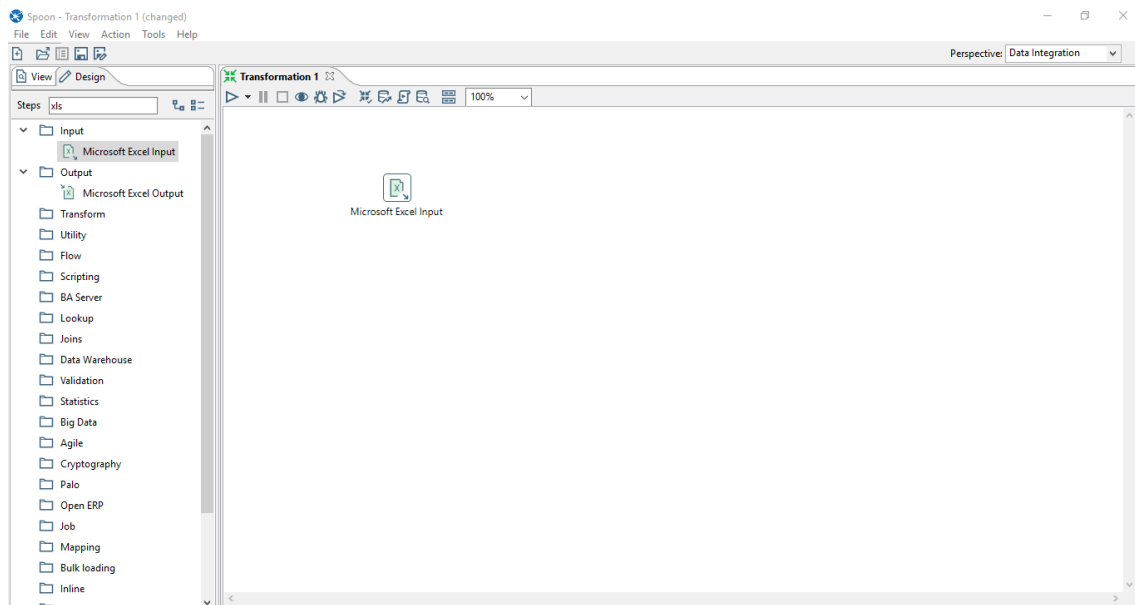
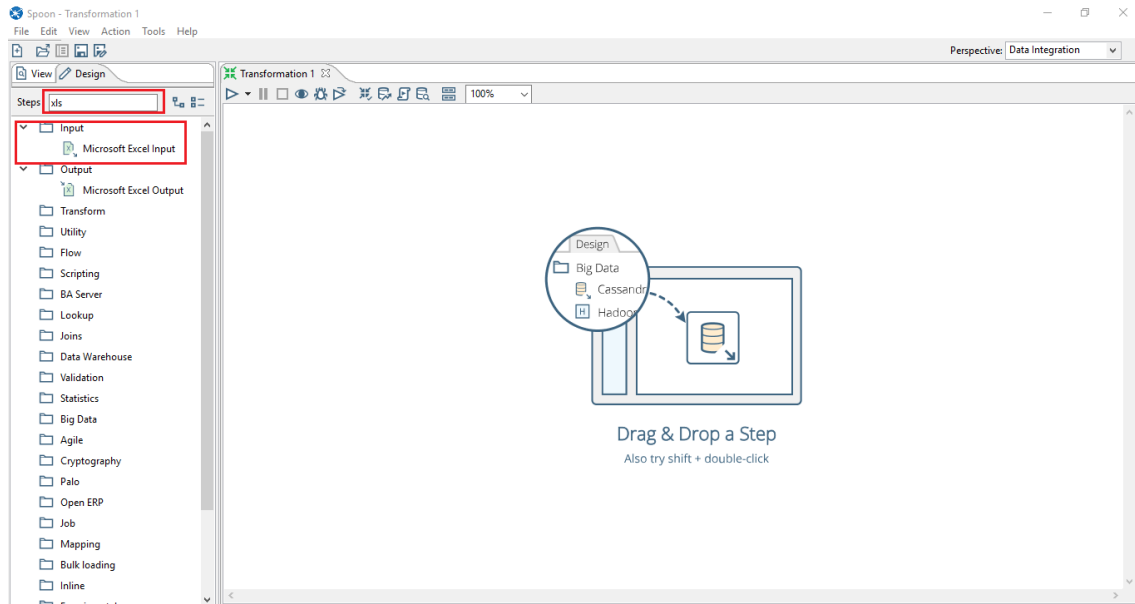
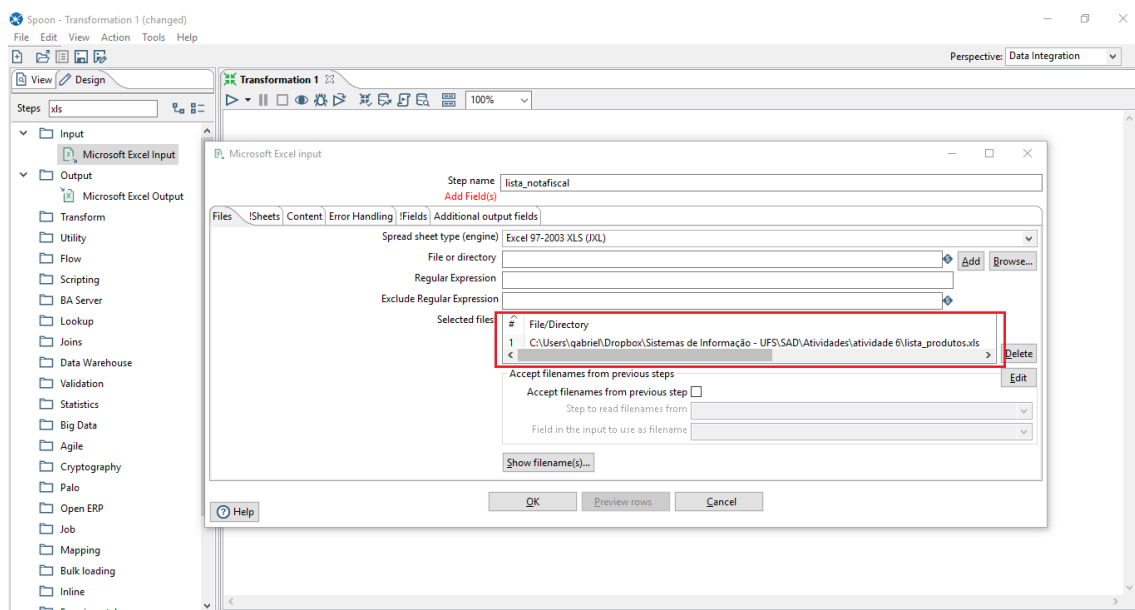
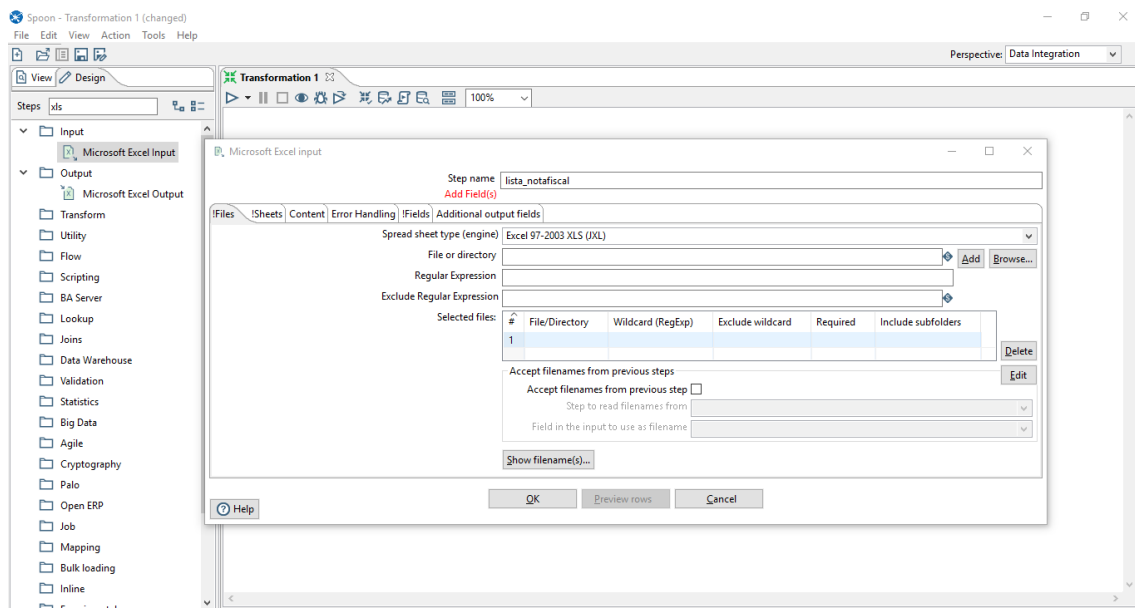


# Tutorial - Realizando o ETL através do PDI Pentaho em um banco de dados relacional

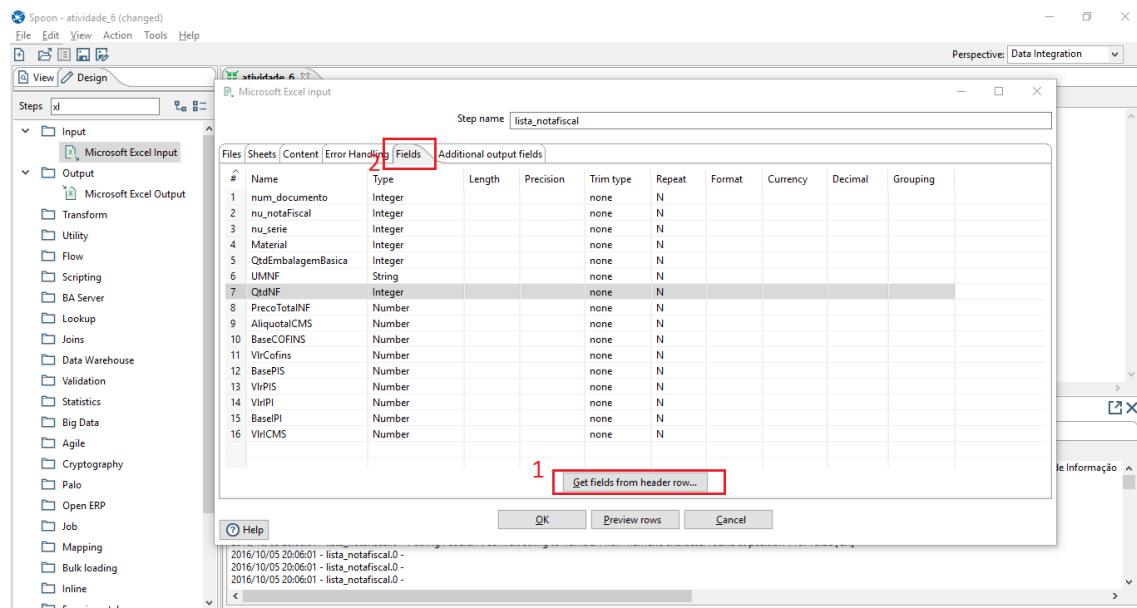
Passo 1: Selecionando o tipo de arquivo para a entrada de dados no formato xls (Excel):



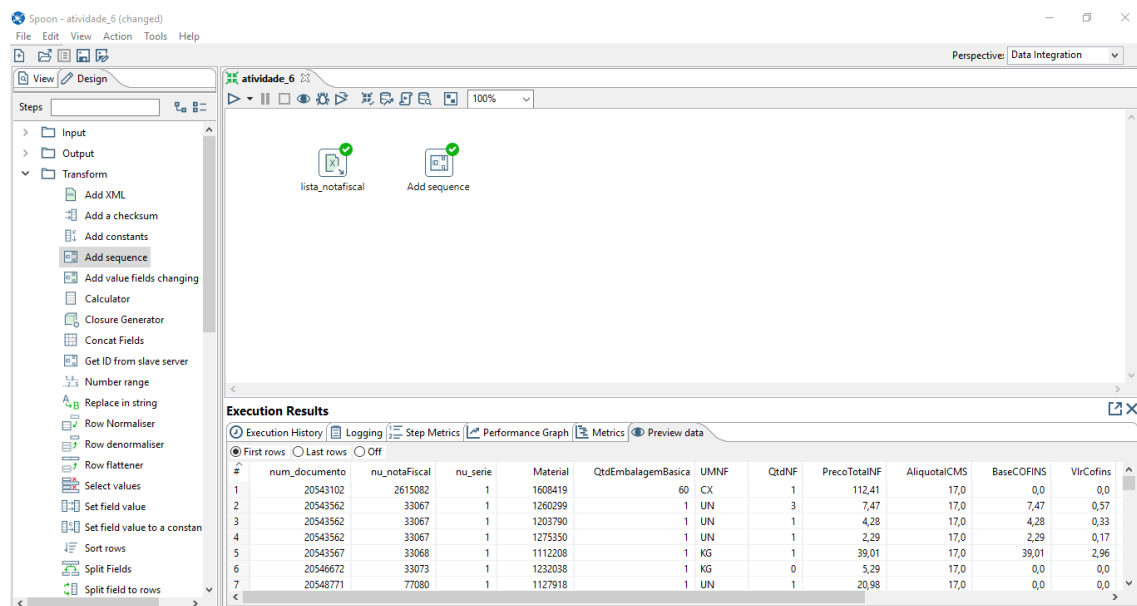
## Passo 2: Configurar step de entrada para receber os dados da planilha



Ir na aba *Fields* para configurar o formato das colunas da planilha



Passo 3: Adicionar um step de transformação *Add sequence*, para adicionar um colunar com números seqüências



The screenshot shows the Spoon IDE interface. On the left is the 'Design' pane with a project tree containing 'Input', 'Output', and 'Transform' folders. The 'Transform' folder is expanded, showing various tools like 'Add XML', 'Add a checksum', 'Add constants', 'Add sequence' (highlighted), 'Add value fields changing', 'Calculator', 'Closure Generator', 'Concat Fields', 'Get ID from slave server', 'Number range', 'Replace in string', 'Row Normaliser', 'Row denormaliser', 'Row flattener', 'Select values', 'Set field value', 'Set field value to a constant', 'Sort rows', 'Split Fields', and 'Split field to rows'.

The main workspace displays a data flow diagram with a table component labeled 'lista\_notas'. A dialog box titled 'Get Value From Sequence' is open, showing the following configuration:

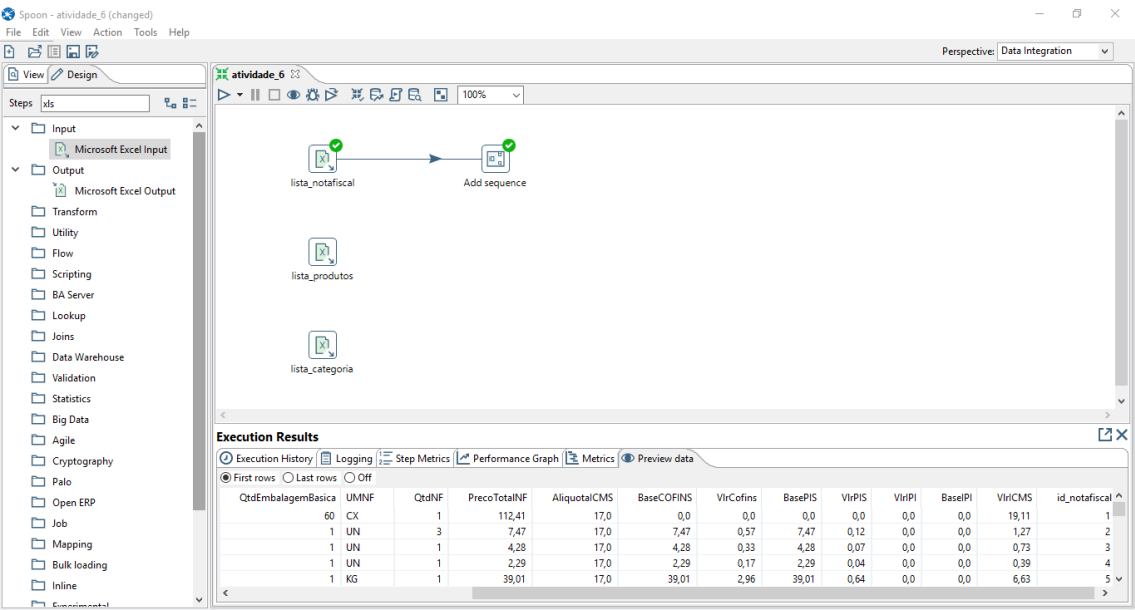
- Step name: Add sequence
- Name of value: id\_notafiscal
- Use a database to generate the sequence: ☐
- Use DB to get sequence?: ☐
- Connection: (empty)
- Schema name: (empty)
- Sequence name: SEQ\_
- Use a transformation counter to generate the sequence: ☒
- Use counter to calculate sequence?: ☒
- Counter name (optional): (empty)
- Start at value: 1
- Increment by: 1
- Maximum value: 999999999

Below the dialog box, the 'Execution Results' table is visible, showing the output of the 'Add sequence' step. The table has columns for 'First rows', 'Last rows', 'num\_document', and several calculated fields: 'PrecoTotalNF', 'AliquotaCMS', 'BaseCOFINS', and 'VlrCofins'.

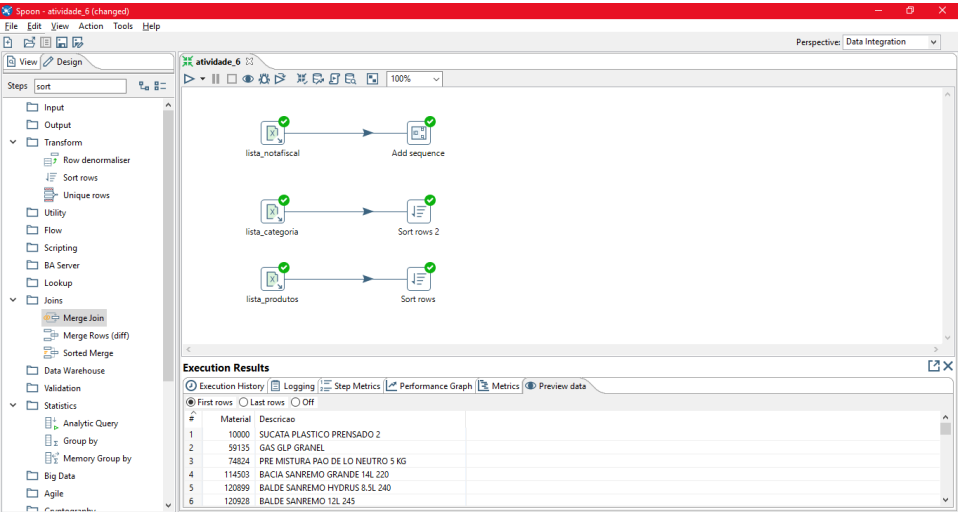
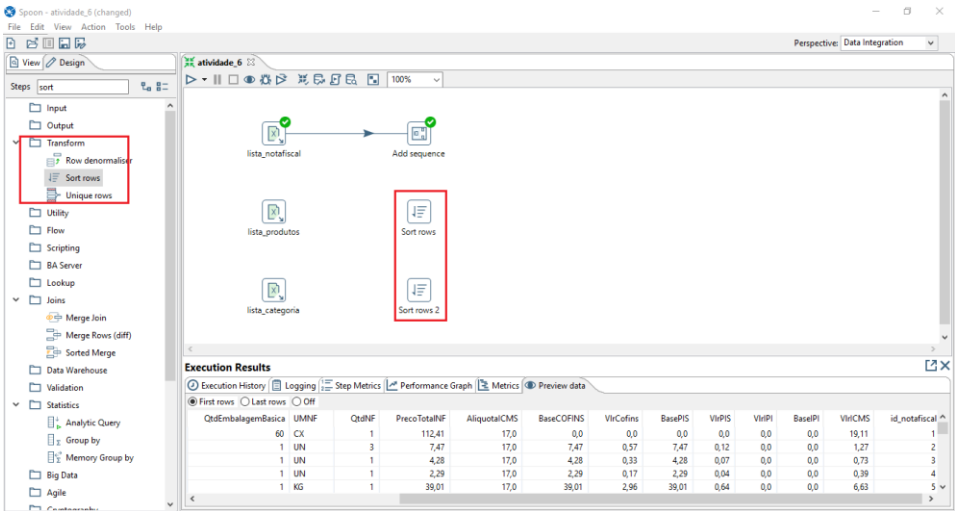
First rows	Last rows	num_document	PrecoTotalNF	AliquotaCMS	BaseCOFINS	VlrCofins					
1	20543102	2615082	1	1608419	60	CX	1	112,41	17,0	0,0	0,0
2	20543562	33067	1	1260299	1	UN	3	7,47	17,0	7,47	0,57
3	20543562	33067	1	1203790	1	UN	1	4,28	17,0	4,28	0,33
4	20543562	33067	1	1275350	1	UN	1	2,29	17,0	2,29	0,17
5	20543567	33068	1	1112208	1	KG	1	39,01	17,0	39,01	2,96
6	20546672	33073	1	1232038	1	KG	0	5,29	17,0	0,0	0,0
7	20548771	77080	1	1127918	1	UN	1	20,98	17,0	0,0	0,0

[illegible]

Adicionar os demais steps de entrada no formato de dados xls



Passo 4: Adicionar os stpes de transformação para ordenar os dados da planilha



Spoon - atividade\_6 (changed)

File Edit View Action Tools Help

Perspective: Data Integration

View Design

Steps sort

- Input
- Output
- Transform
  - Row denormaliser
  - Sort rows
  - Unique rows
- Utility
- Flow
- Scripting
- BA Server
- Lookup
- Joins
  - Merge Join
  - Merge Rows (diff)
  - Sorted Merge
- Data Warehouse
- Validation
- Statistics
  - Analytic Query
  - Group by
  - Memory Group by
- Big Data
- Agile
- Customizable

atividade\_6

100%

lista\_notafiscal → Add sequence → Sort rows 3

lista\_categoria → Sort rows 2

lista\_produtos → Sort rows

Execution Results

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

First rows | Last rows | Off

CtdEmbalagemBasica	UMNF	QtdNF	PrecoTotalNF	AliquotatCMS	BaseCOFINS	VlrCofins	BasePIS	VlrPIS	VlrIPI	BasePI	VlrICMS	id_notafiscal
60	CX	1,0	112,41	17,0	0,0	0,0	0,0	0,0	0,0	0,0	19,11	1
1	UN	1,0	4,28	17,0	4,28	0,33	4,28	0,07	0,0	0,0	0,73	3
1	UN	3,0	7,47	17,0	7,47	0,57	7,47	0,12	0,0	0,0	1,27	2
1	UN	1,0	2,29	17,0	2,29	0,17	2,29	0,04	0,0	0,0	0,39	4
1	KG	1,295	39,01	17,0	39,01	2,96	39,01	0,64	0,0	0,0	6,63	5

Passo 5: Realizar o step de *Merge Join* nas duas planilhas, lista\_notafiscal e lista\_categoria.

Spoon - atividade\_6 (changed)

File Edit View Action Tools Help

Perspective: Data Integration

View Design

Steps sort

- Input
- Output
- Transform
  - Row denormaliser
  - Sort rows
  - Unique rows
- Utility
- Flow
- Scripting
- BA Server
- Lookup
- Joins
  - Merge Join
  - Merge Rows (diff)
  - Sorted Merge
- Data Warehouse
- Validation
- Statistics
  - Analytic Query
  - Group by
  - Memory Group by
- Big Data
- Agile
- Customizable

atividade\_6

100%

lista\_notafiscal → Add sequence → Sort rows 3

lista\_categoria → Sort rows 2

lista\_produtos → Sort rows

Sort rows 3 → Merge Join

Sort rows 2 → Merge Join

Sort rows

Execution Results

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

First rows | Last rows | Off

#	CN	Descrição
1	A1	Cancelamento entrada (A)
2	A2	Cancelamento saída (A)
3	C1	Conhecimento entrada
4	C3	Conhecimento
5	CR	Cancelamento entrada (A)
6	E1	Nota fiscal entrada

Spoon - atividade\_6

File Edit View Action Tools Help

Perspective: Data Integration

View Design

Steps sort

- Input
- Output
- Transform
  - Row denormaliser
  - Sort rows
  - Unique rows
- Utility
- Flow
- Scripting
- BA Server
- Lookup
- Joins
  - Merge Join
  - Merge Rows (diff)
  - Sorted Merge
- Data Warehouse
- Validation
- Statistics
  - Analytic Query
  - Group by
  - Memory Group by
- Big Data
- Agile
- Customizable

atividade\_6

100%

lista\_notafiscal → Add sequence → Sort rows 3

lista\_categoria → Sort rows 2

lista\_produtos → Sort rows

Sort rows 3 → Merge Join

Sort rows 2 → Merge Join

Sort rows

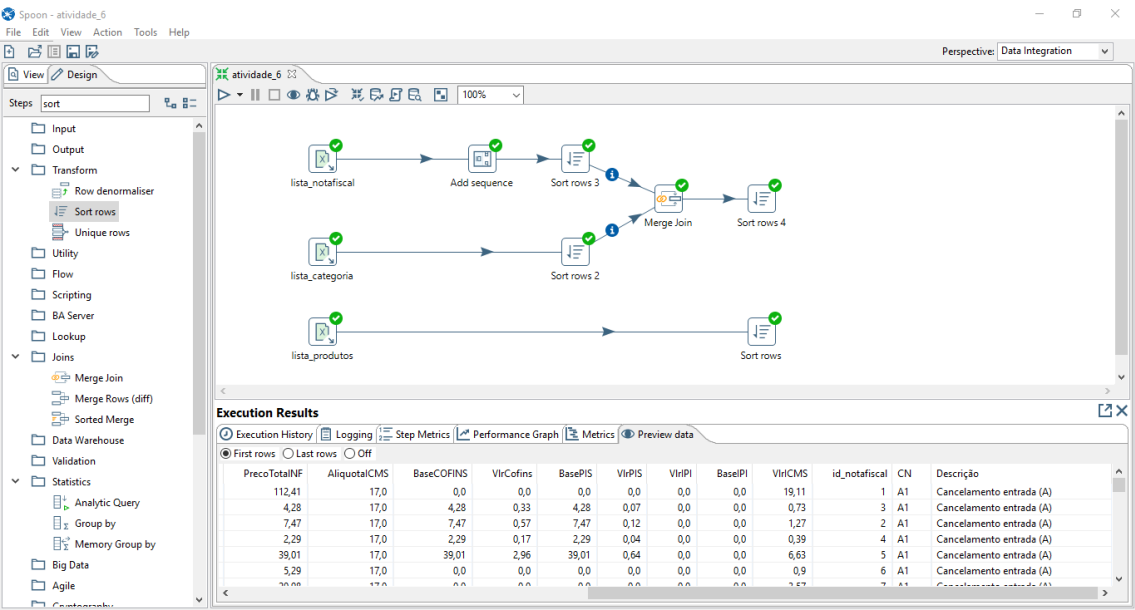
Execution Results

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

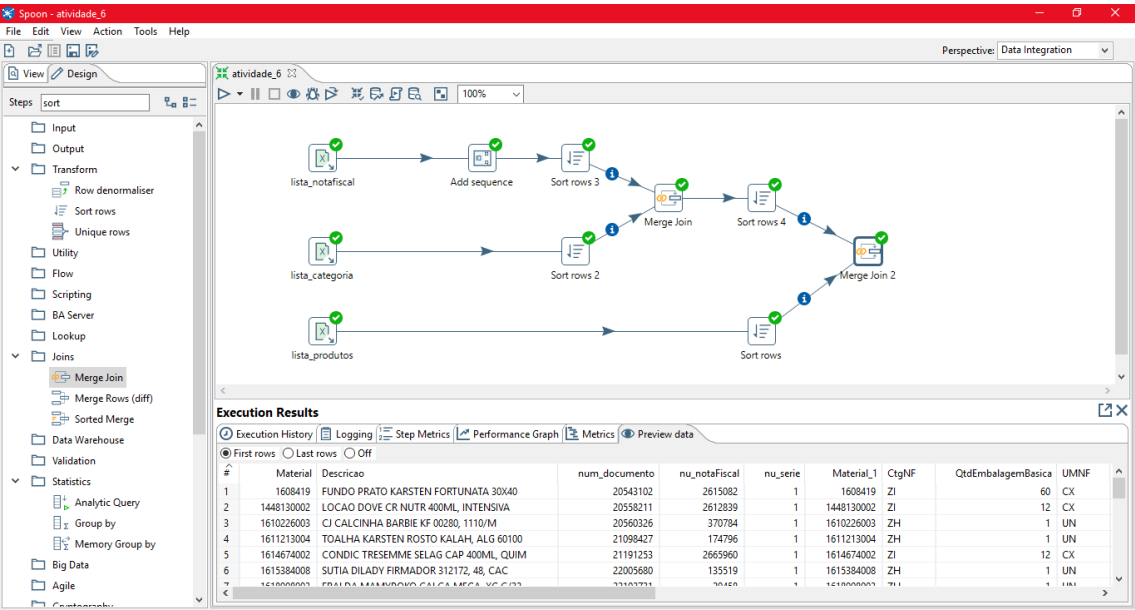
First rows | Last rows | Off

PrecoTotalNF	AliquotatCMS	BaseCOFINS	VlrCofins	BasePIS	VlrPIS	VlrIPI	BasePI	VlrICMS	id_notafiscal	CN	Descrição
179,91	17,0	179,91	13,67	179,91	2,97	0,0	0,0	30,58	17199	A1	Cancelamento entrada (A)
99,91	17,0	99,91	7,59	99,91	1,65	0,0	0,0	16,98	17198	A1	Cancelamento entrada (A)
30,9	17,0	30,9	2,35	30,9	0,51	0,0	0,0	5,25	17200	A1	Cancelamento entrada (A)
24,0	17,0	24,0	0,0	24,0	0,0	0,0	0,0	4,08	17197	A1	Cancelamento entrada (A)
7,0	17,0	7,0	0,53	7,0	0,12	0,0	0,0	1,19	17196	A1	Cancelamento entrada (A)
105,0	17,0	105,0	0,0	105,0	0,0	0,0	0,0	17,85	17195	A1	Cancelamento entrada (A)

Passo 6: Realizar novamente a ordenação no resultado do step *Merge Join*



Passo 7: Realizar o *Merge Join* nas duas saídas ordenadas



Passo 8: Realizar a transformação *Select values*, para selecionar as colunas que irá conter a nova saída

The screenshot shows the SAP Data Integration Studio interface. The left pane displays the 'Steps' list with 'Select values' highlighted. The main workspace shows a data flow diagram with three input tables: 'lista\_notafiscal', 'lista\_categoria', and 'lista\_produtos'. These are connected to 'Sort rows 3', 'Sort rows 2', and 'Sort rows' respectively. These sorted rows are then merged into 'Merge Join' and 'Merge Join 2' steps. The final output is generated by the 'Select values' step. The 'Execution Results' pane at the bottom shows a table with columns: #Fiscal, nu\_serie, cd\_categoria, sd\_categoria, nu\_material, ds\_material, un\_medida, nu\_quantidade, and vl\_total. The table contains 6 rows of data.

#Fiscal	nu_serie	cd_categoria	sd_categoria	nu_material	ds_material	un_medida	nu_quantidade	vl_total
15082	1	A1	Cancelamento entrada (A)	1608419	FUNDO PRATO KARSTEN FORTUNATA 30X40	CX	1,0	112,41
12839	1	A1	Cancelamento entrada (A)	1448130002	LOCAO DOVE CR NUTR 400ML INTENSIVA	CX	1,0	128,39
70784	1	A1	Cancelamento entrada (A)	1610226003	CJ CALCINHA BARBIE KF 00280, 1110/M	UN	1,0	34,9
74796	1	A1	Cancelamento entrada (A)	1611213004	TOALHA KARSTEN ROSTO KALAH, ALG 60100	UN	1,0	6,9
65960	1	A1	Cancelamento entrada (A)	1614674002	CONDIC TRESEMME SELAG CAP 400ML, QUIM	CX	1,0	123,18
35519	1	A1	Cancelamento entrada (A)	1615384008	SUTIA DILADY FIRMADOR 312172, 48, CAC	UN	1,0	44,89

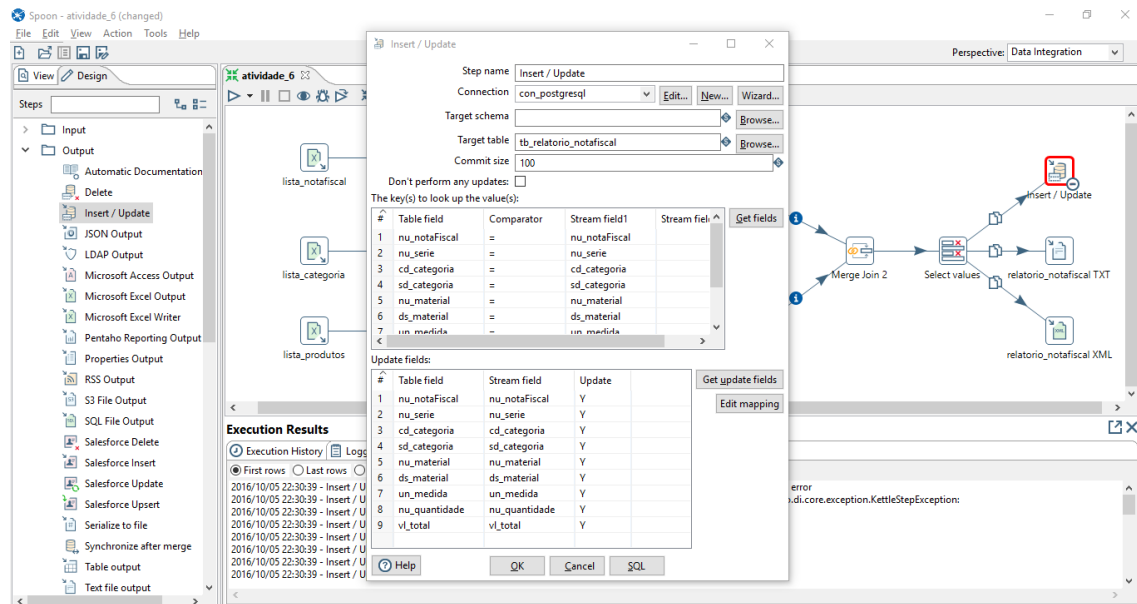
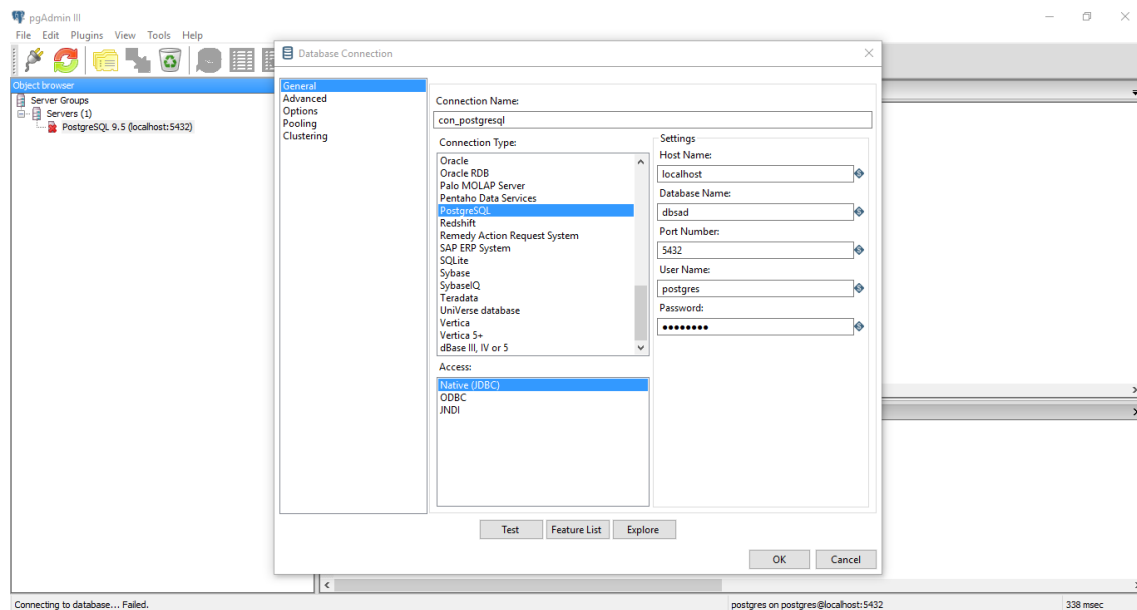
Passo 9: Adicionar os stpes de saída para os formatos XML, TXT e gerar uma carga no banco de dados PostgreSQL

The screenshot shows the SAP Data Integration Studio interface. The left pane displays the 'Steps' list with 'XML Output' highlighted. The main workspace shows a data flow diagram with three input tables: 'lista\_notafiscal', 'lista\_categoria', and 'lista\_produtos'. These are connected to 'Sort rows 3', 'Sort rows 2', and 'Sort rows' respectively. These sorted rows are then merged into 'Merge Join' and 'Merge Join 2' steps. The final output is generated by the 'Select values' step. The 'Execution Results' pane at the bottom shows a table with columns: #, nu\_notafiscal, nu\_serie, cd\_categoria, sd\_categoria, nu\_material, ds\_material, un\_medida, and nu\_quantidade. The table contains 6 rows of data.

#	nu_notafiscal	nu_serie	cd_categoria	sd_categoria	nu_material	ds_material	un_medida	nu_quantidade
1	2615082	1	A1	Cancelamento entrada (A)	1608419	FUNDO PRATO KARSTEN FORTUNATA 30X40	CX	
2	2612839	1	A1	Cancelamento entrada (A)	1448130002	LOCAO DOVE CR NUTR 400ML INTENSIVA	CX	
3	370784	1	A1	Cancelamento entrada (A)	1610226003	CJ CALCINHA BARBIE KF 00280, 1110/M	UN	
4	174796	1	A1	Cancelamento entrada (A)	1611213004	TOALHA KARSTEN ROSTO KALAH, ALG 60100	UN	
5	2665960	1	A1	Cancelamento entrada (A)	1614674002	CONDIC TRESEMME SELAG CAP 400ML, QUIM	CX	
6	125519	1	A1	Cancelamento entrada (A)	1615384008	SUTIA DILADY FIRMADOR 312172, 48, CAC	UN	



## Passo 10: Configurar a conexão com o banco de dados



## Criar a base e a tabela com o script gerado pela ferramenta

The screenshot shows the Apache Kettle Spoon interface. The 'Insert / Update' dialog is open, showing the step name 'Insert / Update', connection 'con\_postgresql', target schema 'tb\_relatorio\_notafiscal', and target table 'tb\_relatorio\_notafiscal'. The commit size is set to 100. Below the dialog, the 'Simple SQL editor' is open, displaying the following SQL script:

```

1 CREATE TABLE tb_relatorio_notafiscal
2 (
3     nu_notafiscal DOUBLE PRECISION
4     , nu_serie DOUBLE PRECISION
5     , cd_categoria TEXT
6     , ds_categoria TEXT
7     , nu_material DOUBLE PRECISION
8     , un_medida TEXT
9     , nu_quantidade DOUBLE PRECISION
10    , vl_total DOUBLE PRECISION
11 )
12 :CREATE INDEX idx_tb_relatorio_notafiscal_lookup ON t

```

The 'Execution Results' table shows the following data:

Line	Column 0
1	
2	
3	
4	
5	
6	ds_material
7	un_medida
8	nu_quantidade
9	vl_total

## Executar os steps

The screenshot shows the Apache Kettle Spoon interface with the job 'atividade\_6' running. The job flow includes steps for reading input files (lista\_notafiscal, lista\_categoria, lista\_produtos), processing (Add sequence, Sort rows, Merge Join, Merge Join 2, Select values), and writing output (relatorio\_notafiscal TXT, relatorio\_notafiscal XML). The 'Execution Results' table shows the following data:

#	nu_notafiscal	nu_serie	cd_categoria	ds_categoria	nu_material	ds_material	un_medida	nu_quantidade
1	2615082	1	A1	Cancelamento entrada (A)	1608419	FUNDO PRATO KARSTEN FORTUNATA 30X40	CX	
2	2612839	1	A1	Cancelamento entrada (A)	1448130002	LOCAO DOVE CR NUTR 400ML INTENSIVA	CX	
3	370784	1	A1	Cancelamento entrada (A)	1610226003	CJ CALCINHA BARBIE KF 00280, 1110/M	UN	
4	174796	1	A1	Cancelamento entrada (A)	1611213004	TOALHA KARSTEN ROSTO KALAH, ALG 60100	UN	
5	2665960	1	A1	Cancelamento entrada (A)	1614674002	CONDIC TRESMEME SELAG CAP 400ML, QUIM	CX	
6	126510	1	A1	Cancelamento entrada (A)	1615384008	SLITIA PU ADV BDMADON 313173, 48, CAC	UN	