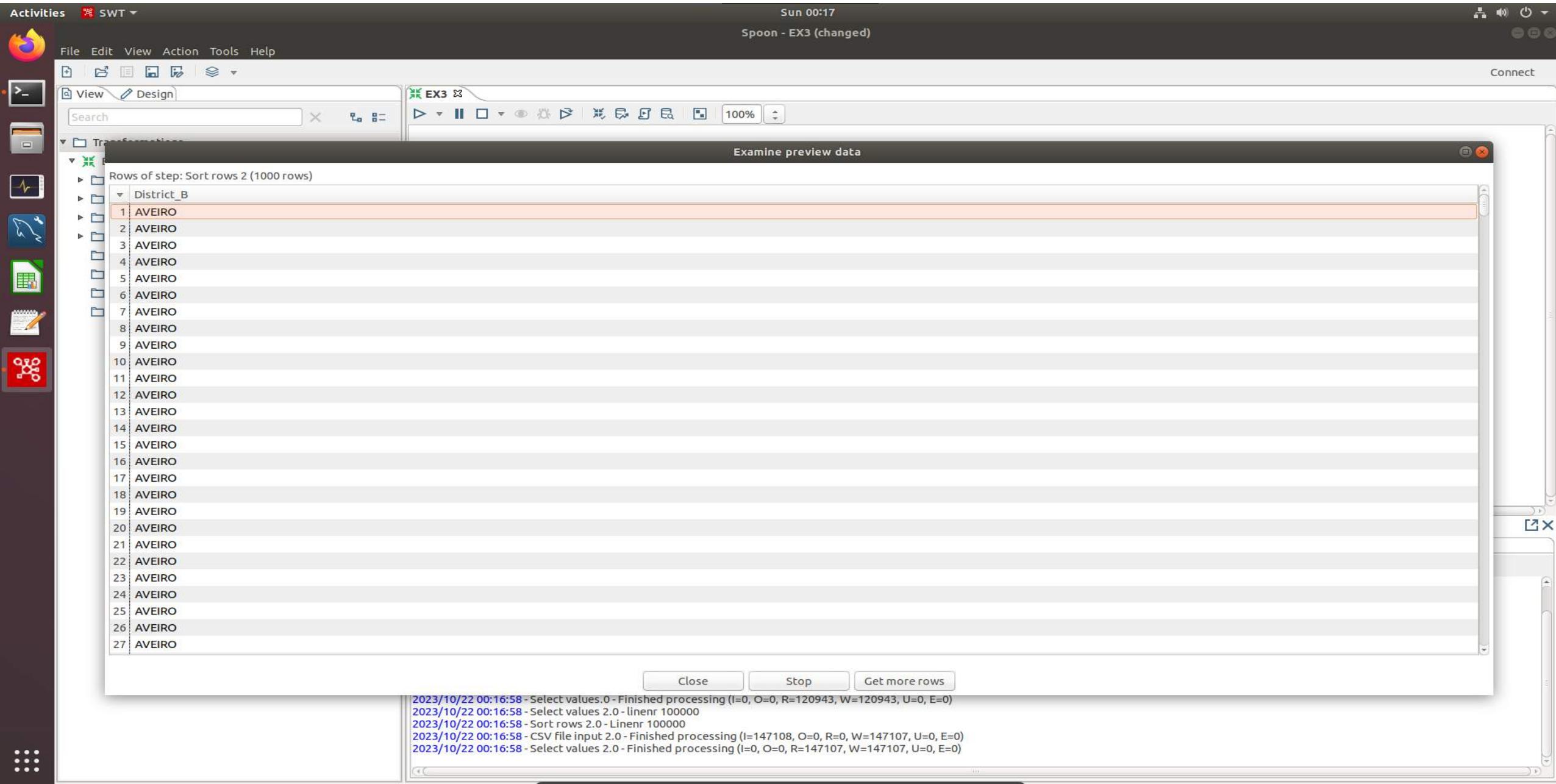
1 AVEIRO  
2 AVEIRO  
3 AVEIRO  
4 AVEIRO  
5 AVEIRO  
6 AVEIRO  
7 AVEIRO  
8 AVEIRO  
9 AVEIRO  
10 AVEIRO  
11 AVEIRO  
12 AVEIRO  
13 AVEIRO  
14 AVEIRO  
15 AVEIRO  
16 AVEIRO  
17 AVEIRO  
18 AVEIRO  
19 AVEIRO  
20 AVEIRO  
21 AVEIRO  
22 AVEIRO  
23 AVEIRO  
24 AVEIRO  
25 AVEIRO  
26 AVEIRO  
27 AVEIRO

Examine preview data

Close Stop Get more rows

2023/10/22 00:16:58 - Select values.0 - Finished processing (I=0, O=0, R=120943, W=120943, U=0, E=0)  
2023/10/22 00:16:58 - Select values 2.0 - linenr 100000  
2023/10/22 00:16:58 - Sort rows 2.0 - Linenr 100000  
2023/10/22 00:16:58 - CSV File input 2.0 - Finished processing (I=147108, O=0, R=0, W=147107, U=0, E=0)  
2023/10/22 00:16:58 - Select values 2.0 - Finished processing (I=0, O=0, R=147107, W=147107, U=0, E=0)

Connect



# Configuration of Group By – Dataset A

Activities SWT

Sat 22:29  
Spoon - EX3

File Edit View Action Tools Help

View Design Search

Transformations

- EX3
  - Run configurations
  - Database connections
  - Steps
  - Hops
  - Partition schemas
  - Slave server
  - Kettle cluster schemas
  - VFS Connections

Connect

**Group by**

Step name: **Group by**

Include all rows?

Temporary files directory: `%%java.io.tmpdir%%`

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 District_A`

Aggregates :

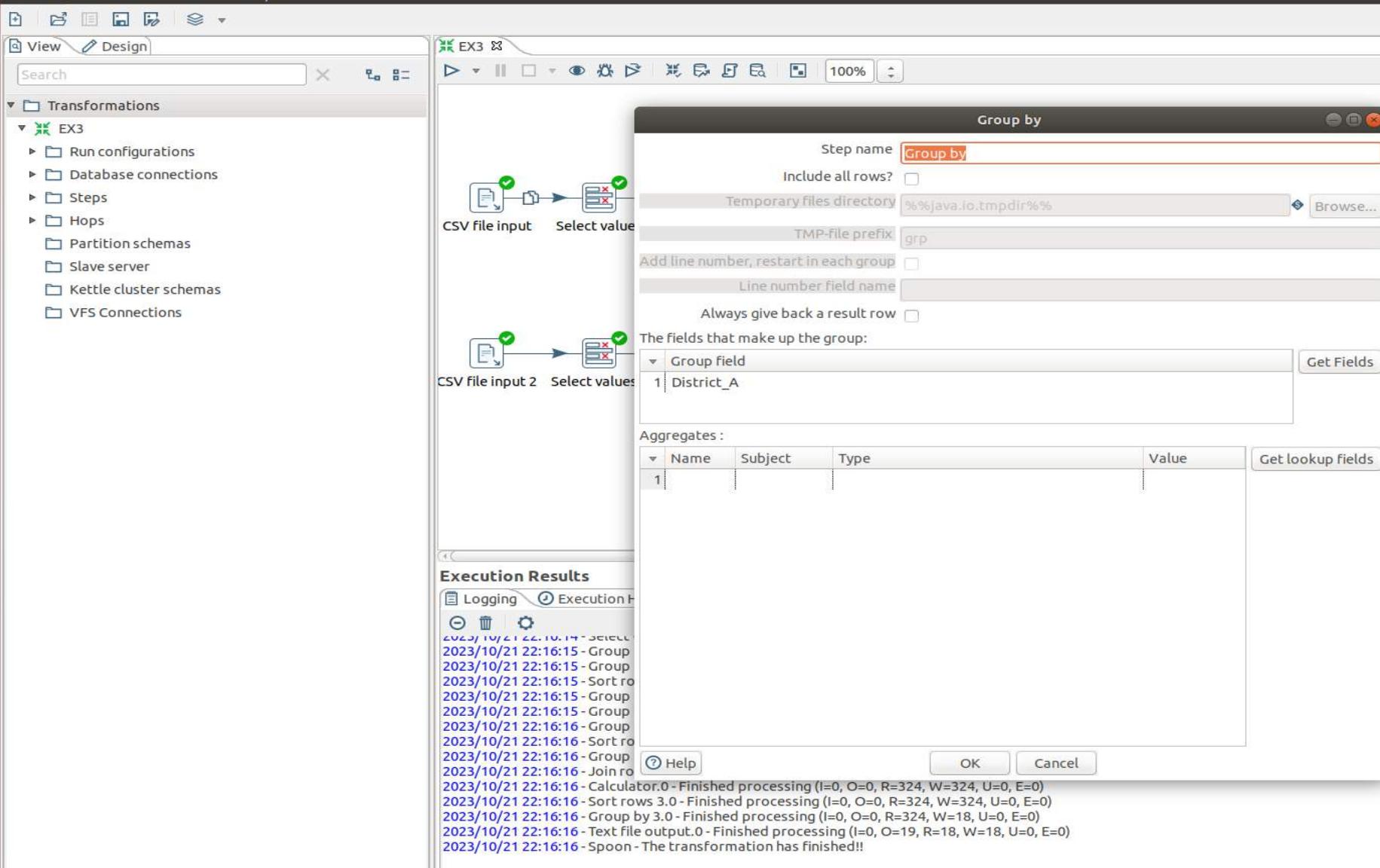
Name	Subject	Type	Value	Get lookup fields
1				

Execution Results

Logging Execution History

```
2023/10/21 22:16:14 - SELECT
2023/10/21 22:16:15 - Group
2023/10/21 22:16:15 - Group
2023/10/21 22:16:15 - Sort ro
2023/10/21 22:16:15 - Group
2023/10/21 22:16:15 - Group
2023/10/21 22:16:16 - Group
2023/10/21 22:16:16 - Sort ro
2023/10/21 22:16:16 - Group
2023/10/21 22:16:16 - Join ro
2023/10/21 22:16:16 - Calculator.0 - Finished processing (I=0, O=0, R=324, W=324, U=0, E=0)
2023/10/21 22:16:16 - Sort rows 3.0 - Finished processing (I=0, O=0, R=324, W=324, U=0, E=0)
2023/10/21 22:16:16 - Group by 3.0 - Finished processing (I=0, O=0, R=324, W=18, U=0, E=0)
2023/10/21 22:16:16 - Text file output.0 - Finished processing (I=0, O=19, R=18, W=18, U=0, E=0)
2023/10/21 22:16:16 - Spoon - The transformation has finished!!
```

OK Cancel Help



# Preview of Group By – Dataset A

Activities SWT

Sun 00:20  
Spoon - EX3 (changed)

File Edit View Action Tools Help

View Design

Search

Connect

EX3

Examine preview data

Rows of step: Group by (18 rows)

District\_A

- 1 Aveiro
- 2 Beja
- 3 Braga
- 4 Bragança
- 5 Castelo Branco
- 6 Coimbra
- 7 Faro
- 8 Guarda
- 9 Leiria
- 10 Lisboa
- 11 Portalegre
- 12 Porto
- 13 Santarém
- 14 Setúbal
- 15 Viana do Castelo
- 16 Vila Real
- 17 Viseu
- 18 Évora

Close

```
2023/10/22 00:20:02 - Calculator.0 - Finished processing (I=0, O=0, R=324, W=324, U=0, E=0)
2023/10/22 00:20:02 - Sort rows 3.0 - Finished processing (I=0, O=0, R=324, W=324, U=0, E=0)
2023/10/22 00:20:02 - Group by 3.0 - Finished processing (I=0, O=0, R=324, W=18, U=0, E=0)
2023/10/22 00:20:02 - Text File output.0 - Finished processing (I=0, O=19, R=18, W=18, U=0, E=0)
2023/10/22 00:20:02 - Spoon - The transformation has finished!!
```

# Configuration of Group By – Dataset B

Sat 22:29  
Spoon - EX3

Activities SWT Connect

File Edit View Action Tools Help

View Design Search Transformations EX3 Run configurations Database connections Steps Hops Partition schemas Slave server Kettle cluster schemas VFS Connections

CSV file input Select value CSV file input 2 Select values

Group by Step name Group by 2

Include all rows?  Temporary files directory %%java.io.tmpdir%%  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 District\_B

Aggregates :

Name	Subject	Type	Value	<input type="Get lookup fields"/>
1				

Execution Results

Logging Execution History

2023/10/21 22:16:14 - SELECT

2023/10/21 22:16:15 - Group

2023/10/21 22:16:15 - Group

2023/10/21 22:16:15 - Sort ro

2023/10/21 22:16:15 - Group

2023/10/21 22:16:15 - Group

2023/10/21 22:16:16 - Group

2023/10/21 22:16:16 - Sort ro

2023/10/21 22:16:16 - Group

2023/10/21 22:16:16 - Join ro

2023/10/21 22:16:16 - Calculator.0 - Finished processing (I=0, O=0, R=324, W=324, U=0, E=0)

2023/10/21 22:16:16 - Sort rows 3.0 - Finished processing (I=0, O=0, R=324, W=324, U=0, E=0)

2023/10/21 22:16:16 - Group by 3.0 - Finished processing (I=0, O=0, R=324, W=18, U=0, E=0)

2023/10/21 22:16:16 - Text file output.0 - Finished processing (I=0, O=19, R=18, W=18, U=0, E=0)

2023/10/21 22:16:16 - Spoon - The transformation has finished!!

# Preview of Group By – Dataset B

Activities SWT

Sun 00:20  
Spoon - EX3 (changed)

File Edit View Action Tools Help

View Design

Search

Connect

Tranformations

Rows of step: Group by 2 (18 rows)

District\_B

1 AVEIRO

2 BEJA

3 BRAGA

4 BRAGANCA

5 CASTELO BRANCO

6 COIMBRA

7 EVORA

8 FARO

9 GUARDA

10 LEIRIA

11 LISBOA

12 PORTALEGRE

13 PORTO

14 SANTAREM

15 SETUBAL

16 VIANA DO CASTELO

17 VILA REAL

18 VISEU

Examine preview data

Close

2023/10/22 00:20:26 - Calculator.0 - Finished processing (I=0, O=0, R=324, W=324, U=0, E=0)  
2023/10/22 00:20:26 - Sort rows 3.0 - Finished processing (I=0, O=0, R=324, W=324, U=0, E=0)  
2023/10/22 00:20:26 - Group by 3.0 - Finished processing (I=0, O=0, R=324, W=18, U=0, E=0)  
2023/10/22 00:20:26 - Text File output.0 - Finished processing (I=0, O=19, R=18, W=18, U=0, E=0)  
2023/10/22 00:20:26 - Spoon - The transformation has finished!!

# Configuration of Join Rows

Activities SWT

Sat 22:30  
Spoon - EX3

File Edit View Action Tools Help

View Design Search

Transformations

- EX3
  - Run configurations
  - Database connections
  - Steps
  - Hops
  - Partition schemas
  - Slave server
  - Kettle cluster schemas
  - VFS Connections

Connect

**EX3**

CSV file input → Select values → Sort rows → Group by

CSV file input 2 → Select values 2 → Sort rows 2 → Group by 2

**Join rows (cartesian product)**

Step name: **Join rows**

Temp directory: `%%java.io.tmpdir%%`

TMP-file prefix: `out`

Max. cache size (in rows): `500`

Main step to read from:

The condition:

`<field> = <value>`

**Execution Results**

Logging Execution History Step Metrics Performance

2023/10/21 22:16:14 - SELECT values 2.0 - Finished processing (I=0, O=0, R=147107, W=147107, U=0, E=0)  
2023/10/21 22:16:15 - Group by 0 - Linenr 50000  
2023/10/21 22:16:15 - Group by 0 - Linenr 100000  
2023/10/21 22:16:15 - Sort rows 0 - Finished processing (I=0, O=0, R=120943, W=120943, U=0, E=0)  
2023/10/21 22:16:15 - Group by 0 - Finished processing (I=0, O=0, R=120943, W=18, U=0, E=0)  
2023/10/21 22:16:15 - Group by 2.0 - Linenr 5000  
2023/10/21 22:16:16 - Group by 2.0 - Linenr 100000  
2023/10/21 22:16:16 - Sort rows 2.0 - Finished processing (I=0, O=0, R=147107, W=147107, U=0, E=0)  
2023/10/21 22:16:16 - Group by 2.0 - Finished processing (I=0, O=0, R=147107, W=18, U=0, E=0)  
2023/10/21 22:16:16 - Join rows 0 - Finished processing (I=0, O=0, R=36, W=324, U=0, E=0)  
2023/10/21 22:16:16 - Calculator 0 - Finished processing (I=0, O=0, R=324, W=324, U=0, E=0)  
2023/10/21 22:16:16 - Sort rows 3.0 - Finished processing (I=0, O=0, R=324, W=324, U=0, E=0)  
2023/10/21 22:16:16 - Group by 3.0 - Finished processing (I=0, O=0, R=324, W=18, U=0, E=0)  
2023/10/21 22:16:16 - Text file output 0 - Finished processing (I=0, O=19, R=18, W=18, U=0, E=0)  
2023/10/21 22:16:16 - Spoon - The transformation has finished!!

# Preview of Join Rows

Activities Sun 00:20

Spoon - EX3 (changed)

File Edit View Action Tools Help

View Design

Search

Connect

Examine preview data

Rows of step: Join rows (324 rows)

	District_B	District_A
1	AVEIRO	Aveiro
2	AVEIRO	Beja
3	AVEIRO	Braga
4	AVEIRO	Bragança
5	AVEIRO	Castelo Branco
6	AVEIRO	Coimbra
7	AVEIRO	Faro
8	AVEIRO	Guarda
9	AVEIRO	Leiria
10	AVEIRO	Lisboa
11	AVEIRO	Portalegre
12	AVEIRO	Porto
13	AVEIRO	Santarém
14	AVEIRO	Setúbal
15	AVEIRO	Viana do Castelo
16	AVEIRO	Vila Real
17	AVEIRO	Viseu
18	AVEIRO	Évora
19	BEJA	Aveiro
20	BEJA	Beja
21	BEJA	Braga
22	BEJA	Bragança
23	BEJA	Castelo Branco
24	BEJA	Coimbra
25	BEJA	Faro
26	BEJA	Guarda
27	BEJA	Leiria

Close

2023/10/22 00:20:53 - Calculator.0 - Finished processing (I=0, O=0, R=324, W=324, U=0, E=0)  
2023/10/22 00:20:53 - Sort rows 3.0 - Finished processing (I=0, O=0, R=324, W=324, U=0, E=0)  
2023/10/22 00:20:53 - Group by 3.0 - Finished processing (I=0, O=0, R=324, W=18, U=0, E=0)  
2023/10/22 00:20:53 - Text File output.0 - Finished processing (I=0, O=19, R=18, W=18, U=0, E=0)  
2023/10/22 00:20:53 - Spoon - The transformation has finished!!

# Configuration of Calculator

Activities SWT Sat 22:35 Spoon - EX3

File Edit View Action Tools Help

View Design Search Transformations EX3

CSV file input Select values Sort rows Group by Join rows Calculator Sort rows 3 Group by 3 Text file output

Calculator

Step name

Throw an error on non existing files

Fields:

	New field	Calculation	Field A	Field B	Field C	Value type	Length	Precision	Remove	Conversion mask	Decimals
1	measure	NeedlemanWunsch distance between String A and String B	District_A	District_B		Number		N			

2023/10/21 22:16:16 - Group by 2.0 - Finished processing (I=0, O=0, R=147107, W=18, U=0, E=0)  
2023/10/21 22:16:16 - Join rows.0 - Finished processing (I=0, O=0, R=36, W=324, U=0, E=0)  
2023/10/21 22:16:16 - Calculator.0 - Finished processing (I=0, O=0, R=324, W=324, U=0, E=0)  
2023/10/21 22:16:16 - Sort rows 3.0 - Finished processing (I=0, O=0, R=324, W=324, U=0, E=0)  
2023/10/21 22:16:16 - Group by 3.0 - Finished processing (I=0, O=0, R=324, W=18, U=0, E=0)  
2023/10/21 22:16:16 - Text file output.0 - Finished processing (I=0, O=19, R=18, W=18, U=0, E=0)  
2023/10/21 22:16:16 - Spoon - The transformation has finished!!

# Preview of Calculator

Activities SWT

Sun 00:21  
Spoon - EX3 (changed)

File Edit View Action Tools Help

View Design Connect

Search

Transformations

Rows of step: Calculator (324 rows)

	District_B	District_A	measure
1	AVEIRO	Aveiro	0.0
2	AVEIRO	Beja	-5.0
3	AVEIRO	Braga	-6.0
4	AVEIRO	Bragança	-7.0
5	AVEIRO	Castelo Branco	-10.0
6	AVEIRO	Coimbra	-6.0
7	AVEIRO	Faro	-4.0
8	AVEIRO	Guarda	-6.0
9	AVEIRO	Leiria	-4.0
10	AVEIRO	Lisboa	-6.0
11	AVEIRO	Portalegre	-7.0
12	AVEIRO	Porto	-5.0
13	AVEIRO	Santarém	-6.0
14	AVEIRO	Setúbal	-7.0
15	AVEIRO	Viana do Castelo	-14.0
16	AVEIRO	Vila Real	-8.0
17	AVEIRO	Viseu	-5.0
18	AVEIRO	Évora	-4.0
19	BEJA	Aveiro	-5.0
20	BEJA	Beja	0.0
21	BEJA	Braga	-3.0
22	BEJA	Bragança	-6.0
23	BEJA	Castelo Branco	-12.0
24	BEJA	Coimbra	-6.0
25	BEJA	Faro	-4.0
26	BEJA	Guarda	-5.0
27	BEJA	Leiria	-4.0

Examine preview data

Close

2023/10/22 00:21:12 - Calculator.0 - Finished processing (I=0, O=0, R=324, W=324, U=0, E=0)  
2023/10/22 00:21:12 - Sort rows 3.0 - Finished processing (I=0, O=0, R=324, W=324, U=0, E=0)  
2023/10/22 00:21:12 - Group by 3.0 - Finished processing (I=0, O=0, R=324, W=18, U=0, E=0)  
2023/10/22 00:21:12 - Text File output.0 - Finished processing (I=0, O=19, R=18, W=18, U=0, E=0)  
2023/10/22 00:21:12 - Spoon - The transformation has finished!!

# Configuration of Sort Rows

Sat 22:36  
Spoon - EX3

Activities SWT

File Edit View Action Tools Help

View Design Search

Transformations

- EX3
  - Run configurations
  - Database connections
  - Steps
  - Hops
  - Partition schemas
  - Slave server
  - Kettle cluster schemas
  - VFS Connections

Connect

EX3

CSV file input → Select values → Sort rows → Group by

CSV file input 2 → Select value

Sort rows configuration dialog:

- Step name: Sort rows 3
- Sort directory: %&java.io.tmpdir%
- 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 District_A	Y	Y	N	0	N
2 measure	N	N	N	0	N

Execution Results

```
2023/10/21 22:16:14 - SELECT
2023/10/21 22:16:15 - Group
2023/10/21 22:16:15 - Group
2023/10/21 22:16:15 - Sort rows
2023/10/21 22:16:15 - Group
2023/10/21 22:16:16 - Group
2023/10/21 22:16:16 - Sort rows
2023/10/21 22:16:16 - Group
2023/10/21 22:16:16 - Join rows
2023/10/21 22:16:16 - Calculations
2023/10/21 22:16:16 - Sort rows 3.0 - Finished processing (I=0, O=0, R=324, W=324, U=0, E=0)
2023/10/21 22:16:16 - Group by 3.0 - Finished processing (I=0, O=0, R=324, W=18, U=0, E=0)
2023/10/21 22:16:16 - Text file output.0 - Finished processing (I=0, O=19, R=18, W=18, U=0, E=0)
2023/10/21 22:16:16 - Spoon - The transformation has finished!!
```

# Preview of Sort Rows

Activities SWT

Sun 00:21  
Spoon - EX3 (changed)

File Edit View Action Tools Help

View Design

Transformations

Rows of step: Sort rows 3 (324 rows)

District\_B District\_A measure

	District_B	District_A	measure
1	AVEIRO	Aveiro	0.0
2	EVORA	Aveiro	-4.0
3	FARO	Aveiro	-4.0
4	LEIRIA	Aveiro	-4.0
5	BEJA	Aveiro	-5.0
6	PORTE	Aveiro	-5.0
7	VISEU	Aveiro	-5.0
8	BRAGA	Aveiro	-6.0
9	COIMBRA	Aveiro	-6.0
10	GUARDA	Aveiro	-6.0
11	LISBOA	Aveiro	-6.0
12	SANTAREM	Aveiro	-6.0
13	BRAGANCA	Aveiro	-7.0
14	PORALEGRE	Aveiro	-7.0
15	SETUBAL	Aveiro	-7.0
16	VILA REAL	Aveiro	-8.0
17	CASTELO BRANCO	Aveiro	-10.0
18	VIANA DO CASTELO	Aveiro	-14.0
19	BEJA	Beja	0.0
20	BRAGA	Beja	-3.0
21	EVORA	Beja	-4.0
22	FARO	Beja	-4.0
23	LEIRIA	Beja	-4.0
24	AVEIRO	Beja	-5.0
25	GUARDA	Beja	-5.0
26	LISBOA	Beja	-5.0
27	PORTE	Beja	-5.0

Examine preview data

Close

2023/10/22 00:21:36 - Calculator.0 - Finished processing (I=0, O=0, R=324, W=324, U=0, E=0)  
2023/10/22 00:21:36 - Sort rows 3.0 - Finished processing (I=0, O=0, R=324, W=324, U=0, E=0)  
2023/10/22 00:21:36 - Group by 3.0 - Finished processing (I=0, O=0, R=324, W=18, U=0, E=0)  
2023/10/22 00:21:36 - Text File output.0 - Finished processing (I=0, O=19, R=18, W=18, U=0, E=0)  
2023/10/22 00:21:36 - Spoon - The transformation has finished!!

# Configuration of Group by

Activities    SWT    Sat 22:37    Spoon - EX3

File Edit View Action Tools Help

View Design Search Transformations EX3 Run configurations Database connections Steps Hops Partition schemas Slave server Kettle cluster schemas VFS Connections

CSV file input Select value CSV file input 2 Select values

Group by Step name Group by 3

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 Get Fields

District\_A

Aggregates :

Name	Subject	Type	Value	Get lookup fields
District_B	District_B	First value		

Execution Results

Logging Execution History

2023/10/21 22:16:14 - SELECT  
2023/10/21 22:16:15 - Group  
2023/10/21 22:16:15 - Group  
2023/10/21 22:16:15 - Sort ro  
2023/10/21 22:16:15 - Group  
2023/10/21 22:16:15 - Group  
2023/10/21 22:16:16 - Group  
2023/10/21 22:16:16 - Sort ro  
2023/10/21 22:16:16 - Group  
2023/10/21 22:16:16 - Join ro  
2023/10/21 22:16:16 - Calculator.0 - Finished processing (I=0, O=0, R=324, W=324, U=0, E=0)  
2023/10/21 22:16:16 - Sort rows 3.0 - Finished processing (I=0, O=0, R=324, W=324, U=0, E=0)  
2023/10/21 22:16:16 - Group by 3.0 - Finished processing (I=0, O=0, R=324, W=18, U=0, E=0)  
2023/10/21 22:16:16 - Text file output.0 - Finished processing (I=0, O=19, R=18, W=18, U=0, E=0)  
2023/10/21 22:16:16 - Spoon - The transformation has finished!!

OK Cancel

```
2023/10/21 22:16:14 - SELECT
2023/10/21 22:16:15 - Group
2023/10/21 22:16:15 - Group
2023/10/21 22:16:15 - Sort ro
2023/10/21 22:16:15 - Group
2023/10/21 22:16:15 - Group
2023/10/21 22:16:16 - Group
2023/10/21 22:16:16 - Sort ro
2023/10/21 22:16:16 - Group
2023/10/21 22:16:16 - Join ro
2023/10/21 22:16:16 - Calculator.0 - Finished processing (I=0, O=0, R=324, W=324, U=0, E=0)
2023/10/21 22:16:16 - Sort rows 3.0 - Finished processing (I=0, O=0, R=324, W=324, U=0, E=0)
2023/10/21 22:16:16 - Group by 3.0 - Finished processing (I=0, O=0, R=324, W=18, U=0, E=0)
2023/10/21 22:16:16 - Text file output.0 - Finished processing (I=0, O=19, R=18, W=18, U=0, E=0)
2023/10/21 22:16:16 - Spoon - The transformation has finished!!
```

# Preview of Group by

Activities SWT

Sun 00:22  
Spoon - EX3 (changed)

File Edit View Action Tools Help

View Design

EX3

Transformations

Rows of step: Group by 3 (18 rows)

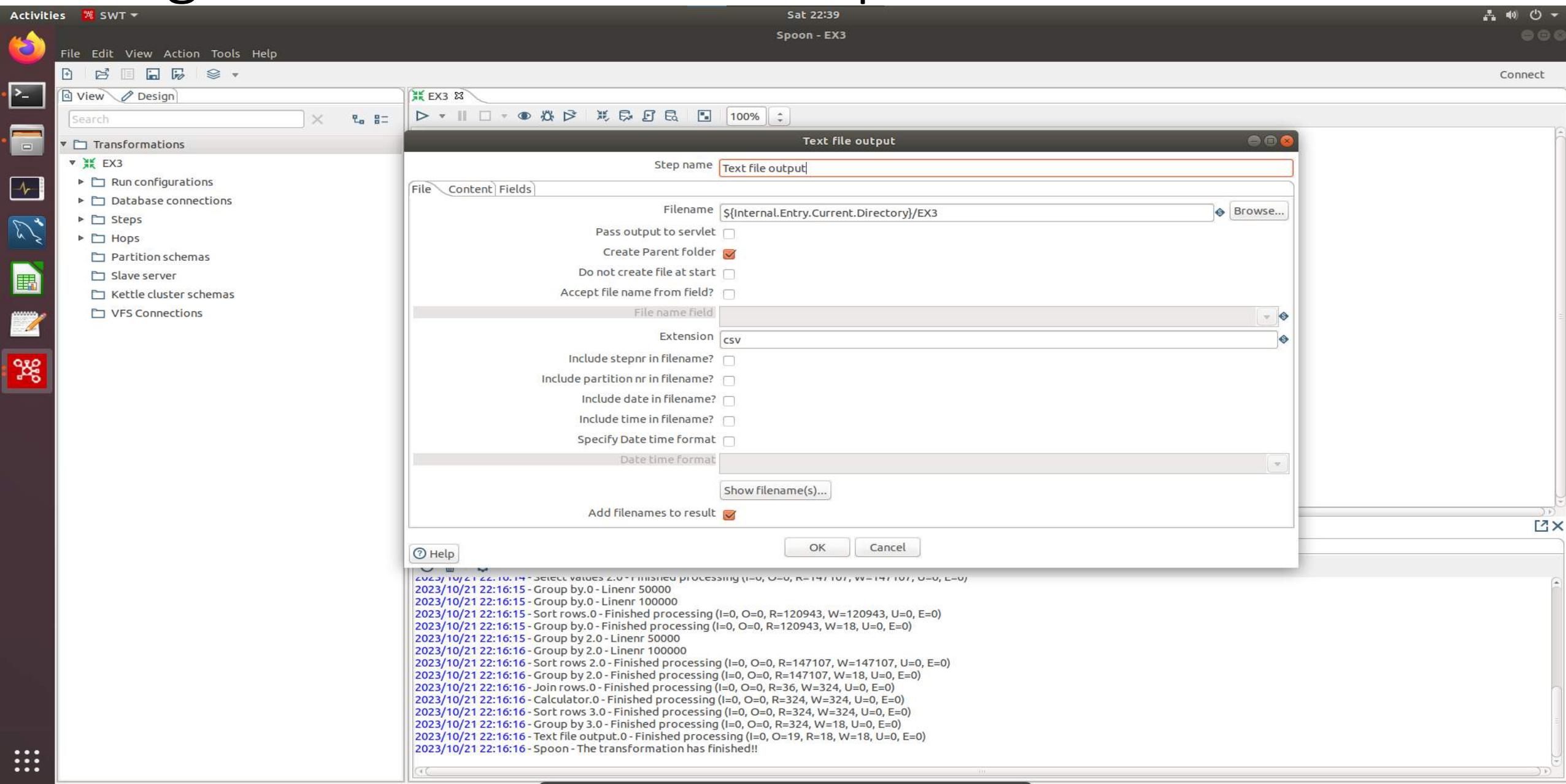
	District_A	District_B
1	Aveiro	AVEIRO
2	Beja	BEJA
3	Braga	BRAGA
4	Bragança	BRAGANCA
5	Castelo Branco	CASTELO BRANCO
6	Coimbra	COIMBRA
7	Faro	FARO
8	Guarda	GUARDA
9	Leiria	LEIRIA
10	Lisboa	LISBOA
11	Portalegre	PORTALEGRE
12	Porto	PORTO
13	Santarém	SANTAREM
14	Setúbal	SETUBAL
15	Viana do Castelo	VIANA DO CASTELO
16	Vila Real	VILA REAL
17	Viseu	VISEU
18	Évora	EVORA

Examine preview data

Close

2023/10/22 00:22:00 - Calculator.0 - Finished processing (I=0, O=0, R=324, W=324, U=0, E=0)  
2023/10/22 00:22:00 - Sort rows 3.0 - Finished processing (I=0, O=0, R=324, W=324, U=0, E=0)  
2023/10/22 00:22:00 - Group by 3.0 - Finished processing (I=0, O=0, R=324, W=18, U=0, E=0)  
2023/10/22 00:22:00 - Text File output.0 - Finished processing (I=0, O=19, R=18, W=18, U=0, E=0)  
2023/10/22 00:22:00 - Spoon - The transformation has finished!!

# Configuration of Test File Output



# Preview of Test File Output

Sun 00:22

Spoon - EX3 (changed)

File Edit View Action Tools Help

View Design

Search

Transformations

Rows of step: Text file output (18 rows)

District\_A District\_B

	District_A	District_B
1	Aveiro	AVEIRO
2	Beja	BEJA
3	Braga	BRAGA
4	Bragança	BRAGANCA
5	Castelo Branco	CASTELO BRANCO
6	Coimbra	COIMBRA
7	Faro	FARO
8	Guarda	GUARDA
9	Leiria	LEIRIA
10	Lisboa	LISBOA
11	Portalegre	PORTALEGRE
12	Porto	PORTO
13	Santarém	SANTAREM
14	Setúbal	SETUBAL
15	Viana do Castelo	VIANA DO CASTELO
16	Vila Real	VILA REAL
17	Viseu	VISEU
18	Évora	EVORA

Examine preview data

Close

2023/10/22 00:22:19 - Calculator.0 - Finished processing (I=0, O=0, R=324, W=324, U=0, E=0)  
2023/10/22 00:22:19 - Sort rows 3.0 - Finished processing (I=0, O=0, R=324, W=324, U=0, E=0)  
2023/10/22 00:22:19 - Group by 3.0 - Finished processing (I=0, O=0, R=324, W=18, U=0, E=0)  
2023/10/22 00:22:19 - Text File output.0 - Finished processing (I=0, O=19, R=18, W=18, U=0, E=0)  
2023/10/22 00:22:19 - Spoon - The transformation has finished!!

# File Output

Activities Text Editor Sun 00:23 EX3.csv ~/Desktop/EX3 Save

District\_A;District\_B  
Aveiro ;AVEIRO  
Beja ;BEJA  
Braga ;BRAGA  
Bragança ;BRAGANCA  
Castelo Branco ;CASTELO BRANCO  
Coimbra ;COIMBRA  
Faro ;FARO  
Guarda ;GUARDA  
Leiria ;LEIRIA  
Lisboa ;LISBOA  
Portalegre ;PORTALEGRE  
Porto ;PORTO  
Santarém ;SANTAREM  
Setúbal ;SETUBAL  
Viana do Castelo;VIANA DO CASTELO  
Vila Real ;VILA REAL  
Viseu ;VISEU  
Évora ;EVORA

# Exercise 4 – Group 50

José Cardoso

Matheus Trindade

Activities Text Editor ▾



Sun 18:34

energy\_dw.sql  
~/Documents/Projeto AID/ExS

Save

energy\_dw.sql

steelwheels\_dw.sql

```
DROP DATABASE IF EXISTS energy_dw;

CREATE DATABASE energy_dw;

USE energy_dw;

CREATE TABLE dim_time (
    YEAR INT,
    MONTH INT,
    MONTH_NAME VARCHAR(255),
    DATE_ID VARCHAR(8),
    PRIMARY KEY (DATE_ID)
);

CREATE TABLE dim_location (
    DISTRICT VARCHAR(255),
    MUNICIPALITY VARCHAR(255),
    PARISH VARCHAR(255),
    PARISH_CODE INT,
    PRIMARY KEY (PARISH_CODE)
);

CREATE TABLE dim_voltage (
    VOLTAGE_LEVEL VARCHAR(255),
    VOLTAGE_ID INT AUTO_INCREMENT PRIMARY KEY
);

CREATE TABLE fact_consumption (
    ACTIVE_ENERGY FLOAT,
    DATE_ID VARCHAR(255),
    PARISH_CODE INT,
    VOLTAGE_ID INT,
    CONSUMPTION_ID INT AUTO_INCREMENT PRIMARY KEY,
    FOREIGN KEY (DATE_ID) REFERENCES dim_time(DATE_ID),
    FOREIGN KEY (PARISH_CODE) REFERENCES dim_location(PARISH_CODE),
    FOREIGN KEY (VOLTAGE_ID) REFERENCES dim_voltage(VOLTAGE_ID)
);
```

SQL ▾ Tab Width: 8 ▾

Ln 6, Col 11

INS

# Exercise 5 – Group 50

José Cardoso

Matheus Trindade

# Dim\_location – Entire transformation & Output table

Activities Terminal Sun 18:37

Spoon - dim\_location

File Edit View Search Terminal Help

mysql> DELETE FROM dim\_location;  
Query OK, 2793 rows affected (0.09 sec)

mysql> select \* from dim\_location LIMIT 40;

DISTRICT	MUNICIPALITY	PARISH	PARISH_CODE
AVEIRO	Águeda	AGUADA DE CIMA	10103
AVEIRO	Águeda	FERMENTELOS	10109
AVEIRO	Águeda	MACINHATA DO VOUGA	10112
AVEIRO	Águeda	VALONGO DO VOUGA	10119
AVEIRO	Águeda	UF AGUEDA E BORRALHA	10121
AVEIRO	Águeda	UF BARRO E AGUDA DE BAIXO	10122
AVEIRO	Águeda	UF BELAZAIMA CAST VOUGA AGADAO	10123
AVEIRO	Águeda	UF RECARDAES E ESPINHEL	10124
AVEIRO	Águeda	UF TRAVASSO E OIS DA RIBEIRA	10125
AVEIRO	Águeda	UF TROFA SEGADAS LAMAS VOUGA	10126
AVEIRO	Águeda	UF PRESTIMO MACIEIRA ALCoba	10127
AVEIRO	Albergaria-a-Velha	ALQUERUBIM	10202
AVEIRO	Albergaria-a-Velha	ANGEJA	10203
AVEIRO	Albergaria-a-Velha	BRANCA	10204
AVEIRO	Albergaria-a-Velha	RIBEIRA DE FRAGUAS	10206
AVEIRO	Albergaria-a-Velha	ALBERGARIA-A-VELHA E VALMAIOR	10209
AVEIRO	Albergaria-a-Velha	SAO JOAO DE LOURE E FROSSOS	10210
AVEIRO	Anadia	VELAS DE CAMINHO	10304
AVEIRO	Anadia	VELAS DE CIMA	10305
AVEIRO	Anadia	MOITA	10307
AVEIRO	Anadia	SANGALHOS	10309
AVEIRO	Anadia	SAO LOURENCO DO BAIRRO	10310
AVEIRO	Anadia	VILA NOVA DE MONSARROS	10312
AVEIRO	Anadia	VILARINHO DO BAIRRO	10313
AVEIRO	Anadia	UF A DA GANDARA BAIRRO E ANCAS	10316
AVEIRO	Anadia	UF ARcos E MOGOFORES	10317
AVEIRO	Anadia	UF TAMENGOS AGUIM OIS BAIRRO	10318
AVEIRO	Arouca	ALVARENGA	10402
AVEIRO	Arouca	CHAVE	10407
AVEIRO	Arouca	ESCARIZ	10409
AVEIRO	Arouca	FERMEDO	10411
AVEIRO	Arouca	MANSORES	10413
AVEIRO	Arouca	MOLDES	10414
AVEIRO	Arouca	ROSSAS	10415
AVEIRO	Arouca	SANTA EULALIA	10416
AVEIRO	Arouca	SAO MIGUEL DO MATO	10417
AVEIRO	Arouca	TROPECO	10418
AVEIRO	Arouca	URRO	10419
AVEIRO	Arouca	VARZEA	10420

a screenshot with the entire contents of the corresponding XML file.

CSV\_input\_location

Filter rows

Select values

Insert / update

Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Preview data

2023/10/22 17:20:56 - Insert / update.0 - Finished processing (I=12347, O=2793, R=12347, W=12346, U=0, E=0)  
2023/10/22 17:20:56 - Select values.0 - Finished processing (I=0, O=0, R=22348, W=22348, U=0, E=0)  
2023/10/22 17:20:56 - Filter rows.0 - Finished processing (I=0, O=0, R=32571, W=32349, U=0, E=0)  
2023/10/22 17:20:56 - CSV\_input\_location.0 - Finished processing (I=42573, O=0, R=0, W=42572, U=0, E=0)  
2023/10/22 17:28:52 - Spoon - Transformation opened.  
2023/10/22 17:28:52 - Spoon - Launching transformation [fact\_consumption]...  
2023/10/22 17:28:52 - Spoon - Started the transformation execution.  
2023/10/22 17:28:53 - Spoon - The transformation has finished!!  
2023/10/22 17:30:32 - Spoon - Spoon  
2023/10/22 17:31:23 - Spoon - Starting job...  
2023/10/22 18:17:08 - Spoon - Job has ended.

# CSV\_input\_location - Configuration

Spoon - dim\_location Sun 18:43

Activities SWT

File Edit View Action Tools Help

View Design

Exercise3 dim\_location /home/aid/Documents/ dim\_voltage 14

10-anos.csv - 1

Transformations dim\_location Run configuration Database connection Steps Hops Partition schema Slave server Kettle cluster connection VFS Connection

CSV file input

Step name: CSV\_input\_location

Filename: \${Internal.Entry.Current.Directory}/3-consumos-faturados-por-municipio.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: UTF-8

Name	Type	Format	Length	Precision	Currency	Decimal
1 District	String		16		\$	.
2 Municipality	String		22		\$	.
3 parish	String		30		\$	.
4 DistrictMunicipalityParishCode	String	#	15	0	\$	.

OK Get Fields Preview Cancel

Pictures

Recent Home Desktop Documents Downloads Music Videos Trash Other Locations

Name Size Modified

Name	Size	Modified
Lab1_Jose_Matheus.png	242.0 kB	21 Sep
Lab2_Jose_Matheus.png	227.9 kB	21 Sep
Lab3_Jose_Matheus.png	425.8 kB	27 Sep
Lab5_Jose_Matheus.png	607.2 kB	3 Oct
Lab6_Final_Jose_Matheus.png	258.2 kB	13 Oct
Lab6_Jose_Matheus.png	418.6 kB	10 Oct
Lab7_Jose_Matheus.png	348.9 kB	13 Oct
Lab8_Jose_Matheus_.png	343.2 kB	Tue
LAB4.PNG	380.5 kB	Fri
Screenshot from 2023-09-19 20-03-23.png	270.6 kB	19 Sep
Screenshot from 2023-09-19 20-04-25.png	269.4 kB	19 Sep
Screenshot from 2023-10-10 23-00-42.png	313.6 kB	10 Oct
Screenshot from 2023-10-15 22-56-25.png	289.3 kB	15 Oct
Screenshot from 2023-10-15 22-56-41.png	275.9 kB	15 Oct
Screenshot from 2023-10-15 22-56-48.png	255.0 kB	15 Oct
Screenshot from 2023-10-15 22-56-52.png	250.6 kB	15 Oct
Screenshot from 2023-10-15 22-56-58.png	249.9 kB	15 Oct
Screenshot from 2023-10-15 22-57-04.png	263.0 kB	15 Oct

"Lab2\_Jose\_Matheus.png" selected (227.9 kB)

# CSV\_input\_location - Preview

Activities Sun 18:43

Spoon - dim\_location

File Edit View Action Tools Help

View Design Exercice? dim\_location /home/bid/Documents/ dim\_voltage ...

Connect Recent Home Pictures

Examine preview data

Rows of step: CSV\_input\_location (100 rows)

	District	Municipality	parish	DistrictMunicipalityParishCode
1	GUARDA	Guarda	ARRIFANA	090706
2	PORTE	Vila do Conde	UF MALTA E CANIDELO	131633
3	SETUBAL	Alcácer do Sal	UF ALCACER SAL E SANTA SUSANA	150107
4	AVEIRO	Anadia	UF ARCOS E MOGOFORES	010317
5	BRAGA	Barcelos	FRAGOSO	030235
6	LISBOA	Sintra	UF SINTRA	111128
7	PORALEGRE	Alter do Chão	CUNHEIRA	120104
8	VIANA DO CASTELO	Valença	BOIVAO	160802
9	AVEIRO	Aveiro	OLIVEIRINHA	010508
10	BRAGANCA	Torre de Moncorvo	UF ADEGANHA E CARDANHA	040918
11	FARO	Albufeira	PADERNE	080103
12	AVEIRO	Arouca	SAO MIGUEL DO MATO	010417
13	GUARDA	Aguiar da Beira	UF SOUTO AGUIAR BEIRA VALVERDE	090116
14	LISBOA	Torres Vedras	UF DOIS PORTOS E RUNA	111324
15	PORTE	Amarante	SALVADOR DO MONTE	130129
16	PORTE	Penafiel	CROCA	131108
17	COIMBRA	Góis	GOIS	060604
18	LISBOA	Alenquer	OLHALVO	110108
19	SANTAREM	Torres Novas	UF BROQUEIRA P Igreja ALCOROCHE	141918
20	VILA REAL	Alijó	VILA CHA	170116
21	AVEIRO	ílhavo	ILHAZO (SAO SALVADOR)	011008
22	BEJA	Aljustrel	ERVIDEL	020102
23	BEJA	Almodôvar	SAO BARNABE	020206
24	BEJA	Beja	BALEIZAO	020502

**Close Stop Get more rows**

Recent Home Pictures

Name Size Modified

PNG Lab1\_Jose\_Matheus.png 242.0 kB 21 Sep

PNG Screenshot from 2023-10-15 22-56-58.png 249.9 kB 15 Oct

PNG Screenshot from 2023-10-15 22-57-04.png 263.0 kB 15 Oct "Lab2\_Jose\_Matheus.png" selected (227.9 kB)

# Filter Rows - Configuration

Activities SWT

Spoon - dim\_location Sun 18:43

File Edit View Action Tools Help

View Design

Transformations

- dim\_location
  - Run configurations
  - Database connections
  - Steps
  - Hops
  - Partition schemas
  - Slave server
  - Kettle cluster schemas
  - VFS Connections

Exercise3 dim\_location /home/aid/Documents/ dim\_voltage »14

CSV\_input\_location → Filter rows → Select values → Insert / update

Filter rows

Step name: Filter rows

Send 'true' data to step:

Send 'false' data to step:

The condition:

DistrictMunicipalityParishCode REGEXP ^[0-9]+\$ (String)

OK Cancel

10-anos.csv - 1 Home Pictures

Recent Home Desktop Documents Downloads Music Pictures Videos Trash Other Locations

Name Size Modified

Name	Size	Modified
Lab1_Jose_Matheus.png	242.0 kB	21 Sep
Lab2_Jose_Matheus.png	222.9 kB	21 Sep
Lab3_Jose_Matheus.png	425.8 kB	27 Sep
Lab5_Jose_Matheus.png	607.2 kB	3 Oct
Lab6_Final_Jose_Matheus.png	258.2 kB	13 Oct
Lab6_Jose_Matheus.png	418.6 kB	10 Oct
Lab7_Jose_Matheus.png	348.9 kB	13 Oct
Lab8_Jose_Matheus_.png	343.2 kB	Tue
LAB4.PNG	380.5 kB	Fri
Screenshot from 2023-09-19 20-03-23.png	270.6 kB	19 Sep
Screenshot from 2023-09-19 20-04-25.png	269.4 kB	19 Sep
Screenshot from 2023-10-10 23-00-42.png	313.6 kB	10 Oct
Screenshot from 2023-10-15 22-56-25.png	289.3 kB	15 Oct
Screenshot from 2023-10-15 22-56-41.png	275.9 kB	15 Oct
Screenshot from 2023-10-15 22-56-48.png	255.0 kB	15 Oct
Screenshot from 2023-10-15 22-56-52.png	250.6 kB	15 Oct
Screenshot from 2023-10-15 22-56-58.png	249.9 kB	15 Oct
Screenshot from 2023-10-15 22-57-04.png	263.0 kB	15 Oct
"Lab2_Jose_Matheus.png"	selected (227.9 kB)	

# Filter Rows - Preview

Activities Sun 18:43

Spoon - dim\_location

File Edit View Action Tools Help

View Design fact order load dw Project1 Exercise3 dim location /home/aid/Documents/ dim voltage fact consumption Job Exercise6 Connect

Rows of step: Filter rows (100 rows)

	District	Municipality	parish	DistrictMunicipalityParishCode
1	GUARDA	Guarda	ARRIFANA	090706
2	PORTO	Vila do Conde	UF MALTA E CANIDELO	131633
3	SETUBAL	Alcácer do Sal	UF ALCACER SAL E SANTA SUSANA	150107
4	AVEIRO	Anadia	UF ARCOS E MOGOFORES	010317
5	BRAGA	Barcelos	FRAGOSO	030235
6	LISBOA	Sintra	UF SINTRA	111128
7	PORALEGRE	Alter do Chão	CUNHEIRA	120104
8	VIANA DO CASTELO	Valença	BOIVAO	160802
9	AVEIRO	Aveiro	OLIVEIRINHA	010508
10	BRAGANCA	Torre de Moncorvo	UF ADEGANHA E CARDANHA	040918
11	FARO	Albufeira	PADERNE	080103
12	AVEIRO	Arouca	SAO MIGUEL DO MATO	010417
13	GUARDA	Aguiar da Beira	UF SOUTO AGUIAR BEIRA VALVERDE	090116
14	LISBOA	Torres Vedras	UF DOIS PORTOS E RUNA	111324
15	PORTO	Amarante	SALVADOR DO MONTE	130129
16	PORTO	Penafiel	CROCA	131108
17	COIMBRA	Góis	GOIS	060604
18	LISBOA	Alenquer	OLHALVO	110108
19	SANTAREM	Torres Novas	UF BROQUEIRA P IGREJA ALCOROCH	141918
20	VILA REAL	Alijó	VILA CHA	170116
21	AVEIRO	Ílhavo	ILHAVO (SAO SALVADOR)	011008
22	BEJA	Aljustrel	ERVIDEL	020102
23	BEJA	Almodôvar	SAO BARNABE	020206
24	BEJA	Beja	BALEIZAO	020502

[Close](#) [Stop](#) [Get more rows](#)

# Select Values- Configuration (1st part)

Activities Sun 18:44

Spoon - dim\_location

File Edit View Action Tools Help Connect

View Design Search

Transformations

- dim\_location
  - Run configurations
  - Database connections
  - Steps
  - Hops
  - Partition schemas
  - Slave server
  - Kettle cluster schemas
  - VFS Connections

fact\_order load\_dw Project1 Exercise3 dim\_location /home/aid/Documents/ dim\_voltage fact\_consumption Job\_Exercise6

100%

Output Updated

Step name Select values

Select & Alter Remove Meta-data

Fields :

Fieldname	Rename to	Length	Precision
1 District			
2 Municipality			
3 parish			
4 DistrictMunicipalityParishCode			

Get Fields to select Edit Mapping

Execution Results

Logging Execut

Include unspecified fields, ordered |

OK Cancel Help

2023/10/22 18:43:51 - dim\_location

2023/10/22 18:43:51 - CSV\_input\_location.0 - Header row skipped in file 'file:///home/aid/Documents/Projeto%20AID/Ex5/3-consumos-faturados-por-municipio-ultimos-10-anos.csv'

The screenshot shows the Kettle Spoon interface with a 'Select values' configuration dialog open. The dialog is titled 'Select values' and has a step name 'Select values'. It contains a table with four fields: District, Municipality, parish, and DistrictMunicipalityParishCode. There are buttons for 'Get Fields to select' and 'Edit Mapping'. Below the dialog, the execution results show a log message indicating the header row was skipped from a CSV file.

# Select Values- Configuration (2nd part)

Activities SWT

Sun 18:44

Spoon - dim\_location

File Edit View Action Tools Help

View Design

Search

Transformations

- dim\_location
  - Run configurations
  - Database connections
  - Steps
  - Hops
  - Partition schemas
  - Slave server
  - Kettle cluster schemas
  - VFS Connections

fact\_order load\_dw Project1 Exercise3 dim\_location /home/aid/Documents/ dim\_voltage fact\_consumption Job\_Exercise6

Connect

**Select values**

Step name Select values

CSV\_input\_location

Select & Alter Remove Meta-data

Fields to alter the meta-data for:

Fieldname	Rename to	Type
1 DistrictMunicipalityParishCode		Integer

Get fields to change

Execution Results

Logging Execut

2023/10/22 18:43:51 - di

2023/10/22 18:43:51 - CSV\_input\_location.0 - Header row skipped in file 'file:///home/aid/Documents/Projeto%20AID/Ex5/3-consumos-faturados-por-municipio-ultimos-10-anos.csv'

OK Cancel

The screenshot shows the Kettle Spoon interface with a 'Select values' configuration dialog open. The dialog is titled 'Select values' and has a step name 'Select values'. It contains a table to alter meta-data for a single field, 'DistrictMunicipalityParishCode', changing its type to Integer. The background shows various other transformation steps and a project overview.

# Select Values- Preview

Activities Sun 18:44

Spoon - dim\_location

File Edit View Action Tools Help

View Design fact order load dw Project1 Exercise3 dim location /home/aid/Documents/ dim voltage fact consumption Job Exercise6 Connect

Rows of step: Select values (100 rows)

	District	Municipality	parish	DistrictMunicipalityParishCode
1	GUARDA	Guarda	ARRIFANA	90706
2	PORTO	Vila do Conde	UF MALTA E CANIDELO	131633
3	SETUBAL	Alcácer do Sal	UF ALCACER SAL E SANTA SUSANA	150107
4	AVEIRO	Anadia	UF ARCOS E MOGOFORES	10317
5	BRAGA	Barcelos	FRAGOSO	30235
6	LISBOA	Sintra	UF SINTRA	111128
7	PORTALEGRE	Alter do Chão	CUNHEIRA	120104
8	VIANA DO CASTELO	Valença	BOIVAO	160802
9	AVEIRO	Aveiro	OLIVEIRINHA	10508
10	BRAGANCA	Torre de Moncorvo	UF ADEGANHA E CARDANHA	40918
11	FARO	Albufeira	PADERNE	80103
12	AVEIRO	Arouca	SAO MIGUEL DO MATO	10417
13	GUARDA	Aguiar da Beira	UF SOUTO AGUIAR BEIRA VALVERDE	90116
14	LISBOA	Torres Vedras	UF DOIS PORTOS E RUNA	111324
15	PORTO	Amarante	SALVADOR DO MONTE	130129
16	PORTO	Penafiel	CROCA	131108
17	COIMBRA	Góis	GOIS	60604
18	LISBOA	Alenquer	OLHALVO	110108
19	SANTAREM	Torres Novas	UF BROQUEIRA P IGREJA ALCOROCHE	141918
20	VILA REAL	Alijó	VILA CHA	170116
21	AVEIRO	Ílhavo	ILHAZO (SAO SALVADOR)	11008
22	BEJA	Aljustrel	ERVIDEL	20102
23	BEJA	Almodôvar	SAO BARNABE	20206
24	BEJA	Beja	BALEIZAO	20502

# Insert/Update - Configuration

Activities SWT

Tue 20:45

Spoon - dim\_location

File Edit View Action Tools Help

View Design

database loo

- Input
- Output
- Streaming
- Transform
- Utility
- Flow
- Scripting
- Pentaho Server
- Lookup
- Joins
- Data Warehouse
- Validation
- Statistics
- Big Data
- Cryptography
- Job
- Mapping
- Bulk loading
- Inline
- Experimental
- Deprecated
- History

fact\_order load\_dw Project1 Exercise3 dim\_location /home/aid/Documents/ dim\_voltage fact\_consumption Job\_Exercise6

100%

CSV\_input\_location → Filter rows → Select values → Insert / update

Output  
Updated  
Rejected  
Errors  
Active Paused  
Time 2h 58mn 51s  
Speed (r/s) 0  
input/output 4205/0

Step name: Insert / update  
Connection: energy\_dw  
Target schema: energy\_dw  
Target table: dim\_location  
Commit size: 100  
 Don't perform any updates:  
The key(s) to look up the value(s):  

Table field	Comparator	Stream field1	Stream field2
1 PARISH_CODE	=	DistrictMunicipalityParishCode	

  
Update fields:  

Table field	Stream field	Update
1 DISTRICT	District	Y
2 MUNICIPALITY	Municipality	Y
3 PARISH	parish	Y
4 PARISH_CODE	DistrictMunicipalityParishCode	Y

Execution Results

Logging Execution History Step Metrics Performance Graph Metrics

2023/10/24 20:10:10 - Spoon - Job has ended.  
2023/10/24 20:29:18 - Spoon - Processing of transformation stopped.  
2023/10/24 20:29:53 - Spoon - Processing of transformation stopped.  
2023/10/24 20:30:22 - Spoon - Processing of transformation stopped.  
2023/10/24 20:30:59 - Spoon - Processing of transformation stopped.  
2023/10/24 20:32:05 - Spoon - Processing of transformation stopped.  
2023/10/24 20:33:16 - Spoon - Processing of transformation stopped.  
2023/10/24 20:34:06 - Spoon - Processing of transformation stopped.  
2023/10/24 20:43:20 - Spoon - Spoon  
2023/10/24 20:43:26 - Spoon - Spoon  
2023/10/24 20:43:30 - Spoon - Spoon

# Insert/Update - Preview

Activities    SWT    Examine preview data

Tue 20:50

Rows of step: insert / update (100 rows)

	District	Municipality	parish	District	Municipality	ParishCode
1	GUARDA	Guarda	ARRIFANA			90706
2	PORTO	Vila do Conde	UF MALTA E CANIDELO			131633
3	SETUBAL	Alcácer do Sal	UF ALCACER SAL E SANTA SUSANA			150107
4	AVEIRO	Anadia	UF ARCOS E MOGOFORES			10317
5	BRAGA	Barcelos	FRAGOSO			30235
6	LISBOA	Sintra	UF SINTRA			111128
7	PORTALEGRE	Alter do Chão	CUNHEIRA			120104
8	VIANA DO CASTELO	Valença	BOIVAO			160802
9	AVEIRO	Aveiro	OLIVEIRINHA			10508
10	BRAGANCA	Torre de Moncorvo	UF ADEGANHA E CARDANHA			40918
11	FARO	Albufeira	PADERNE			80103
12	AVEIRO	Arouca	SAO MIGUEL DO MATO			10417
13	GUARDA	Aguiar da Beira	UF SOUTO AGUIAR BEIRA VALVERDE			90116
14	LISBOA	Torres Vedras	UF DOIS PORTOS E RUNA			111324
15	PORTO	Amarante	SALVADOR DO MONTE			130129
16	PORTO	Penafiel	CROCA			131108
17	COIMBRA	Góis	GOIS			60604
18	LISBOA	Alenquer	OLHALVO			110108
19	SANTAREM	Torres Novas	UF BROQUEIRA P IGREJA ALCOROCH			141918
20	VILA REAL	Alijó	VILA CHA			170116
21	AVEIRO	Ílhavo	ILHAZO (SAO SALVADOR)			11008
22	BEJA	Aljustrel	ERVIDEL			20102
23	BEJA	Almodôvar	SAO BARNABE			20206
24	BEJA	Beja	BALEIZAO			20502
25	BRAGA	Braga	UF NOGUEIRA FRAIAO E LAMACAES			30377
26	PORTO	Vila do Conde	ARVORE			131602
27	SETUBAL	Santiago do Cacém	ABELA			150901
28	VIANA DO CASTELO	Ponte da Barca	UF P BARCA MUIA P V MAGALHAES			160628
29	BRAGA	Vila Nova de Famalicão	RIBEIRAO			31235
30	GUARDA	Guarda	FAIA			90716
31	LISBOA	Oeiras	UF CARNAXIDE E QUEIJAS			111013
32	PORTALEGRE	Ponte de Sor	GALVEIAS			121301
33	PORTO	Felgueiras	UF UNHAO E LORDELO			130338
34	PORTO	Lousada	LODARES			130510

Close

Stop

Get more rows

# Dim\_time – Entire transformation & Output table

Activities Terminal Sun 18:38

Spoon - dim\_time (changed)

Project1 Exercise3 dim\_location /home/aid/Documents/ 14

View Design

Transformations

- dim\_time
  - Run configurations
  - Database connections
  - Steps
  - Hops
  - Partition schemas
  - Slave server
  - Kettle cluster schemas
  - VFS Connections

CSV\_input\_time

Insert / update

Value mapper 2

File Edit View Search Terminal Help

AVEIRO | Arouca | UF AROUCA E BURGO | 10421 |

40 rows in set (0.10 sec)

```
mysql> select * from dim_time LIMIT 40;
```

YEAR	MONTH	MONTH_NAME	DATE_ID
2020	11	Nov	2020-11
2020	12	Dec	2020-12
2021	1	Jan	2021-01
2021	2	Feb	2021-02
2021	3	Mar	2021-03
2021	4	Apr	2021-04
2021	5	May	2021-05
2021	6	Jun	2021-06
2021	7	Jul	2021-07
2021	8	Aug	2021-08
2021	9	Sep	2021-09
2021	10	Oct	2021-10
2021	11	Nov	2021-11
2021	12	Dec	2021-12
2022	1	Jan	2022-01
2022	2	Feb	2022-02
2022	3	Mar	2022-03
2022	4	Apr	2022-04
2022	5	May	2022-05
2022	6	Jun	2022-06
2022	7	Jul	2022-07
2022	8	Aug	2022-08
2022	9	Sep	2022-09
2022	10	Oct	2022-10
2022	11	Nov	2022-11
2022	12	Dec	2022-12
2023	1	Jan	2023-01
2023	2	Feb	2023-02
2023	3	Mar	2023-03
2023	4	Apr	2023-04
2023	5	May	2023-05
2023	6	Jun	2023-06
2023	7	Jul	2023-07
2023	8	Aug	2023-08

34 rows in set (0.01 sec)

```
mysql>
```

a screenshot with the entire contents of the corresponding XML file.

# CSV\_input\_time - Configuration

Activities SWT

Sun 18:51

Spoon - dim\_time (changed)

File Edit View Action Tools Help

View Design

Search

Transformations

- dim\_time
  - Run configurations
  - Database connections
  - Steps
  - Hops
  - Partition schemas
  - Slave server
  - Kettle cluster schemas
  - VFS Connections

fact\_order load\_dw Project1 Exercise3 dim\_location /home/aid/Documents/ dim\_voltage fact\_consumption Job\_Exercise6

100%

CSV file input

Step name: CSV\_input\_time

Filename: /home/aid/Documents/Projeto AID/Ex5/3-consumos-faturados-po

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 Year	Integer	#	15	0	\$	.	,	none
2 Month	Integer	#	15	0	\$	.	,	none
3 Date	String		7		\$	.	,	none

Help OK Get Fields Preview Cancel

```
graph TD; CSV[CSV_input_time] --> VM[Value mapper 2]; VM --> Insert[Insert / update]
```

# CSV\_input\_time - Preview

Activities SWT Sun 18:52 Spoon - dim\_time (changed)

File Edit View Action Tools Help Connect

View Design Search

Transformations dim\_time

- Run configurations
- Database connections
- Steps
- Hops
- Partition schemas
- Slave server
- Kettle cluster schemas
- VFS Connections

fact\_order load\_dw Project1 Exercise3 dim\_location /home/aid/Documents/ dim\_voltage fact\_consumption Job\_Exercise6

Errors 0 Active Paused Time 22.1s Speed (r/s) 5 input/output 0/100

CSV\_input\_time Insert / update Value mapper 2

Examine preview data

Rows of step: CSV\_input\_time (100 rows)

	Year	Month	Date
1	2020	11	2020-11
2	2020	11	2020-11
3	2020	11	2020-11
4	2020	12	2020-12
5	2020	12	2020-12
6	2020	12	2020-12
7	2020	12	2020-12
8	2020	12	2020-12
9	2021	1	2021-01
10	2021	1	2021-01
11	2021	1	2021-01
12	2021	2	2021-02
13	2021	2	2021-02
14	2021	2	2021-02
15	2021	2	2021-02
16	2021	2	2021-02
17	2021	3	2021-03
18	2021	3	2021-03
19	2021	3	2021-03
20	2021	3	2021-03
21	2021	4	2021-04
22	2021	4	2021-04
23	2021	4	2021-04
24	2021	4	2021-04

Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Preview data

2023/10/22 17:28:52 - Spoon - Launching transformation [fact\_consumption]...  
2023/10/22 17:28:52 - Spoon - Started the transformation execution.  
2023/10/22 17:28:53 - Spoon - The transformation has finished!!  
2023/10/22 17:30:32 - Spoon - Spoon  
2023/10/22 17:31:23 - Spoon - Starting job...  
2023/10/22 18:17:08 - Spoon - Job has ended.  
2023/10/22 18:43:29 - Spoon - Processing of transformation stopped.  
2023/10/22 18:44:28 - Spoon - Processing of transformation stopped.  
2023/10/22 18:45:07 - Spoon - Processing of transformation stopped.  
2023/10/22 18:51:50 - dim\_time - Dispatching started for transformation [dim\_time]  
2023/10/22 18:51:50 - CSV\_input\_time.0 - Header row skipped in file '/home/aid/Documents/Projeto AID/Exercícios/exercício 6/fact\_consumption/dados/por\_mês/dim\_time/dim\_time.csv'

Close Stop Get more rows

# Value Mapper - Configuration

Activities SWT

Sun 18:52

Spoon - dim\_time (changed)

File Edit View Action Tools Help

View Design Search

Transformations

- dim\_time
  - Run configurations
  - Database connections
  - Steps
  - Hops
  - Partition schemas
  - Slave server
  - Kettle cluster schemas
  - VFS Connections

fact\_order load\_dw Project1 Exercise3 dim\_location

Copynr 0  
Read 1  
Written 0  
Input 0  
Output 0  
Updated 0  
Rejected 0  
Errors 0  
Active Paused  
Time 34.1s  
Speed (r/s) 0  
input/output 100/0

Insert / update

Value mapper 2

Step name: Value mapper 2

Fieldname to use: Month

Target field name (empty=overwrite): Month\_name

Default upon non-matching:

Field values:

Source value	Target value
1 1	Jan
2 2	Feb
3 3	Mar
4 4	Apr
5 5	May
6 6	Jun
7 7	Jul
8 8	Aug
9 9	Sep
10 10	Oct
11 11	Nov
12 12	Dec

Execution Results

Logging Execution History Step Metrics Performance Graph

2023/10/22 17:28:52 - Spoon - Launching transformation [fact\_consumption]...  
2023/10/22 17:28:52 - Spoon - Started the transformation execution.  
2023/10/22 17:28:53 - Spoon - The transformation has finished!!  
2023/10/22 17:30:32 - Spoon - Spoon  
2023/10/22 17:31:23 - Spoon - Starting job...  
2023/10/22 18:17:08 - Spoon - Job has ended.  
2023/10/22 18:43:29 - Spoon - Processing of transformation stopped.  
2023/10/22 18:44:28 - Spoon - Processing of transformation stopped.  
2023/10/22 18:45:07 - Spoon - Processing of transformation stopped.  
2023/10/22 18:51:50 - dim\_time - Dispatching started for transformation [dim\_time]  
2023/10/22 18:51:50 - CSV\_input\_time.0 - Header row skipped in file '/home/aid/D

OK Cancel

# Value Mapper - Preview

Activities SWT

Sun 18:53

Spoon - dim\_time (changed)

File Edit View Action Tools Help

View Design

Search

Transformations

- dim\_time
  - Run configurations
  - Database connections
  - Steps
  - Hops
  - Partition schemas
  - Slave server
  - Kettle cluster schemas
  - VFS Connections

fact\_order load\_dw Project1 Exercise3 dim\_location /home/aid/Documents/ dim\_voltage fact\_consumption Job\_Exercise6

100%

Copynr 0  
Read 101  
Written 101  
Input 0  
Output 0  
Updated 0  
Rejected 0  
Errors 0  
Active Paused  
Time 9.1s  
Speed (r/s) 11  
input/output 1389/100

Insert/update

Value mapper 2

Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Preview data

2023/10/22 18:52:57 - dim\_time - Dispatching started for transformation [dim\_time]  
2023/10/22 18:52:57 - CSV\_input\_time.0 - Header row skipped in file '/home/aid/Documents/Projeto AID/Ex'

Examine preview data

Rows of step: Value mapper 2 (100 rows)

	Year	Month	Date	Month_name
1	2020	11	2020-11	Nov
2	2020	11	2020-11	Nov
3	2020	11	2020-11	Nov
4	2020	12	2020-12	Dec
5	2020	12	2020-12	Dec
6	2020	1	Integer(15)	Dec
7	2020	12	2020-12	Dec
8	2020	12	2020-12	Dec
9	2021	1	2021-01	Jan
10	2021	1	2021-01	Jan
11	2021	1	2021-01	Jan
12	2021	2	2021-02	Feb
13	2021	2	2021-02	Feb
14	2021	2	2021-02	Feb
15	2021	2	2021-02	Feb
16	2021	2	2021-02	Feb
17	2021	3	2021-03	Mar
18	2021	3	2021-03	Mar
19	2021	3	2021-03	Mar
20	2021	3	2021-03	Mar
21	2021	4	2021-04	Apr
22	2021	4	2021-04	Apr
23	2021	4	2021-04	Apr
24	2021	4	2021-04	Apr

Close Stop Get more rows

# Insert/Update - Configuration

Activities SWT

Sun 18:53

Spoon - dim\_time (changed)

File Edit View Action Tools Help

View Design

Search

Transformations

- dim\_time
  - Run configurations
  - Database connections
  - Steps
  - Hops
  - Partition schemas
  - Slave server
  - Kettle cluster schemas
  - VFS Connections

fact\_order load\_dw Project1 Exercise3 dim\_location /home/aid/Documents/ dim\_voltage fact\_consumption Job\_Exercise6

Connect

**Insert / update**

Step name: Insert / update

Connection: energy\_dw

Target schema: energy\_dw

Target table: dim\_time

Commit size: 100

Don't perform any updates:

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

Table field	Comparator	Stream field1	Stream field2
1 DATE_ID	=	Date	

**Update fields:**

Table field	Stream field	Update
1 YEAR	Year	Y
2 MONTH	Month	Y
3 DATE_ID	Date	Y
4 MONTH_NAME	Month_name	Y

Help OK Cancel SQL

view data

objeto AID/Ex5/3-consumos-faturados-por-municipio-ultimos-10-anos.csv'

# Insert/Update - Preview

Activities SWT

Sun 18:53

Spoon - dim\_time (changed)

File Edit View Action Tools Help

View Design

Search

Transformations

dim\_time

- Run configurations
- Database connections
- Steps
- Hops
- Partition schemas
- Slave server
- Kettle cluster schemas
- VFS Connections

fact\_order load\_dw Project1 Exercise3 dim\_location /home/aid/Documents/ dim\_voltage fact\_consumption Job\_Exercise6

Connect

CSV\_input\_time

Value mapper 2

Insert / update

Errors 0  
Active Paused  
Time 7.1s  
Speed (r/s) 14  
input/output 9971/0

Examine preview data

Rows of step: Insert / update (100 rows)

	Year	Month	Date	Month_name
1	2020	11	2020-11	Nov
2	2020	11	2020-11	Nov
3	2020	11	2020-11	Nov
4	2020	12	2020-12	Dec
5	2020	12	2020-12	Dec
6	2020	12	2020-12	Dec
7	2020	12	2020-12	Dec
8	2020	12	2020-12	Dec
9	2021	1	2021-01	Jan
10	2021	1	2021-01	Jan
11	2021	1	2021-01	Jan
12	2021	2	2021-02	Feb
13	2021	2	2021-02	Feb
14	2021	2	2021-02	Feb
15	2021	2	2021-02	Feb
16	2021	2	2021-02	Feb
17	2021	3	2021-03	Mar
18	2021	3	2021-03	Mar
19	2021	3	2021-03	Mar
20	2021	3	2021-03	Mar
21	2021	4	2021-04	Apr
22	2021	4	2021-04	Apr
23	2021	4	2021-04	Apr
24	2021	4	2021-04	Apr

Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Preview data

2023/10/22 18:53:35 - dim\_time - Dispatching started for transformation [dim\_time]  
2023/10/22 18:53:35 - CSV\_input\_time.0 - Header row skipped in file '/home/aid/Documents/Projeto AID/Ex'

Close Stop Get more rows

# Dim\_voltage – Entire transformation & Output table

Activities Terminal Sun 18:39

Spoon - dim\_voltage

File Edit View Action Tools Help Connect

View Design Search Transformations dim\_voltage Run configurations Database connections Steps Hops Partition schemas Slave server Kettle cluster schemas VFS Connections

fact\_order load\_dw Project1 Exercise3 dim\_location /home/aid/Documents/ dim\_voltage fact\_consumption Job\_Exercise6

CSV\_input\_voltage Insert/update Value mapper

aid@virtualbox: ~/Documents/Projeto AID/Ex5

```
mysql> select * from dim_voltage LIMIT 40;
+-----+-----+
| VOLTAGE_LEVEL | VOLTAGE_ID |
+-----+-----+
| Muito Alta, Alta e Média Tensões | 1 |
| Baixa Tensão | 2 |
+-----+-----+
2 rows in set (0.00 sec)

mysql>
```

Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Preview data

2023/10/22 17:10:46 - Spoon - Processing of transformation stopped.  
2023/10/22 17:10:46 - Insert / update.0 - Finished processing (I=3295, O=0, R=3295, W=3295, U=0, E=0)  
2023/10/22 17:10:46 - Value mapper.0 - Finished processing (I=0, O=0, R=13296, W=13296, U=0, E=0)  
2023/10/22 17:10:46 - CSV\_input\_voltage.0 - Finished processing (I=23298, O=0, R=0, W=23297, U=0, E=0)  
2023/10/22 17:28:52 - Spoon - Transformation opened.  
2023/10/22 17:28:52 - Spoon - Launching transformation [fact\_consumption]...  
2023/10/22 17:28:52 - Spoon - Started the transformation execution.  
2023/10/22 17:28:53 - Spoon - The transformation has finished!!  
2023/10/22 17:30:32 - Spoon - Spoon  
2023/10/22 17:31:23 - Spoon - Starting job...  
2023/10/22 18:17:08 - Spoon - Job has ended.

# CSV\_input\_voltage - Configuration

Activities Sun 18:56

Spoon - dim\_voltage

File Edit View Action Tools Help

View Design Search

Transformations

- dim\_voltage
  - Run configurations
  - Database connections
  - Steps
  - Hops
  - Partition schemas
  - Slave server
  - Kettle cluster schemas
  - VFS Connections

fact\_order load\_dw Project1 Exercise3 dim\_location /home/aid/Documents/ dim\_voltage fact\_consumption Job\_Exercise6

100%

CSV file input

Step name: CSV\_input\_voltage

Filename: /home/aid/Documents/Projeto AID/Ex5/3-consumos-faturados-po

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: UTF-8

Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim type
1 Voltage level	String		32		\$	.	,	none

Execution Results

Logging Execution History Step Metrics Performance

2023/10/22 17:28:52 - Spoon - Launching transformation [fact\_consum...  
2023/10/22 17:28:52 - Spoon - Started the transformation execution.  
2023/10/22 17:28:53 - Spoon - The transformation has finished!!  
2023/10/22 17:30:32 - Spoon - Spoon  
2023/10/22 17:31:23 - Spoon - Starting job...  
2023/10/22 18:17:08 - Spoon - Job has ended.  
2023/10/22 18:43:29 - Spoon - Processing of transformation stopped.  
2023/10/22 18:44:28 - Spoon - Processing of transformation stopped.  
2023/10/22 18:45:07 - Spoon - Processing of transformation stopped.  
2023/10/22 18:52:50 - Spoon - Processing of transformation stopped.  
2023/10/22 18:53:28 - Spoon - Processing of transformation stopped.

Help OK Get Fields Preview Cancel

```
graph TD; CSV[CSV_input_voltage] --> ValueMapper[Value mapper]; ValueMapper --> Insert[Insert/update];
```

# CSV\_input\_voltage - Preview

Activities Sun 18:57

File Edit View Action Tools Help

Spoon - dim\_voltage

View Design Search

Transformations

- dim\_voltage
  - Run configurations
  - Database connections
  - Steps
  - Hops
  - Partition schemas
  - Slave server
  - Kettle cluster schemas
  - VFS Connections

fact\_order load\_dw Project1 Exercise3 dim\_location /home/aid/Documents/ dim\_voltage fact\_consumption Job\_Exercise6

Connect

Active Paused  
Time 19.1s  
Speed (r/s) 5  
input/output 0/100

CSV\_input\_voltage Insert/update Value mapper

Examine preview data

Rows of step: CSV\_input\_voltage (100 rows)

Voltage level
1 Muito Alta, Alta e Média Tensões
2 Baixa Tensão
3 Muito Alta, Alta e Média Tensões
4 Baixa Tensão
5 Muito Alta, Alta e Média Tensões
6 Baixa Tensão
7 Muito Alta, Alta e Média Tensões
8 Baixa Tensão
9 Baixa Tensão
10 Baixa Tensão
11 Baixa Tensão
12 Baixa Tensão
13 Baixa Tensão
14 Muito Alta, Alta e Média Tensões
15 Baixa Tensão
16 Muito Alta, Alta e Média Tensões
17 Baixa Tensão
18 Baixa Tensão
19 Baixa Tensão
20 Muito Alta, Alta e Média Tensões
21 Muito Alta, Alta e Média Tensões
22 Baixa Tensão
23 Baixa Tensão
24 Muito Alta, Alta e Média Tensões

Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Preview data

2023/10/22 18:56:59 - dim\_voltage - Dispatching started for transformation [dim\_voltage]  
2023/10/22 18:56:59 - CSV\_input\_voltage.0 - Header row skipped in file '/home/aid/Documents/Projeto AID'

Close Stop Get more rows

# Value Mapper - Configuration

Activities Sun 18:57

Spoon - dim\_vo

File Edit View Action Tools Help

View Design Search

Transformations

- dim\_voltage
  - Run configurations
  - Database connections
  - Steps
  - Hops
  - Partition schemas
  - Slave server
  - Kettle cluster schemas
  - VFS Connections

fact\_order load\_dw Project1 Exercise3 dim\_lc

CSV\_input\_voltage

Input	0
Output	0
Updated	0
Rejected	0
Errors	0
Active	Paused
Time	35.1s
Speed (r/s)	0
input/output	100/0

Value mapper

Step name: Value mapper

Fieldname to use: Voltage level

Target field name (empty=overwrite): Voltage\_Id

Default upon non-matching: Sentinel

Field values:

Source value	Target value
1 Baixa Tensão	2
2 Muito Alta, Alta e Média Tensões	1

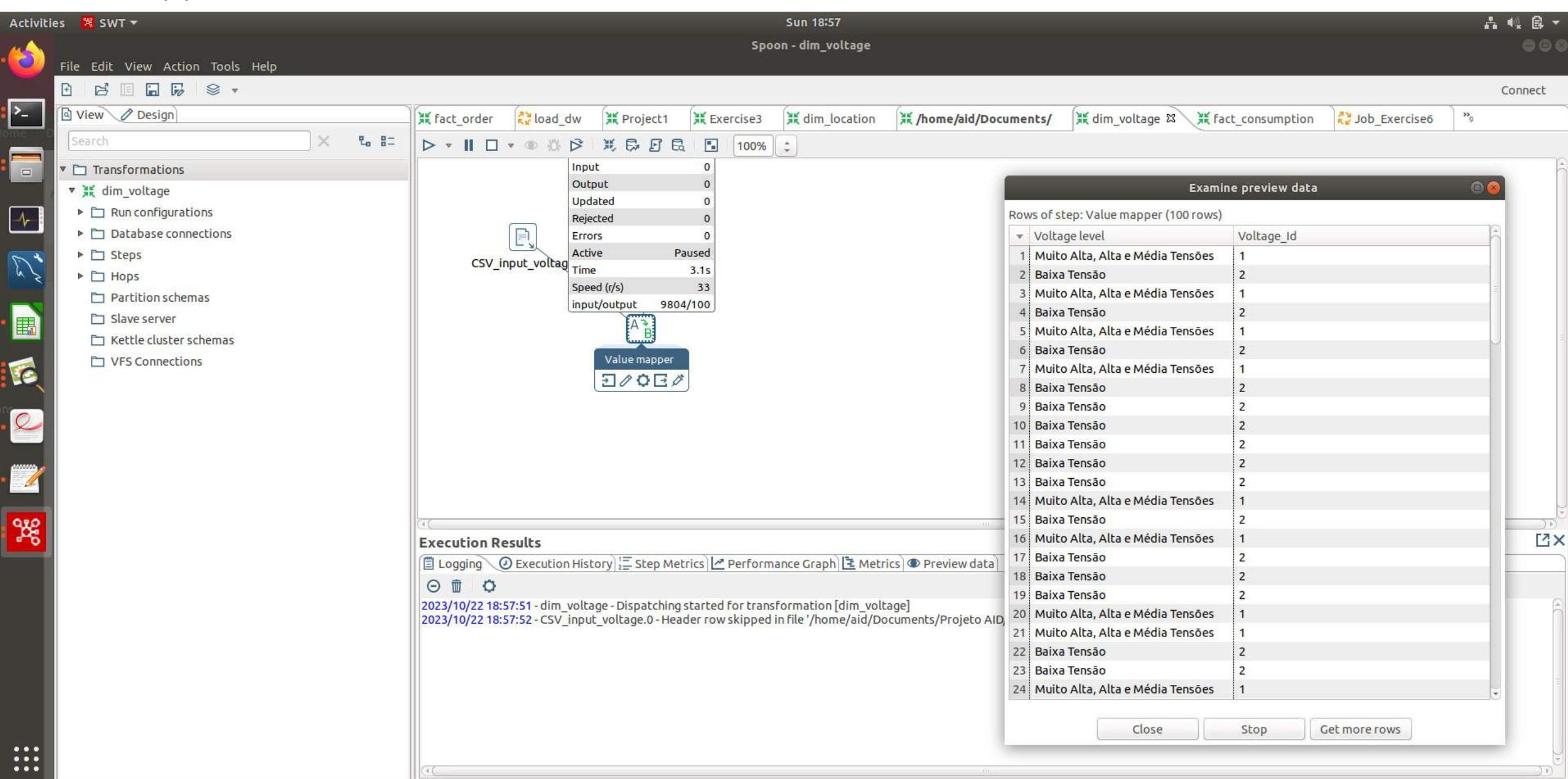
Execution Results

Logging Execution History Step Metrics Performance Graph

2023/10/22 18:56:59 - dim\_voltage - Dispatching started for transformation [2023/10/22 18:56:59 - CSV\_input\_voltage.0 - Header row skipped in file '/hom

OK Cancel

# Value Mapper - Preview



# Insert/Update - Configuration

Activities Sun 18:58

Spoon - dim\_voltage

File Edit View Action Tools Help

View Design Search

Transformations

- dim\_voltage
  - Run configurations
  - Database connections
  - Steps
  - Hops
  - Partition schemas
  - Slave server
  - Kettle cluster schemas
  - VFS Connections

fact\_order load\_dw Project1 Exercise3 dim\_location /home/aid/Documents/ dim\_voltage fact\_consumption Job\_Exercise6

Connect

**Execution Results**

Logging Execution History Step Metrics Performance Graph

2023/10/22 18:57:51 - dim\_voltage - Dispatching started for transformation [dim\_voltage]

2023/10/22 18:57:52 - CSV\_input\_voltage.0 - Header row skipped in file '/home/aid/Documents/dim\_voltage.csv'

2023/10/22 18:57:59 - Spoon - Processing of transformation stopped.

2023/10/22 18:57:59 - Value mapper.0 - Finished processing (I=0, O=0, R=101, W=0)

2023/10/22 18:57:59 - CSV\_input\_voltage.0 - Finished processing (I=9907, O=0, W=0)

2023/10/22 18:57:59 - Insert / update.0 - Finished processing (I=1, O=0, R=1, W=0)

**Insert / update**

Step name: **Insert / update**

Connection: **energy\_dw**

Target schema: **energy\_dw**

Target table: **dim\_voltage**

Commit size: **100**

Don't perform any updates:

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

Table field	Comparator	Stream field1	Stream field2
1 VOLTAGE_ID	=	Voltage_Id	

Get fields

Update fields:

Table field	Stream field	Update
1 VOLTAGE_LEVEL	Voltage level	Y
2 VOLTAGE_ID	Voltage_Id	Y

Get update fields

Help OK Cancel SQL

```
graph TD; CSV[CSV_input_voltage] --> IU[Insert / update]; IU --> VM[Value mapper A B];
```

# Insert/Update - Preview

Activities Sun 18:58

Spoon - dim\_voltage

File Edit View Action Tools Help

View Design Search

Transformations

- dim\_voltage
  - Run configurations
  - Database connections
  - Steps
  - Hops
  - Partition schemas
  - Slave server
  - Kettle cluster schemas
  - VFS Connections

fact\_order load\_dw Project1 Exercise3 dim\_location /home/aid/Documents/ dim\_voltage fact\_consumption Job\_Exercise6

Connect

CSV\_input\_voltage → Insert / update → Value mapper

Active: Paused: Time: 3.1s Speed (r/s): 33 input/output: 9995/0

Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Preview data

2023/10/22 18:58:20 - dim\_voltage - Dispatching started for transformation [dim\_voltage]  
2023/10/22 18:58:20 - CSV\_input\_voltage.0 - Header row skipped in file '/home/aid/Documents/Projeto AID'

Examine preview data

Rows of step: Insert / update (100 rows)	Voltage level	Voltage_Id
1	Muito Alta, Alta e Média Tensões	1
2	Baixa Tensão	2
3	Muito Alta, Alta e Média Tensões	1
4	Baixa Tensão	2
5	Muito Alta, Alta e Média Tensões	1
6	Baixa Tensão	2
7	Muito Alta, Alta e Média Tensões	1
8	Baixa Tensão	2
9	Baixa Tensão	2
10	Baixa Tensão	2
11	Baixa Tensão	2
12	Baixa Tensão	2
13	Baixa Tensão	2
14	Muito Alta, Alta e Média Tensões	1
15	Baixa Tensão	2
16	Muito Alta, Alta e Média Tensões	1
17	Baixa Tensão	2
18	Baixa Tensão	2
19	Baixa Tensão	2
20	Muito Alta, Alta e Média Tensões	1
21	Muito Alta, Alta e Média Tensões	1
22	Baixa Tensão	2
23	Baixa Tensão	2
24	Muito Alta, Alta e Média Tensões	1

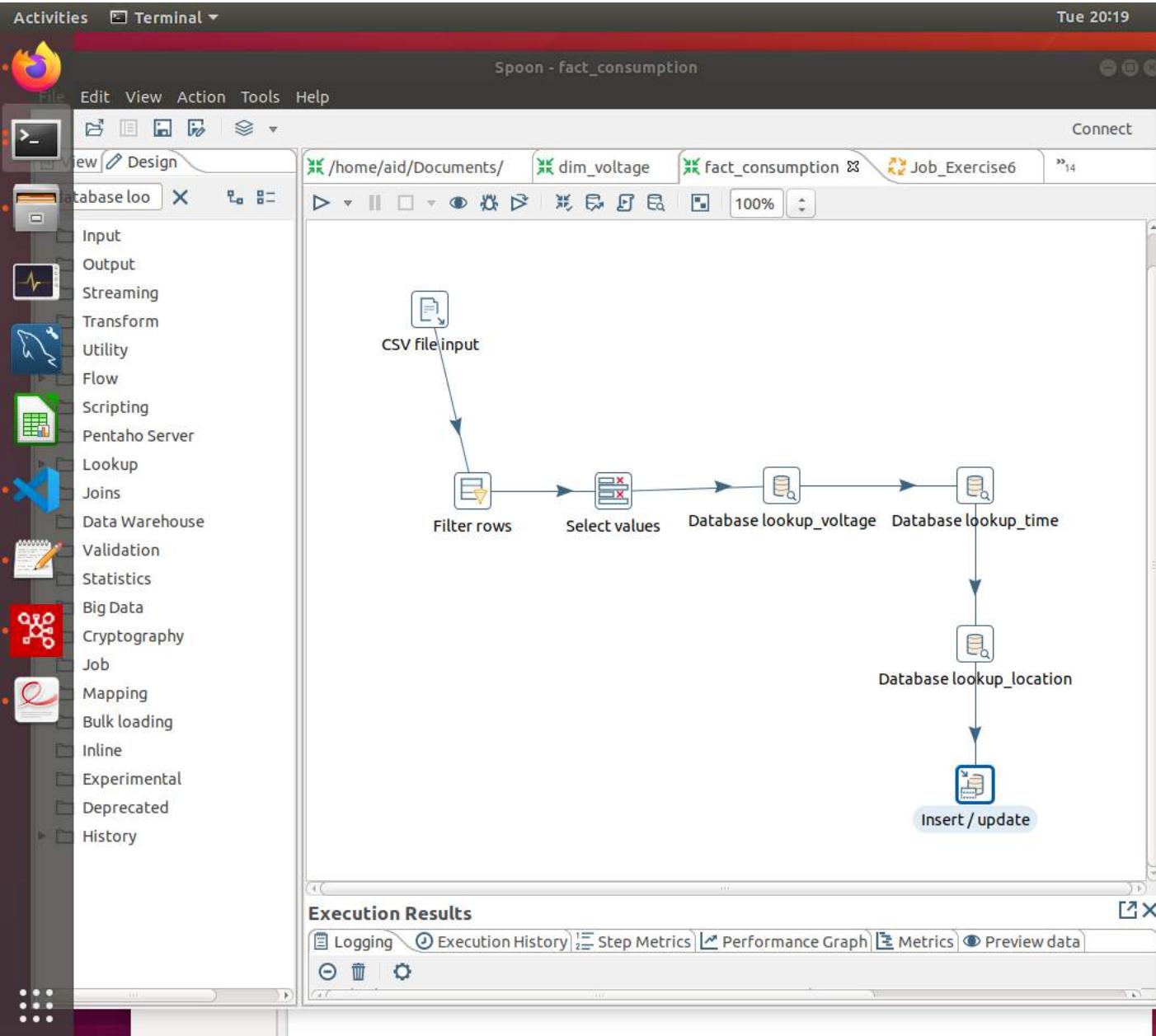
Close Stop Get more rows

# Exercise 6 – Group 50

José Cardoso

Matheus Trindade

# Fact\_consumption – Entire transformation & Output table



aid@virtualbox: ~

```
mysql> DELETE FROM dim_voltage;
Query OK, 2 rows affected (0.04 sec)

mysql> select * from fact_consumption LIMIT 40;
```

ACTIVE_ENERGY	DATE_ID	PARISH_CODE	VOLTAGE_ID	CONSUMPTION_ID
57149.7	2020-11	90706	1	55588129
303182	2020-11	131633	2	55588130
2994080	2020-11	150107	1	55588131
1599650	2020-12	10317	2	55588132
18873.2	2020-12	30235	1	55588133
9924020	2020-12	111128	2	55588134
17042.9	2020-12	120104	1	55588135
22110.5	2020-12	160802	2	55588136
859496	2021-01	10508	2	55588137
88215.5	2021-01	40918	2	55588138
1025360	2021-01	80103	2	55588139
74644.6	2021-02	10417	2	55588140
93260.4	2021-02	90116	2	55588141
297382	2021-02	111324	1	55588142
118805	2021-02	130129	2	55588143
43728.2	2021-02	131108	1	55588144
419798	2021-03	60604	2	55588145
274569	2021-03	110108	2	55588146
426617	2021-03	141918	2	55588147
125765	2021-03	170116	1	55588148
3817418	2021-04	11008	1	55588149
135292	2021-04	20102	2	55588150
76417.8	2021-04	20206	2	55588151
972687	2021-04	20502	1	55588152
2210890	2021-04	30377	2	55588153
349936	2021-04	131602	1	55588154
189779	2021-04	150901	2	55588155
54819.8	2021-04	160628	1	55588156
1558100	2021-05	31235	2	55588157
8121.79	2021-05	90716	1	55588158
5793450	2021-05	111013	2	55588159
36218.7	2021-05	121301	1	55588160
116918	2021-05	130338	2	55588161
283550	2021-05	130510	2	55588162
199985	2021-05	130532	1	55588163
203887	2021-05	131311	1	55588164
16138	2021-05	160307	2	55588165
267187	2021-05	171409	2	55588166
181396	2021-05	171420	2	55588167

# CSV\_file\_input - Configuration

Activities SWT

Tue 20:21

Spoon - fact\_consumption

File Edit View Action Tools Help

View Design

dim\_product /home/aid/Downloads/ fact\_order load\_dw Project1 Exercise3 dim\_location /home/aid/Documents/ dim\_voltage fact\_consumption Job\_Exercise6

Connect

loo X

Input Output Streaming Transform Utility Flow Scripting Pentaho Server Lookup Joins Data Warehouse Validation Statistics Big Data Cryptography Job Mapping Bulk loading Inline Experimental Deprecated History

CSV file input

Filter rows Select values Database lookup\_voltage Database lookup\_time Database lookup\_location Insert / update

CSV file input

Step name CSV File input  
Filename /home/aid/Documents/Projeto AID/Ex5/3-consumos-faturados-po  
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 UTF-8

Name	Type	Format	Length	Precision	Currency	Decimal
1 Year	Integer	#	15	0	\$	.
2 Month	Integer	#	15	0	\$	.
3 Date	String		7		\$	.
4 District	String		16		\$	.
5 Municipality	String		17		\$	.
6 parish	String		29		\$	.
7 Voltage level	String		32		\$	.
8 Active Energy (kwh)	Number	#.#	11	3	\$	.
9 DistrictMunicipalityParishCode	String	#	15	0	\$	.

Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Preview data

Help OK Get Fields Preview Cancel

# CSV\_file\_input - Preview

Activities SWT

Tue 20:29

Spoon - fact\_consumption

File Edit View Action Tools Help

View Design

loo X

Input Output Streaming Transform Utility Flow Scripting Pentaho Server Lookup Joins Data Warehouse Validation Statistics Big Data Cryptography Job Mapping Bulk loading Inline Experimental Deprecated History

dim\_product /home/aid/Downloads/ fact\_order load\_dw Project1 Exercise3 dim\_location /home/aid/Documents/ dim\_voltage fact\_consumption Job\_Exercise6

Connect

Examine preview data

Rows of step: CSV file input (100 rows)

	Year	Month	Date	District	Municipality	parish	Voltage level	Active Energy (kWh)	DistrictMunicipalityParishCode
1	2020	11	2020-11	GUARDA	Guarda	ARRIFANA	Muito Alta, Alta e Média Tensões	57149.7	090706
2	2020	11	2020-11	PORTO	Vila do Conde	UF MALTA E CANIDELO	Baixa Tensão	303181.9	131633
3	2020	11	2020-11	SETUBAL	Alcácer do Sal	UF ALCACER SAL E SANTA SUSANA	Muito Alta, Alta e Média Tensões	2994085.1	150107
4	2020	12	2020-12	AVEIRO	Anadia	UF ARCOS E MOGOFORES	Baixa Tensão	1599645.3	010317
5	2020	12	2020-12	BRAGA	Barcelos	FRAGOSO	Muito Alta, Alta e Média Tensões	18873.2	030235
6	2020	12	2020-12	LISBOA	Sintra	UF SINTRA	Baixa Tensão	9924017.6	111128
7	2020	12	2020-12	PORTALEGRE	Alter do Chão	CUNHEIRA	Muito Alta, Alta e Média Tensões	17042.9	120104
8	2020	12	2020-12	VIANA DO CASTELO	Valença	BOIAO	Baixa Tensão	22110.5	160802
9	2021	1	2021-01	AVEIRO	Aveiro	OLIVEIRINHA	Baixa Tensão	859496.3	010508
10	2021	1	2021-01	BRAGANCA	Torre de Moncorvo	UF ADEGANHA E CARDANHA	Baixa Tensão	88215.5	040918
11	2021	1	2021-01	FARO	Albufeira	PADERNE	Baixa Tensão	1025357.6	080103
12	2021	2	2021-02	AVEIRO	Arouca	SAO MIGUEL DO MATO	Baixa Tensão	74644.6	010417
13	2021	2	2021-02	GUARDA	Aguiar da Beira	UF SOUTO AGUIAR BEIRA VALVERDE	Baixa Tensão	93260.4	090116
14	2021	2	2021-02	LISBOA	Torres Vedras	UF DOIS PORTOS E RUNA	Muito Alta, Alta e Média Tensões	297381.6	111324
15	2021	2	2021-02	PORTO	Amarante	SALVADOR DO MONTE	Baixa Tensão	118804.7	130129
16	2021	2	2021-02	PORTO	Penafiel	CROCA	Muito Alta, Alta e Média Tensões	43728.2	131108
17	2021	3	2021-03	COIMBRA	Góis	GOIS	Baixa Tensão	419797.5	060604
18	2021	3	2021-03	LISBOA	Alenquer	OLHALVO	Baixa Tensão	274569.1	110108
19	2021	3	2021-03	SANTAREM	Torres Novas	UF BROQUEIRA P IGREJA ALCOROCH	Baixa Tensão	426617.5	141918
20	2021	3	2021-03	VILA REAL	Alijó	VILA CHA	Muito Alta, Alta e Média Tensões	125764.7	170116
21	2021	4	2021-04	AVEIRO	Ilhavo	ILHAZO (SAO SALVADOR)	Muito Alta, Alta e Média Tensões	3817408.5	011008
22	2021	4	2021-04	BEJA	Aljustrel	ERVIDEL	Baixa Tensão	135291.7	020102
23	2021	4	2021-04	BEJA	Almodôvar	SAO BARNABE	Baixa Tensão	76417.8	020206
24	2021	4	2021-04	BEJA	Paiz	PALEIZAO	Muito Alta, Alta e Média Tensões	072607.5	020502

Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Preview data

Close Stop Get more rows

# Filter Rows – Configuration (Filtering only numeric values in the strings)

Activities SWT

Tue 20:21

Spoon - fact\_consumption

File Edit View Action Tools Help

View Design

loo X

- Input
- Output
- Streaming
- Transform
- Utility
- Flow
- Scripting
- Pentaho Server
- Lookup
- Joins
- Data Warehouse
- Validation
- Statistics
- Big Data
- Cryptography
- Job
- Mapping
- Bulk loading
- Inline
- Experimental
- Deprecated
- History

dim\_product /home/aid/Downloads/ fact\_order load\_dw Project1 Exercise3 dim\_location /home/aid/Documents/ dim\_voltage fact\_consumption Job\_Exercise6

100%

CSV fileinput → Filter rows → Select values → Database lookup\_voltage → Database lookup\_time → Database lookup\_location → Insert / update

Filter rows

Step name: Filter rows

Send 'true' data to step:

Send 'false' data to step:

The condition:

District Municipality ParishCode REGEXP ^[0-9]+\$ (String)

OK Cancel

Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Preview data

# Filter Rows - Preview

Activities SWT

Tue 20:29

Spoon - fact\_consumption

File Edit View Action Tools Help

View Design

dim\_product /home/aid/Downloads/ fact\_order load\_dw Project1 Exercise3 dim\_location /home/aid/Documents/ dim\_voltage fact\_consumption Job\_Exercise6

Connect

Examine preview data

Rows of step: Filter rows (100 rows)

	Year	Month	Date	District	Municipality	parish	Voltage level	Active Energy (kWh)	DistrictMunicipalityParishCode
1	2020	11	2020-11	GUARDA	Guarda	ARRIFANA	Muito Alta, Alta e Média Tensões	57149.7	090706
2	2020	11	2020-11	PORTO	Vila do Conde	UF MALTA E CANIDELO	Baixa Tensão	303181.9	131633
3	2020	11	2020-11	SETUBAL	Alcácer do Sal	UF ALCACER SAL E SANTA SUSANA	Muito Alta, Alta e Média Tensões	2994085.1	150107
4	2020	12	2020-12	AVEIRO	Anadia	UF ARCASTRO E MOGOFORES	Baixa Tensão	1599645.3	010317
5	2020	12	2020-12	BRAGA	Barcelos	FRAGOSO	Muito Alta, Alta e Média Tensões	18873.2	030235
6	2020	12	2020-12	LISBOA	Sintra	UF SINTRA	Baixa Tensão	9924017.6	111128
7	2020	12	2020-12	PORTALEGRE	Alter do Chão	CUNHEIRA	Muito Alta, Alta e Média Tensões	17042.9	120104
8	2020	12	2020-12	VIANA DO CASTELO	Valença	BOIVAO	Baixa Tensão	22110.5	160802
9	2021	1	2021-01	AVEIRO	Aveiro	OLIVEIRINHA	Baixa Tensão	859496.3	010508
10	2021	1	2021-01	BRAGANCA	Torre de Moncorvo	UF ADEGANHA E CARDANHA	Baixa Tensão	88215.5	040918
11	2021	1	2021-01	FARO	Albufeira	PADERNE	Baixa Tensão	1025357.6	080103
12	2021	2	2021-02	AVEIRO	Arouca	SAO MIGUEL DO MATO	Baixa Tensão	74644.6	010417
13	2021	2	2021-02	GUARDA	Aguilar da Beira	UF SOUTO AGUIAR BEIRA VALVERDE	Baixa Tensão	93260.4	090116
14	2021	2	2021-02	LISBOA	Torres Vedras	UF DOIS PORTOS E RUNA	Muito Alta, Alta e Média Tensões	297381.6	111324
15	2021	2	2021-02	PORTO	Amarante	SALVADOR DO MONTE	Baixa Tensão	118804.7	130129
16	2021	2	2021-02	PORTO	Penafiel	CROCA	Muito Alta, Alta e Média Tensões	43728.2	131108
17	2021	3	2021-03	COIMBRA	Góis	GOIS	Baixa Tensão	419797.5	060604
18	2021	3	2021-03	LISBOA	Alenquer	OLHALVO	Baixa Tensão	274569.1	110108
19	2021	3	2021-03	SANTAREM	Torres Novas	UF BROQUEIRA P Igreja Alcoroch	Baixa Tensão	426617.5	141918
20	2021	3	2021-03	VILA REAL	Alijó	VILA CHA	Muito Alta, Alta e Média Tensões	125764.7	170116
21	2021	4	2021-04	AVEIRO	Ilhavo	ILHAVO (SAO SALVADOR)	Muito Alta, Alta e Média Tensões	3817408.5	011008
22	2021	4	2021-04	BEJA	Aljustrel	ERVIDEL	Baixa Tensão	135291.7	020102
23	2021	4	2021-04	BEJA	Almodôvar	SAO BARNABE	Baixa Tensão	76417.8	020206
24	2021	4	2021-04	BEJA	Beja	BALEIZAO	Muito Alta, Alta e Média Tensões	673607.5	020503

Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Preview data

Close Stop Get more rows

# Select Values- Configuration (1st part)

Activities SWT

Tue 20:21

Spoon - fact\_consumption

File Edit View Action Tools Help

View Design

dim\_product /home/aid/Downloads/ fact\_order load\_dw Project1 Exercise3 dim\_location /home/aid/Documents/ dim\_voltage fact\_consumption Job\_Exercise6

Connect

loo X

Input Output Streaming Transform Utility Flow Scripting Pentaho Server Lookup Joins Data Warehouse Validation Statistics Big Data Cryptography Job Mapping Bulk loading Inline Experimental Deprecated History

CSV file input

Filter rows Select values Database lookup\_voltage Database lookup\_time Database lookup\_location Insert / update

Select values

Step name **Select values**

Select & Alter Remove Meta-data

Fields :

Fieldname	Rename to	Length	Precision	Get fields to select
1 Year				Edit Mapping
2 Month				
3 Date				
4 District				
5 Municipality				
6 parish				
7 Voltage level				
8 Active Energy (kWh)				
9 DistrictMunicipalityParishCode				

Include unspecified fields, ordered |

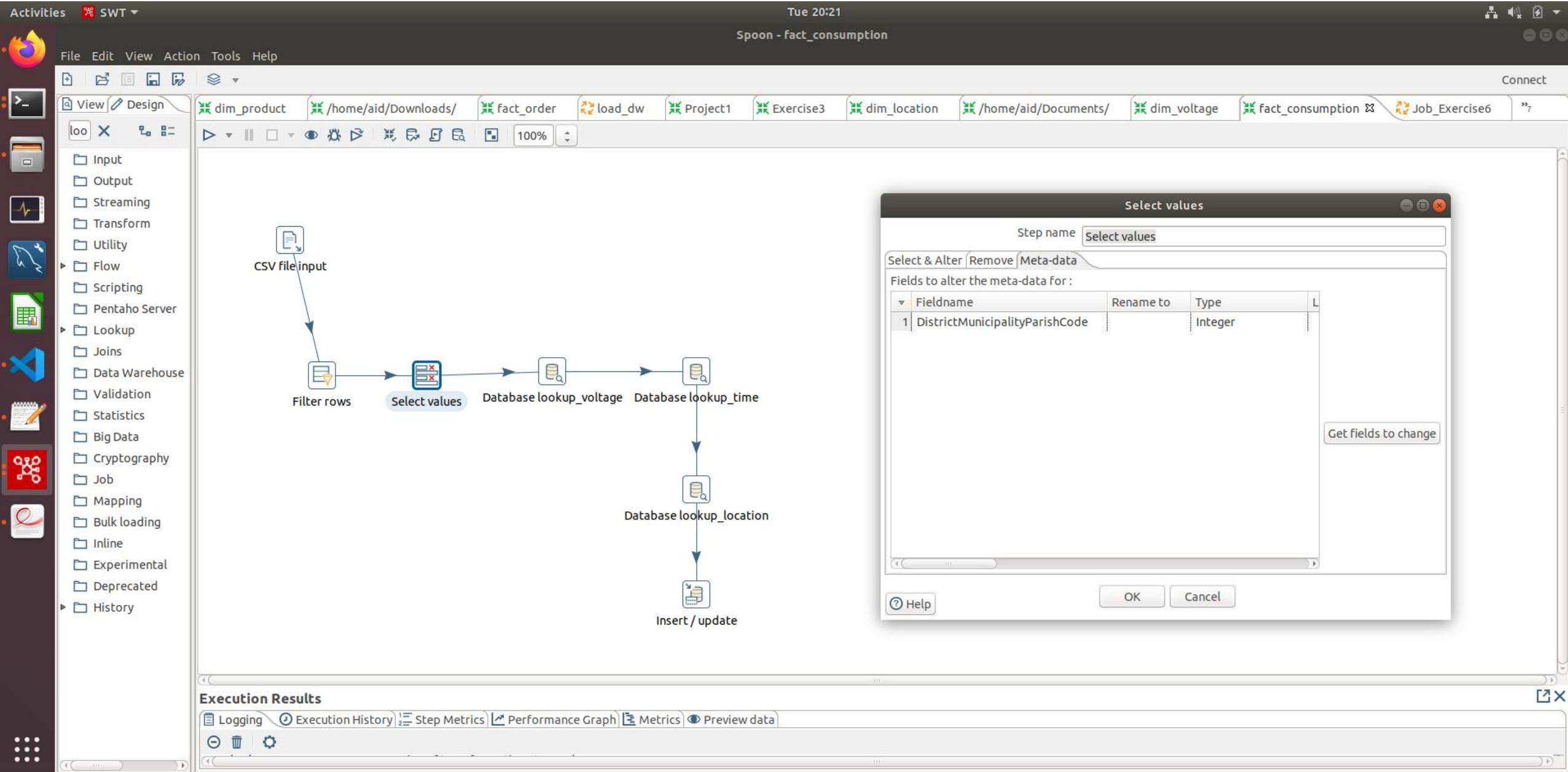
OK Cancel Help

Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Preview data

Cancel

## Select Values- Configuration (2nd part – Transforming Parish Code into integers)



# Select Values- Preview

Activities SWT

Tue 20:30

Spoon - fact\_consumption

File Edit View Action Tools Help

View Design

dim\_product /home/aid/Downloads/ fact\_order load\_dw Project1 Exercise3 dim\_location /home/aid/Documents/ dim\_voltage fact\_consumption Job\_Exercise6

Connect

loo X

Input Output Streaming Transform Utility Flow Scripting Pentaho Server Lookup Joins Data Warehouse Validation Statistics Big Data Cryptography Job Mapping Bulk loading Inline Experimental Deprecated History

Examine preview data

Rows of step: Select values (100 rows)

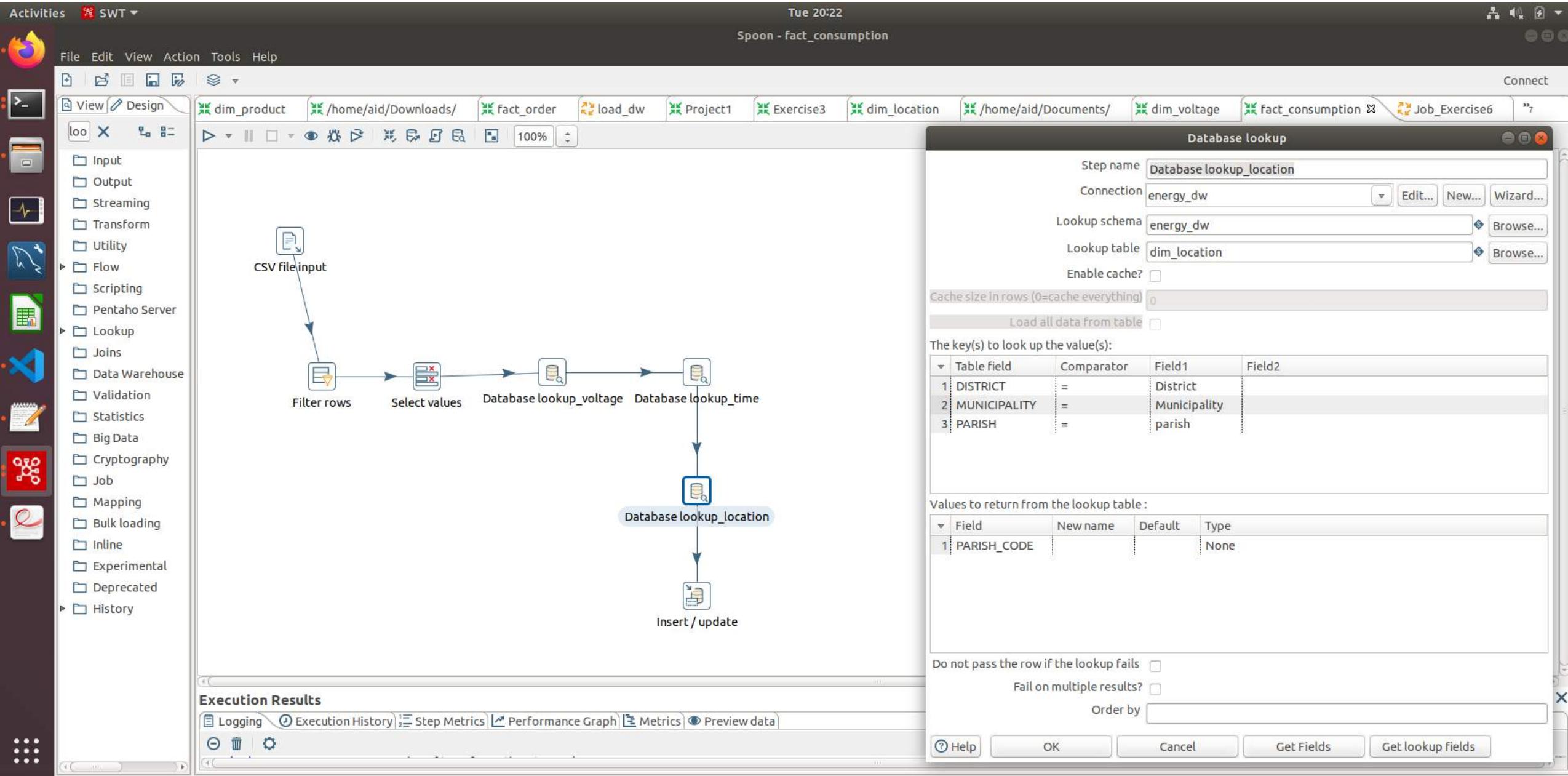
	Year	Month	Date	District	Municipality	parish	Voltage level	Active Energy (kwh)	DistrictMunicipalityParishCode
1	2020	11	2020-11	GUARDA	Guarda	ARRIFANA	Muito Alta, Alta e Média Tensões	57149.7	90706
2	2020	11	2020-11	PORUTO	Vila do Conde	UF MALTA E CANIDELO	Baixa Tensão	303181.9	131633
3	2020	11	2020-11	SETUBAL	Alcácer do Sal	UF ALCACER SAL E SANTA SUSANA	Muito Alta, Alta e Média Tensões	2994085.1	150107
4	2020	12	2020-12	AVEIRO	Anadia	UF ARCOS E MOGOFORES	Baixa Tensão	1599645.3	10317
5	2020	12	2020-12	BRAGA	Barcelos	FRAGOSO	Muito Alta, Alta e Média Tensões	18873.2	30235
6	2020	12	2020-12	LISBOA	Sintra	UF SINTRA	Baixa Tensão	9924017.6	111128
7	2020	12	2020-12	PORTALEGRE	Alter do Chão	CUNHEIRA	Muito Alta, Alta e Média Tensões	17042.9	120104
8	2020	12	2020-12	VIANA DO CASTELO	Valença	BOIVAO	Baixa Tensão	22110.5	160802
9	2021	1	2021-01	AVEIRO	Aveiro	OLIVEIRINHA	Baixa Tensão	859496.3	10508
10	2021	1	2021-01	BRAGANCA	Torre de Moncorvo	UF ADEGANHA E CARDANHA	Baixa Tensão	88215.5	40918
11	2021	1	2021-01	FARO	Albufeira	PADERNE	Baixa Tensão	1025357.6	80103
12	2021	2	2021-02	AVEIRO	Arouca	SAO MIGUEL DO MATO	Baixa Tensão	74644.6	10417
13	2021	2	2021-02	GUARDA	Aguiar da Beira	UF SOUTO AGUIAR BEIRA VALVERDE	Baixa Tensão	93260.4	90116
14	2021	2	2021-02	LISBOA	Torres Vedras	UF DOIS PORTOS E RUNA	Muito Alta, Alta e Média Tensões	297381.6	111324
15	2021	2	2021-02	PORUTO	Amarante	SALVADOR DO MONTE	Baixa Tensão	118804.7	130129
16	2021	2	2021-02	PORUTO	Penafiel	CROCA	Muito Alta, Alta e Média Tensões	43728.2	131108
17	2021	3	2021-03	COIMBRA	Góis	GOIS	Baixa Tensão	419797.5	60604
18	2021	3	2021-03	LISBOA	Alenquer	OLHALVO	Baixa Tensão	274569.1	110108
19	2021	3	2021-03	SANTAREM	Torres Novas	UF BROQUEIRA P IGREJA ALCOROCHE	Baixa Tensão	426617.5	141918
20	2021	3	2021-03	VILA REAL	Alijó	VILA CHA	Muito Alta, Alta e Média Tensões	125764.7	170116
21	2021	4	2021-04	AVEIRO	Ílhavo	ILHAZO (SAO SALVADOR)	Muito Alta, Alta e Média Tensões	3817408.5	11008
22	2021	4	2021-04	BEJA	Aljustrel	ERVIDEL	Baixa Tensão	135291.7	20102
23	2021	4	2021-04	BEJA	Almodôvar	SAO BARNABE	Baixa Tensão	76417.8	20206
24	2021	4	2021-04	BEJA	Beja	BALEIZAO	Muito Alta, Alta e Média Tensões	972687.5	20502

Close Stop Get more rows

Logging Execution History Step Metrics Performance Graph Metrics Preview data

Cancel Delete Settings

# Database Lookup Location - Configuration



# Database Lookup Location - Preview

Activities Tue 20:33

Spoon - fact\_consumption

File Edit View Action Tools Help

View Design dim\_product /home/aid/Downloads/ fact\_order load\_dw Project1 Exercise3 dim\_location /home/aid/Documents/ dim\_voltage fact\_consumption Job\_Exercise6 ?

Connect

Examine preview data

Rows of step: Database lookup\_location (100 rows)

	Year	Month	Date	District	Municipality	parish	Voltage level	Active Energy (kWh)	DistrictMunicipalityParishCode	VOLTAGE_ID	DATE_ID	MONTH_NAME	PARISH_CODE
1	2020	11	2020-11	GUARDA	Guarda	ARRIFANA	Muito Alta, Alta e Média Tensões	57149.7	90706	1	2020-11	Nov	90706
2	2020	11	2020-11	PORTO	Vila do Conde	UF MALTA E CANIDELO	Baixa Tensão	303181.9	131633	2	2020-11	Nov	131633
3	2020	11	2020-11	SETUBAL	Alcácer do Sal	UF ALCACER SAL E SANTA SUSANA	Muito Alta, Alta e Média Tensões	2994085.1	150107	1	2020-11	Nov	150107
4	2020	12	2020-12	AVEIRO	Anadia	UF ARCOS E MOGOFORES	Baixa Tensão	1599645.3	10317	2	2020-12	Dec	10317
5	2020	12	2020-12	BRAGA	Barcelos	FRAGOSO	Muito Alta, Alta e Média Tensões	18873.2	30235	1	2020-12	Dec	30235
6	2020	12	2020-12	LISBOA	Sintra	UF SINTRA	Baixa Tensão	9924017.6	111128	2	2020-12	Dec	111128
7	2020	12	2020-12	PORALEGRE	Alter do Chão	CUNHEIRA	Muito Alta, Alta e Média Tensões	17042.9	120104	1	2020-12	Dec	120104
8	2020	12	2020-12	VIANA DO CASTELO	Valença	BOIVAO	Baixa Tensão	22110.5	160802	2	2020-12	Dec	160802
9	2021	1	2021-01	AVEIRO	Aveiro	OLIVEIRINHA	Baixa Tensão	859496.3	10508	2	2021-01	Jan	10508
10	2021	1	2021-01	BRAGANCA	Torre de Moncorvo	UF ADEGANHA E CARDANHA	Baixa Tensão	88215.5	40918	2	2021-01	Jan	40918
11	2021	1	2021-01	FARO	Albufeira	PADERNE	Baixa Tensão	1025357.6	80103	2	2021-01	Jan	80103
12	2021	2	2021-02	AVEIRO	Arouca	SAO MIGUEL DO MATO	Baixa Tensão	74644.6	10417	2	2021-02	Feb	10417
13	2021	2	2021-02	GUARDA	Aguiar da Beira	UF SOUTO AGUIAR BEIRA VALVERDE	Baixa Tensão	93260.4	90116	2	2021-02	Feb	90116
14	2021	2	2021-02	LISBOA	Torres Vedras	UF DOIS PORTOS E RUNA	Muito Alta, Alta e Média Tensões	297381.6	111324	1	2021-02	Feb	111324
15	2021	2	2021-02	PORTO	Amarante	SAVADOR DO MONTE	Baixa Tensão	118804.7	130129	2	2021-02	Feb	130129
16	2021	2	2021-02	PORTO	Penafiel	CROCA	Muito Alta, Alta e Média Tensões	43728.2	131108	1	2021-02	Feb	131108
17	2021	3	2021-03	COIMBRA	Góis	GOIS	Baixa Tensão	419797.5	60604	2	2021-03	Mar	60604
18	2021	3	2021-03	LISBOA	Alenquer	OLHALVO	Baixa Tensão	274569.1	110108	2	2021-03	Mar	110108
19	2021	3	2021-03	SANTAREM	Torres Novas	UF BROQUEIRA P IGREJA ALCOROCHE	Baixa Tensão	426617.5	141918	2	2021-03	Mar	141918
20	2021	3	2021-03	VILA REAL	Alijó	VILA CHA	Muito Alta, Alta e Média Tensões	125764.7	170116	1	2021-03	Mar	170116
21	2021	4	2021-04	AVEIRO	ílhavo	ILHAZO (SAO SALVADOR)	Muito Alta, Alta e Média Tensões	3817408.5	11008	1	2021-04	Apr	11008
22	2021	4	2021-04	BEJA	Aljustrel	ERVIDEL	Baixa Tensão	135291.7	20102	2	2021-04	Apr	20102
23	2021	4	2021-04	BEJA	Almodôvar	SAO BARNABE	Baixa Tensão	76417.8	20206	2	2021-04	Apr	20206
24	2021	4	2021-04	BEJA	Beja	BALEIZAO	Muito Alta, Alta e Média Tensões	972687.5	20502	1	2021-04	Apr	20502

[Close](#) [Stop](#) [Get more rows](#)

Logging Execution History Step Metrics Performance Graph Metrics Preview data

# Database Lookup Time - Configuration

Activities SWT Tue 20:22 Spoon - fact\_consumption

File Edit View Action Tools Help Connect

View Design

dim\_product /home/aid/Downloads/ fact\_order load\_dw Project1 Exercise3 dim\_location /home/aid/Documents/ dim\_voltage fact\_consumption Job\_Exercise6

loo X

Input Output Streaming Transform Utility Flow Scripting Pentaho Server Lookup Joins Data Warehouse Validation Statistics Big Data Cryptography Job Mapping Bulk loading Inline Experimental Deprecated History

CSV file input

Filter rows Select values Database lookup\_voltage Database lookup\_time

Database lookup\_location

Insert / update

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

Database lookup

Step name: Database lookup\_time  
Connection: energy\_dw  
Lookup schema: energy\_dw  
Lookup table: dim\_time  
Enable cache?   
Cache size in rows (0=cache everything) 0  
Load all data from table   
The key(s) to look up the value(s):

Table field	Comparator	Field1	Field2
1 YEAR	=	Year	
2 MONTH	=	Month	

Values to return from the lookup table:

Field	New name	Default	Type
1 DATE_ID			None
2 MONTH_NAME			None

Do not pass the row if the lookup fails   
Fail on multiple results?   
Order by

Help OK Cancel Get Fields Get lookup fields

```
graph LR; CSV[CSV file input] --> Filter[Filter rows]; Filter --> Select[Select values]; Select --> Voltage[Database lookup_voltage]; Voltage --> Time[Database lookup_time]; Time --> Location[Database lookup_location]; Location --> Insert[Insert / update]
```

# Database Lookup Time - Preview

Activities Tue 20:31

Spoon - fact\_consumption

File Edit View Action Tools Help Connect

View Design dim\_product /home/aid/Downloads/ fact\_order load\_dw Project1 Exercise3 dim\_location /home/aid/Documents/ dim\_voltage fact\_consumption Job\_Exercise6 ?

loo X 100% Examine preview data

Rows of step: Database lookup\_time (100 rows)

	Year	Month	Date	District	Municipality	parish	Voltage level	Active Energy (kWh)	DistrictMunicipalityParishCode	VOLTAGE_ID	DATE_ID	MONTH_NAME
1	2020	11	2020-11	GUARDA	Guarda	ARRIFANA	Muito Alta, Alta e Média Tensões	57149.7	90706	1	2020-11	Nov
2	2020	11	2020-11	PORTO	Vila do Conde	UF MALTA E CANIDELO	Baixa Tensão	303181.9	131633	2	2020-11	Nov
3	2020	11	2020-11	SETUBAL	Alcácer do Sal	UF ALCACER SAL E SANTA SUSANA	Muito Alta, Alta e Média Tensões	2994085.1	150107	1	2020-11	Nov
4	2020	12	2020-12	AVEIRO	Anadia	UF ARCOS E MOGOFORES	Baixa Tensão	1599645.3	10317	2	2020-12	Dec
5	2020	12	2020-12	BRAGA	Barcelos	FRAGOSO	Muito Alta, Alta e Média Tensões	18873.2	30235	1	2020-12	Dec
6	2020	12	2020-12	LISBOA	Sintra	UF SINTRA	Baixa Tensão	9924017.6	111128	2	2020-12	Dec
7	2020	12	2020-12	PORCALEGRE	Alter do Chão	CUNHEIRA	Muito Alta, Alta e Média Tensões	17042.9	120104	1	2020-12	Dec
8	2020	12	2020-12	VIANA DO CASTELO	Valença	BOIVAO	Baixa Tensão	22110.5	160802	2	2020-12	Dec
9	2021	1	2021-01	AVEIRO	Aveiro	OLIVEIRINHA	Baixa Tensão	859496.3	10508	2	2021-01	Jan
10	2021	1	2021-01	BRAGANCA	Torre de Moncorvo	UF ADEGANHA E CARDANHA	Baixa Tensão	88215.5	40918	2	2021-01	Jan
11	2021	1	2021-01	FARO	Albufeira	PADERNE	Baixa Tensão	1025357.6	80103	2	2021-01	Jan
12	2021	2	2021-02	AVEIRO	Arouca	SAO MIGUEL DO MATO	Baixa Tensão	74644.6	10417	2	2021-02	Feb
13	2021	2	2021-02	GUARDA	Aguiar da Beira	UF SOUTO AGUIAR BEIRA VALVERDE	Baixa Tensão	93260.4	90116	2	2021-02	Feb
14	2021	2	2021-02	LISBOA	Torres Vedras	UF DOIS PORTOS E RUNA	Muito Alta, Alta e Média Tensões	297381.6	111324	1	2021-02	Feb
15	2021	2	2021-02	PORTO	Amarante	SALVADOR DO MONTE	Baixa Tensão	118804.7	130129	2	2021-02	Feb
16	2021	2	2021-02	PORTO	Penafiel	CROCA	Muito Alta, Alta e Média Tensões	43728.2	131108	1	2021-02	Feb
17	2021	3	2021-03	COIMBRA	Góis	GOIS	Baixa Tensão	419797.5	60604	2	2021-03	Mar
18	2021	3	2021-03	LISBOA	Alenquer	OLHALVO	Baixa Tensão	274569.1	110108	2	2021-03	Mar
19	2021	3	2021-03	SANTAREM	Torres Novas	UF BROQUEIRA P IGREJA ALCOROCHE	Baixa Tensão	426617.5	141918	2	2021-03	Mar
20	2021	3	2021-03	VILA REAL	Alijó	VILA CHA	Muito Alta, Alta e Média Tensões	125764.7	170116	1	2021-03	Mar
21	2021	4	2021-04	AVEIRO	Ilhavo	ILHAVO (SAO SALVADOR)	Muito Alta, Alta e Média Tensões	3817408.5	11008	1	2021-04	Apr
22	2021	4	2021-04	BEJA	Aljustrel	ERVIDEL	Baixa Tensão	135291.7	20102	2	2021-04	Apr
23	2021	4	2021-04	BEJA	Almodôvar	SAO BARNABE	Baixa Tensão	76417.8	20206	2	2021-04	Apr
24	2021	4	2021-04	BEJA	Beja	BALEIZAO	Muito Alta, Alta e Média Tensões	972687.5	20502	1	2021-04	Apr

[Close](#) [Stop](#) [Get more rows](#)

Logging Execution History Step Metrics Performance Graph Metrics Preview data

Activities SWT Tue 20:31

# Database Lookup Voltage - Configuration

Activities SWT Tue 20:22 Spoon - fact\_consumption

File Edit View Action Tools Help

View Design

dim\_product /home/aid/Downloads/ fact\_order load\_dw Project1 Exercise3 dim\_location /home/aid/Documents/ dim\_voltage fact\_consumption Job\_Exercise6

CSV file input

Filter rows Select values Database lookup\_voltage Database lookup\_time Database lookup\_location Insert / update

Database lookup

Step name: Database lookup\_voltage  
Connection: energy\_dw  
Lookup schema: energy\_dw  
Lookup table: dim\_voltage  
Enable cache?   
Cache size in rows (0=cache everything) 0  
Load all data from table

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

Table field	Comparator	Field1	Field2
1 VOLTAGE_LEVEL	=	Voltage level	

Values to return from the lookup table:

Field	New name	Default	Type
1 VOLTAGE_ID			None

Do not pass the row if the lookup fails   
Fail on multiple results?   
Order by   
Help OK Cancel Get Fields Get lookup fields

Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Preview data

Activities SWT Tue 20:22 Spoon - fact\_consumption

File Edit View Action Tools Help

View Design

dim\_product /home/aid/Downloads/ fact\_order load\_dw Project1 Exercise3 dim\_location /home/aid/Documents/ dim\_voltage fact\_consumption Job\_Exercise6

CSV file input

Filter rows Select values Database lookup\_voltage Database lookup\_time Database lookup\_location Insert / update

Database lookup

Step name: Database lookup\_voltage  
Connection: energy\_dw  
Lookup schema: energy\_dw  
Lookup table: dim\_voltage  
Enable cache?   
Cache size in rows (0=cache everything) 0  
Load all data from table

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

Table field	Comparator	Field1	Field2
1 VOLTAGE_LEVEL	=	Voltage level	

Values to return from the lookup table:

Field	New name	Default	Type
1 VOLTAGE_ID			None

Do not pass the row if the lookup fails   
Fail on multiple results?   
Order by   
Help OK Cancel Get Fields Get lookup fields

Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Preview data

# Database Lookup Voltage - Preview

Activities SWT

Tue 20:30

Spoon - fact\_consumption

File Edit View Action Tools Help

View Design dim\_product /home/aid/Downloads/ fact\_order load\_dw Project1 Exercise3 dim\_location /home/aid/Documents/ dim\_voltage fact\_consumption Job\_Exercise6 ?

Connect

loo X

Input Output Streaming Transform Utility Flow Scripting Pentaho Server Lookup Joins Data Warehouse Validation Statistics Big Data Cryptography Job Mapping Bulk loading Inline Experimental Deprecated History

Examine preview data

Rows of step: Database lookup\_voltage (100 rows)

	Year	Month	Date	District	Municipality	parish	Voltage level	Active Energy (kWh)	DistrictMunicipalityParishCode	VOLTAGE_ID
1	2020	11	2020-11	GUARDA	Guarda	ARRIFANA	Muito Alta, Alta e Média Tensões	57149.7	90706	1
2	2020	11	2020-11	PORTO	Vila do Conde	UF MALTA E CANIDELO	Baixa Tensão	303181.9	131633	2
3	2020	11	2020-11	SETUBAL	Alcácer do Sal	UF ALCACER SAL E SANTA SUSANA	Muito Alta, Alta e Média Tensões	2994085.1	150107	1
4	2020	12	2020-12	AVEIRO	Anadia	UF ARCASTRO E MOGOFORES	Baixa Tensão	1599645.3	10317	2
5	2020	12	2020-12	BRAGA	Barcelos	FRAGOSO	Muito Alta, Alta e Média Tensões	18873.2	30235	1
6	2020	12	2020-12	LISBOA	Sintra	UF SINTRA	Baixa Tensão	9924017.6	111128	2
7	2020	12	2020-12	PORTALEGRE	Alter do Chão	CUNHEIRA	Muito Alta, Alta e Média Tensões	17042.9	120104	1
8	2020	12	2020-12	VIANA DO CASTELO	Valença	BOIVAO	Baixa Tensão	22110.5	160802	2
9	2021	1	2021-01	AVEIRO	Aveiro	OLIVEIRINHA	Baixa Tensão	859496.3	10508	2
10	2021	1	2021-01	BRAGANCA	Torre de Moncorvo	UF ADEGANHA E CARDANHA	Baixa Tensão	88215.5	40918	2
11	2021	1	2021-01	FARO	Albufeira	PADERNE	Baixa Tensão	1025357.6	80103	2
12	2021	2	2021-02	AVEIRO	Arouca	SAO MIGUEL DO MATO	Baixa Tensão	74644.6	10417	2
13	2021	2	2021-02	GUARDA	Aguiar da Beira	UF SOUTO AGUIAR BEIRA VALVERDE	Baixa Tensão	93260.4	90116	2
14	2021	2	2021-02	LISBOA	Torres Vedras	UF DOIS PORTOS E RUNA	Muito Alta, Alta e Média Tensões	297381.6	111324	1
15	2021	2	2021-02	PORTO	Amarante	SALVADOR DO MONTE	Baixa Tensão	118804.7	130129	2
16	2021	2	2021-02	PORTO	Penafiel	CROCA	Muito Alta, Alta e Média Tensões	43728.2	131108	1
17	2021	3	2021-03	COIMBRA	Góis	GOIS	Baixa Tensão	419797.5	60604	2
18	2021	3	2021-03	LISBOA	Alenquer	OLHALVO	Baixa Tensão	274569.1	110108	2
19	2021	3	2021-03	SANTAREM	Torres Novas	UF BROQUEIRA P IGREJA ALCOROCHE	Baixa Tensão	426617.5	141918	2
20	2021	3	2021-03	VILA REAL	Alijó	VILA CHA	Muito Alta, Alta e Média Tensões	125764.7	170116	1
21	2021	4	2021-04	AVEIRO	Ilhavo	ILHAVO (SAO SALVADOR)	Muito Alta, Alta e Média Tensões	3817408.5	11008	1
22	2021	4	2021-04	BEJA	Aljustrel	ERVIDEL	Baixa Tensão	135291.7	20102	2
23	2021	4	2021-04	BEJA	Almodôvar	SAO BARNABE	Baixa Tensão	76417.8	20206	2
24	2021	4	2021-04	BEJA	Beja	BALEIZAO	Muito Alta, Alta e Média Tensões	972687.5	20502	1

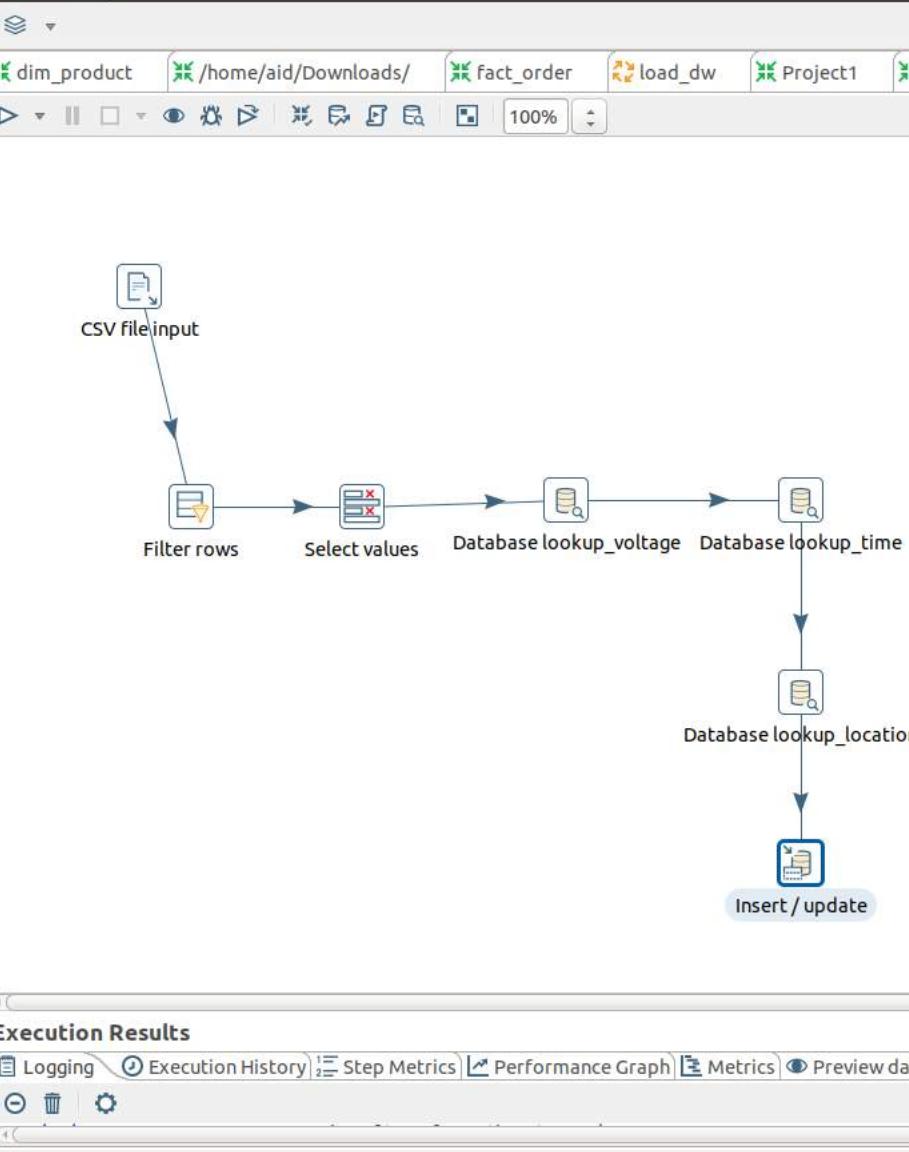
Logging Execution History Step Metrics Performance Graph Metrics Preview data

## Insert/Update - Configuration

Activities  SWT ▾



The screenshot shows the Pentaho Data Integration (Kettle) interface. On the left, there's a vertical toolbar with icons for various tools like File, View, Design, etc. Below it is a palette containing icons for Input, Output, Streaming, Transform, Utility, Flow, Scripting, Pentaho Server, Lookup, Joins, Data Warehouse, Validation, Statistics, Big Data, Cryptography, Job, Mapping, Bulk loading, Inline, Experimental, Deprecated, and History. On the right, there's a large text editor window with some Kettle script code.



**Insert / update**

Step name	Insert / update																											
Connection	energy_dw	<input type="button" value="Edit..."/>	<input type="button" value="New..."/>																									
Target schema	energy_dw <input type="button" value="Browse..."/>																											
Target table	fact_consumption <input type="button" value="Browse..."/>																											
Commit size	100 <input type="button" value=""/>																											
Don't perform any updates: <input type="checkbox"/>																												
The key(s) to look up the value(s):																												
<table border="1"> <thead> <tr> <th>▼</th> <th>Table field</th> <th>Comparator</th> <th>Stream field1</th> <th>Stream field2</th> <th><input type="button" value="Get fields"/></th> </tr> </thead> <tbody> <tr> <td>1</td> <td>CONSUMPTION_ID</td> <td>=</td> <td>Active Energy (kwh)</td> <td></td> <td></td> </tr> </tbody> </table>				▼	Table field	Comparator	Stream field1	Stream field2	<input type="button" value="Get fields"/>	1	CONSUMPTION_ID	=	Active Energy (kwh)															
▼	Table field	Comparator	Stream field1	Stream field2	<input type="button" value="Get fields"/>																							
1	CONSUMPTION_ID	=	Active Energy (kwh)																									
Update fields:																												
<table border="1"> <thead> <tr> <th>▼</th> <th>Table field</th> <th>Stream field</th> <th>Update</th> <th><input type="button" value="Get update fields"/></th> </tr> </thead> <tbody> <tr> <td>1</td> <td>ACTIVE_ENERGY</td> <td>Active Energy (kwh)</td> <td>Y</td> <td></td> </tr> <tr> <td>2</td> <td>PARISH_CODE</td> <td>PARISH_CODE</td> <td>Y</td> <td></td> </tr> <tr> <td>3</td> <td>VOLTAGE_ID</td> <td>VOLTAGE_ID</td> <td>Y</td> <td></td> </tr> <tr> <td>4</td> <td>DATE_ID</td> <td>DATE_ID</td> <td>Y</td> <td></td> </tr> </tbody> </table>				▼	Table field	Stream field	Update	<input type="button" value="Get update fields"/>	1	ACTIVE_ENERGY	Active Energy (kwh)	Y		2	PARISH_CODE	PARISH_CODE	Y		3	VOLTAGE_ID	VOLTAGE_ID	Y		4	DATE_ID	DATE_ID	Y	
▼	Table field	Stream field	Update	<input type="button" value="Get update fields"/>																								
1	ACTIVE_ENERGY	Active Energy (kwh)	Y																									
2	PARISH_CODE	PARISH_CODE	Y																									
3	VOLTAGE_ID	VOLTAGE_ID	Y																									
4	DATE_ID	DATE_ID	Y																									

# Insert/Update - Preview

Activities SWT

Tue 20:34

Spoon - fact\_consumption

File Edit View Action Tools Help

View Design dim\_product /home/aid/Downloads/ fact\_order load\_dw Project1 Exercise3 dim\_location /home/aid/Documents/ dim\_voltage fact\_consumption Job\_Exercise6 ?

Connect

Examine preview data

Rows of step: Insert / update (100 rows)

	Year	Month	Date	District	Municipality	parish	Voltage level	Active Energy (kwh)	DistrictMunicipalityParishCode	VOLTAGE_ID	DATE_ID	MONTH_NAME	PARISH_CODE
1	2020	11	2020-11	GUARDA	Guarda	ARRIFANA	Muito Alta, Alta e Média Tensões	57149.7	90706	1	2020-11	Nov	90706
2	2020	11	2020-11	PORTO	Vila do Conde	UF MALTA E CANIDELO	Baixa Tensão	303181.9	131633	2	2020-11	Nov	131633
3	2020	11	2020-11	SETUBAL	Alcácer do Sal	UF ALCACER SAL E SANTA SUSANA	Muito Alta, Alta e Média Tensões	2994085.1	150107	1	2020-11	Nov	150107
4	2020	12	2020-12	AVEIRO	Anadia	UF ARCOS E MOGOFORES	Baixa Tensão	1599645.3	10317	2	2020-12	Dec	10317
5	2020	12	2020-12	BRAGA	Barcelos	FRAGOSO	Muito Alta, Alta e Média Tensões	18873.2	30235	1	2020-12	Dec	30235
6	2020	12	2020-12	LISBOA	Sintra	UF SINTRA	Baixa Tensão	9924017.6	111128	2	2020-12	Dec	111128
7	2020	12	2020-12	PORTELEGRE	Alter do Chão	CUNHEIRA	Muito Alta, Alta e Média Tensões	17042.9	120104	1	2020-12	Dec	120104
8	2020	12	2020-12	VIANA DO CASTELO	Valença	BOIVAO	Baixa Tensão	22110.5	160802	2	2020-12	Dec	160802
9	2021	1	2021-01	AVEIRO	Aveiro	OLIVEIRINHA	Baixa Tensão	859496.3	10508	2	2021-01	Jan	10508
10	2021	1	2021-01	BRAGANCA	Torre de Moncorvo	UF ADEGANHA E CARDANHA	Baixa Tensão	88215.5	40918	2	2021-01	Jan	40918
11	2021	1	2021-01	FARO	Albufeira	PADERNE	Baixa Tensão	1025357.6	80103	2	2021-01	Jan	80103
12	2021	2	2021-02	AVEIRO	Arouca	SAO MIGUEL DO MATO	Baixa Tensão	74644.6	10417	2	2021-02	Feb	10417
13	2021	2	2021-02	GUARDA	Aguiar da Beira	UF SOUTO AGUIAR BEIRA VALVERDE	Baixa Tensão	93260.4	90116	2	2021-02	Feb	90116
14	2021	2	2021-02	LISBOA	Torres Vedras	UF DOIS PORTOS E RUNA	Muito Alta, Alta e Média Tensões	297381.6	111324	1	2021-02	Feb	111324
15	2021	2	2021-02	PORTO	Amarante	SAVADOR DO MONTE	Baixa Tensão	118804.7	130129	2	2021-02	Feb	130129
16	2021	2	2021-02	PORTO	Penafiel	CROCA	Muito Alta, Alta e Média Tensões	43728.2	131108	1	2021-02	Feb	131108
17	2021	3	2021-03	COIMBRA	Góis	GOIS	Baixa Tensão	419797.5	60604	2	2021-03	Mar	60604
18	2021	3	2021-03	LISBOA	Alenquer	OLHALVO	Baixa Tensão	274569.1	110108	2	2021-03	Mar	110108
19	2021	3	2021-03	SANTAREM	Torres Novas	UF BROQUEIRA P IGREJA ALCOROCHE	Baixa Tensão	426617.5	141918	2	2021-03	Mar	141918
20	2021	3	2021-03	VILA REAL	Alijó	VILA CHA	Muito Alta, Alta e Média Tensões	125764.7	170116	1	2021-03	Mar	170116
21	2021	4	2021-04	AVEIRO	Ilhavo	ILHAVO (SAO SALVADOR)	Muito Alta, Alta e Média Tensões	3817408.5	11008	1	2021-04	Apr	11008
22	2021	4	2021-04	BEJA	Aljustrel	ERVIDEL	Baixa Tensão	135291.7	20102	2	2021-04	Apr	20102
23	2021	4	2021-04	BEJA	Almodôvar	SAO BARNABE	Baixa Tensão	76417.8	20206	2	2021-04	Apr	20206
24	2021	4	2021-04	BEJA	Beja	BALEIZAO	Muito Alta, Alta e Média Tensões	972687.5	20502	1	2021-04	Apr	20502

[Close](#) [Stop](#) [Get more rows](#)

Logging Execution History Step Metrics Performance Graph Metrics Preview data

# Job Execution

Activities SWT Tue 20:13 aid@virtualbox: ~/Pentaho/data-integration

Spoon - Job\_Exercise6

File Edit View Action Tools Help

View Design

/home/aid/Documents/ dim\_voltage fact\_consumption Job\_Exercise6 14

Connect

database loo

- General
- Mail
- File management
- Conditions
- Scripting
- Bulk loading
- XML
- Utility
- Repository
- File transfer
- File encryption
- Deprecated

Start dim\_location dim\_voltage dim\_time fact\_consumption

Execution Results

Logging History Job metrics Metrics

2023/10/24 20:09:01 - Database lookup\_location.0 - linenr 150000  
2023/10/24 20:09:05 - Database lookup\_location.0 - Finished processing (I=150623, O=0, R=150623, W=150623, U=0, E=0)  
2023/10/24 20:10:06 - Insert / update.0 - linenr 150000  
2023/10/24 20:10:10 - Insert / update.0 - Finished processing (I=150623, O=150623, R=150623, W=150623, U=0, E=0)  
2023/10/24 20:10:10 - Job\_Exercise6 - Finished job entry [fact\_consumption] (result=[true])  
2023/10/24 20:10:10 - Job\_Exercise6 - Finished job entry [dim\_time] (result=[true])  
2023/10/24 20:10:10 - Job\_Exercise6 - Finished job entry [dim\_voltage] (result=[true])  
2023/10/24 20:10:10 - Job\_Exercise6 - Finished job entry [dim\_location] (result=[true])  
2023/10/24 20:10:10 - Job\_Exercise6 - Job execution finished  
2023/10/24 20:10:10 - Spoon - Job has ended.

aid@virtualbox: ~/Pentaho/data-integration

19:55:18 - CSV file input.0 - Header row skipped in file '/home/aid/Documents/Projeto'

Aveiro	Arouca	ESCARIZ	10409
AVEIRO	Arouca	FERMEDO	10411
AVEIRO	Arouca	MANSORES	10413
AVEIRO	Arouca	MOLDES	10414
AVEIRO	Arouca	ROSSAS	10415
AVEIRO	Arouca	SANTA EULALIA	10416
AVEIRO	Arouca	SAO MIGUEL DO MATO	10417
AVEIRO	Arouca	TROPECO	10418
AVEIRO	Arouca	URRO	10419
AVEIRO	Arouca	VARZEA	10420
AVEIRO	Arouca	UF AROUCA E BURGO	10421

+-----+-----+-----+-----+-----+-----+-----+-----+

40 rows in set (0.02 sec)

mysql> DELETE FROM dim\_location;  
Query OK, 2874 rows affected (0.15 sec)

mysql> DELETE FROM dim\_time;  
Query OK, 34 rows affected (0.09 sec)

mysql> DELETE FROM dim\_voltage;  
Query OK, 2 rows affected (0.04 sec)

mysql>

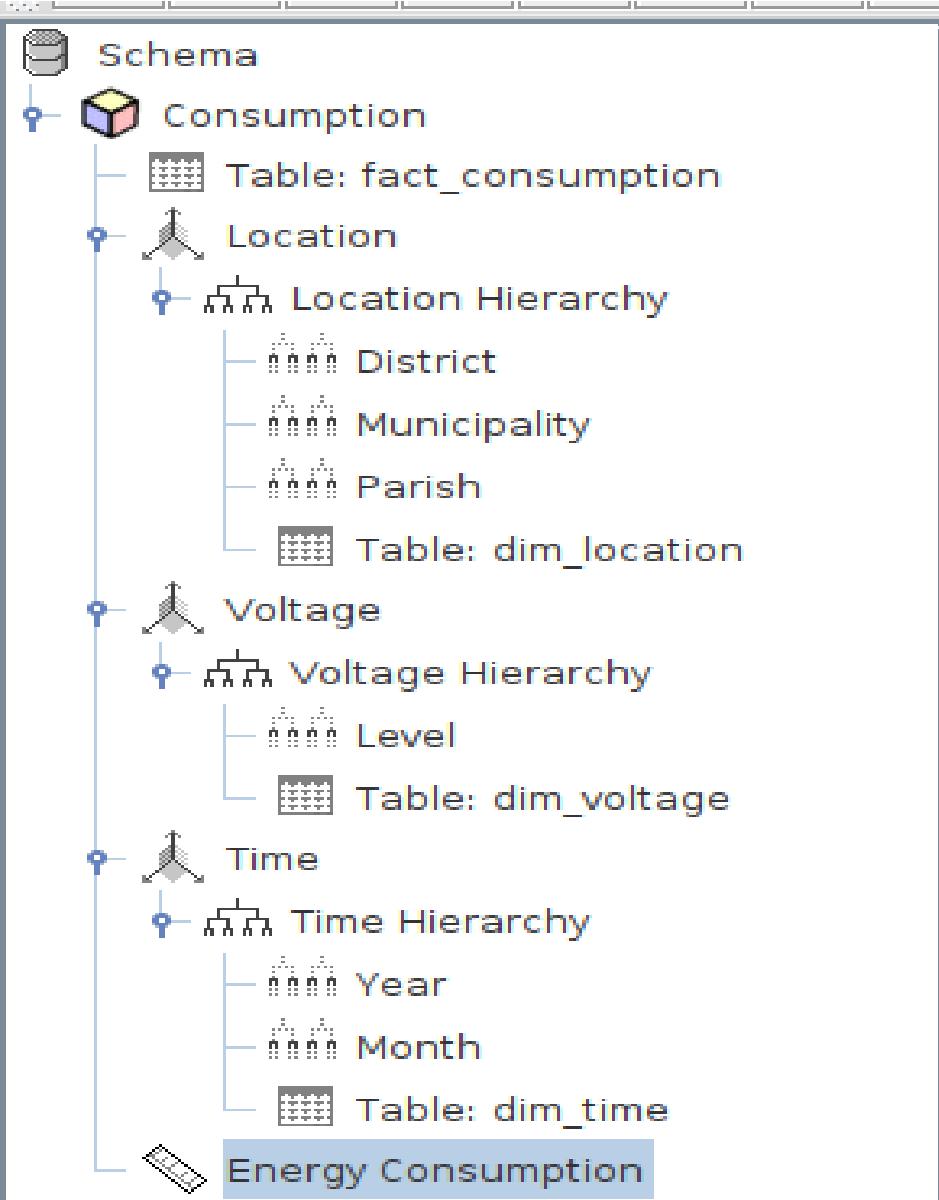
20:07:59 - Database lookup\_time.0 - linenr 150000  
20:08:02 - Database lookup\_time.0 - Finished processing (I=150623, O=0, R=150623, W=150623, U=0, E=0)  
20:09:01 - Database lookup\_location.0 - linenr 150000  
20:09:05 - Database lookup\_location.0 - Finished processing (I=150623, O=0, R=150623, W=150623, U=0, E=0)  
20:10:06 - Insert / update.0 - linenr 150000  
20:10:10 - Insert / update.0 - Finished processing (I=150623, O=150623, R=150623, W=150623, U=0, E=0)  
20:10:10 - Job\_Exercise6 - Finished job entry [fact\_consumption] (result=[true])  
20:10:10 - Job\_Exercise6 - Finished job entry [dim\_time] (result=[true])  
20:10:10 - Job\_Exercise6 - Finished job entry [dim\_voltage] (result=[true])  
20:10:10 - Job\_Exercise6 - Finished job entry [dim\_location] (result=[true])  
20:10:10 - Job\_Exercise6 - Job execution finished  
20:10:10 - Spoon - Job has ended.

# Exercise 7 – Group 50

José Cardoso

Matheus Trindade

# OLAP Cube Definition



# XML File

Activities Text Editor ▾

Thu 20:13

EX7.xml  
~/Desktop

Open Save

Save

XML ▾ Tab Width: 8 ▾ Ln 1, Col 1 ▾ INS

```
<Schema name="energy_dw">
  <Cube name="Consumption" visible="true" cache="true" enabled="true">
    <Table name="fact_consumption">
      </Table>
    <Dimension type="StandardDimension" visible="true" foreignKey="PARISH_CODE" name="Location">
      <Hierarchy name="Location Hierarchy" visible="true" hasAll="true" allMemberName="All Locations" primaryKey="PARISH_CODE">
        <Table name="dim_location">
          </Table>
        <Level name="District" visible="true" column="DISTRICT" type="String" uniqueMembers="false" levelType="Regular">
        </Level>
        <Level name="Municipality" visible="true" column="MUNICIPALITY" type="String" uniqueMembers="false" levelType="Regular">
        </Level>
        <Level name="Parish" visible="true" column="PARISH" type="String" uniqueMembers="false" levelType="Regular">
        </Level>
      </Hierarchy>
    </Dimension>
    <Dimension type="StandardDimension" visible="true" foreignKey="VOLTAGE_ID" name="Voltage">
      <Hierarchy name="Voltage Hierarchy" visible="true" hasAll="true" allMemberName="All Voltages" primaryKey="VOLTAGE_ID">
        <Table name="dim_voltage">
          </Table>
        <Level name="Level" visible="true" column="VOLTAGE_LEVEL" type="String" uniqueMembers="false" levelType="Regular">
        </Level>
      </Hierarchy>
    </Dimension>
    <Dimension type="TimeDimension" visible="true" foreignKey="DATE_ID" name="Time">
      <Hierarchy name="Time Hierarchy" visible="true" hasAll="true" allMemberName="All Years" primaryKey="DATE_ID">
        <Table name="dim_time">
          </Table>
        <Level name="Year" visible="true" column="YEAR" type="Integer" uniqueMembers="false" levelType="TimeYears">
        </Level>
        <Level name="Month" visible="true" column="MONTH_NAME" ordinalColumn="MONTH" type="String" uniqueMembers="false" levelType="TimeMonths">
        </Level>
      </Hierarchy>
    </Dimension>
    <Measure name="Energy Consumption" column="ACTIVE_ENERGY" datatype="Numeric" formatString="###.## kWh" aggregator="sum" visible="true">
    </Measure>
  </Cube>
</Schema>
```

# Exercise 8 – Group 50

José Cardoso

Matheus Trindade

# Energy Consumption by district and year

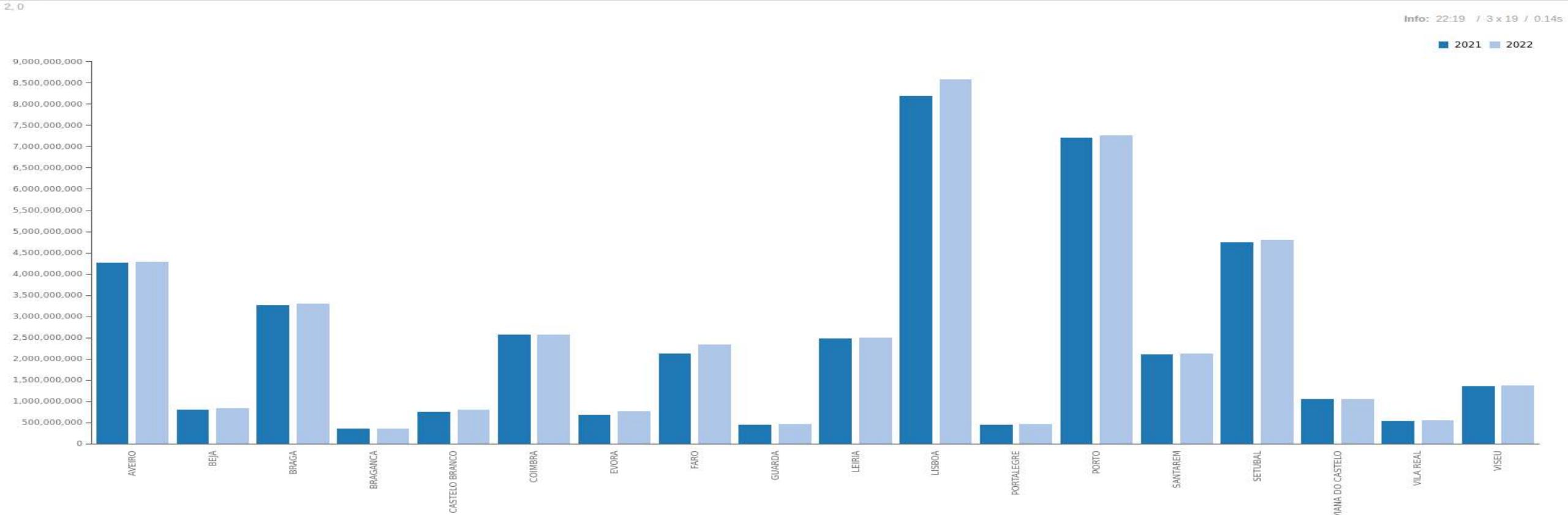
```
1 SELECT
2 FILTER(Time.Year.Members, COUNT(DESCENDANTS(Time.CurrentMember, Time.Month)) = 12) ON COLUMNS,
3 Location.District.Members ON ROWS
4 FROM Consumption
```

1, 6

District	2021	2022
AVEIRO	4272573288.19 kWh	4286627235.61 kWh
BEJA	814485369.53 kWh	851155700.74 kWh
BRAGA	3267879169.85 kWh	3305974196.46 kWh
BRAGANCA	354303873.36 kWh	365683592.35 kWh
CASTELO BRANCO	762689588.7 kWh	806863865.5 kWh
COIMBRA	2580360669.81 kWh	2573327552.58 kWh
EVORA	688852061.92 kWh	772678171.86 kWh
FARO	2123210910.91 kWh	2349456140.74 kWh
GUARDA	450255805.45 kWh	476789661.33 kWh
LEIRIA	2491604190.94 kWh	2510481629.14 kWh
LISBOA	8188862012.68 kWh	8579095335.21 kWh
PORTALEGRE	456919863.4 kWh	472541081.28 kWh
PORTO	7211118567.12 kWh	7264995596.6 kWh
SANTAREM	2118325270.23 kWh	2129435881.77 kWh
SETUBAL	4747012610.88 kWh	4804637935.36 kWh
VIANA DO CASTELO	1052203384.9 kWh	1063191957.6 kWh
VILA REAL	541544926.92 kWh	566820894.16 kWh
VISEU	1362244090.93 kWh	1379225857.4 kWh

# Analysis of Energy Consumption by district and year

```
1 SELECT
2 FILTER(Time.Year.Members, COUNT(DESCENDANTS(Time.CurrentMember, Time.Month)) = 12) ON COLUMNS,
3 Location.District.Members ON ROWS
4 FROM Consumption
```



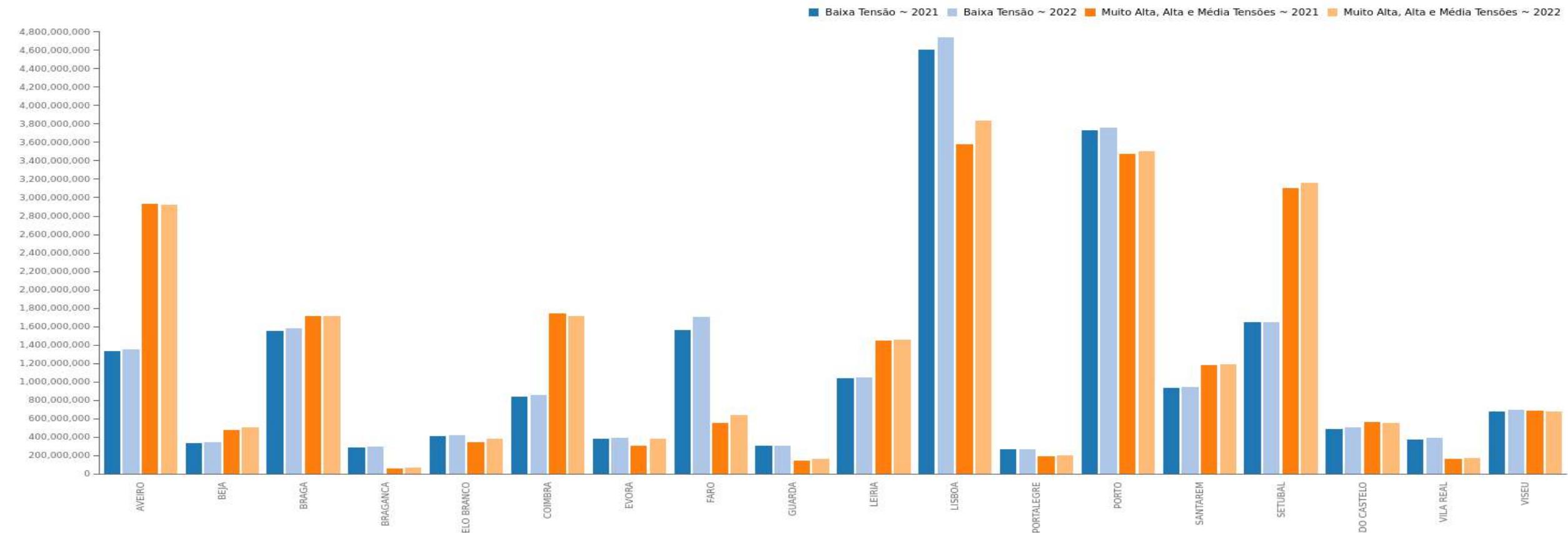
As we can observe when comparing the comsumption of every district between 2021 and 2022 we can notice that almost every single district increased except Coimbra.

# Analysis of Energy Consumption by district, year and voltage level

```
1 SELECT
2 Voltage.[Level].Members * FILTER(Time.Year.Members, COUNT(DESCENDANTS(Time.CurrentMember, Time.Month)) = 12) ON COLUMNS,
3 Location.District.Members ON ROWS
4 FROM Consumption
```

4\_16

Info: 22:20 / 5 x 20 / 0.09s



As we can observe when comparing the consumption of every district between 2021 and 2022 but this time taking into account the voltage level we can notice that every single district's low voltage consumption increased. In contrast the district's hight voltage consumption differences are mixed. Taking everything into account, we can reach the conclusion that the increase in consumption between 2021 and 2022 are due to the increase in the low voltage consumption.