

Sistemas de Apoio a Decisão

(Inteligência nos Negócios - *Business Inteligente*)

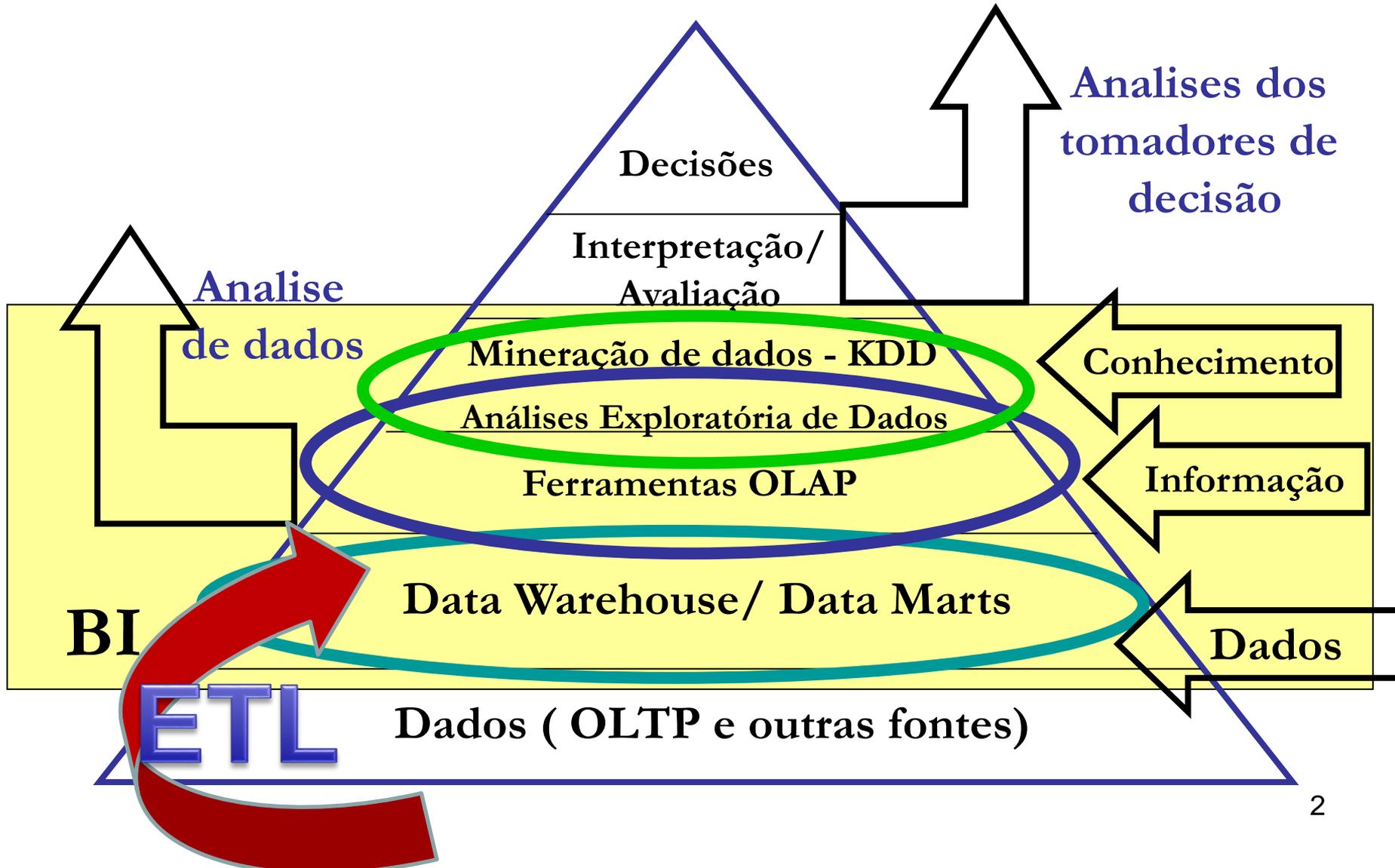


Sistemas de Informação

Aran Bey Tcholakian Morales, Dr. Eng.

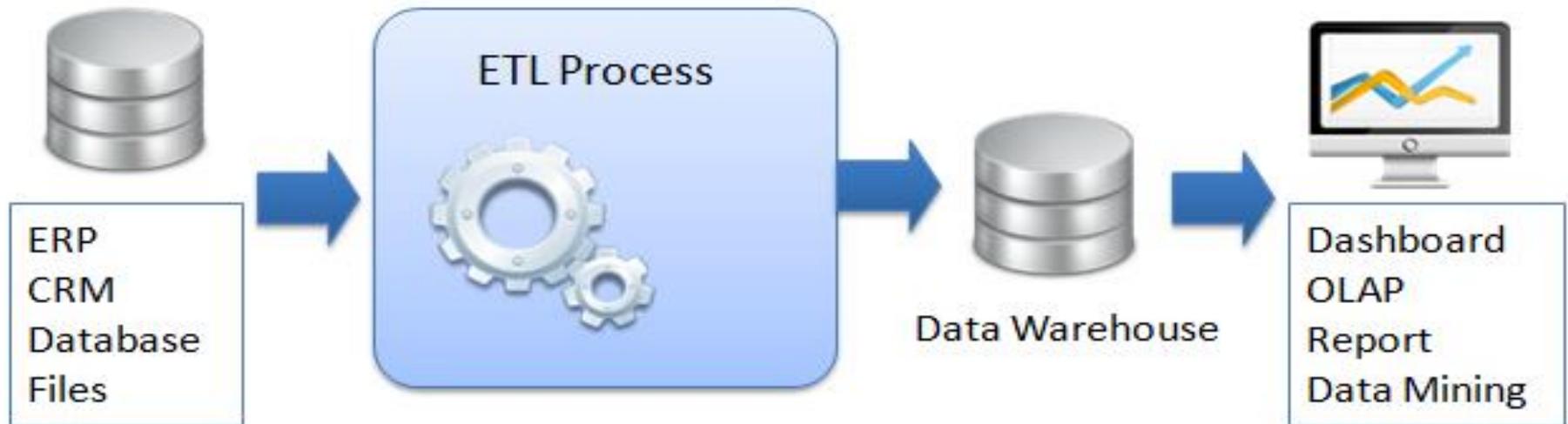
(Apostila 4: Processo ETL – Pentaho Data Integration (Kettle))

Fundamentação da disciplina



Processo ETL

(Extração, Transformação e Carga)



ETL - Extração, Transformação, Carga

- O **ETL** é um sistema ou conjunto de scripts SQLs para extrair os dados das bases de operação e carregá-las no modelo dimensional construído depois da transformação.
- O objetivo da etapa de **ETL** é fazer a integração de informações de fontes múltiplas e complexas, portanto, torna-se uma etapa bastante crítica, já que uma informação carregada erroneamente pode trazer consequências imprevisíveis nas fases posteriores. Esta etapa divide-se basicamente em três passos: **extração**, **transformação** e **carga**.

ETL - Extração, Transformação, Carga

- **Extração:** implica muitas vezes na **seleção** de **diferentes** fontes de dados e na **integração** de tais dados com o objetivo de obter uma única fonte.

As origens das fontes de dados podem ser várias e também podem possuir diferentes formatos, tais como sistemas transacionais, ERP, planilhas, arquivos-textos e até fontes externas

ETL - Extração, Transformação, Carga

- **Transformação:** a origem dos dados muitas vezes pode ser de sistemas diferentes, então é necessário padronizar os diferentes formatos, pois o usuário não pode ver informações iguais em formatos diferentes.

A limpeza dos dados (uma forma de transformação) se dá porque os dados normalmente advêm de uma fonte muitas vezes desconhecida, concebida há muito tempo e contendo muito lixo e inconsistência.

Operações de **remoção de ruídos**, de **atributos incompletos**, tratamento de **erros de digitação** ou **erros nos sistemas de captura de dados**, são tratados neste etapa.

ETL - Extração, Transformação, Carga

Podem ser as seguintes transformações:

Resumo ou agregação: é o processo de eliminar atributos ou reduzir o número de valores de um determinado atributo - por exemplo reduzir o atributo data com granularidade em dias para mês (redução de 365 para 12).

ETL - Extração, Transformação, Carga

Redução do número de valores de um atributo: é feita “discretizando” os valores de um atributo contínuo.

Muitos algoritmos de classificação, por exemplo, necessitam de dados categorizados (discretos), algoritmos de associação de dados binários.

Assim, características que assumem valores contínuos podem ser transformados em valores discretos (em classes) e os valores discretos em binários.

ETL - Extração, Transformação, Carga

- **Carga:** uma vez que a origem dos dados pode ser de sistemas diferentes, padronizam-se os diferentes formatos de modo que fiquem num formato uniforme, normalmente sugerido pelo próprio usuário.
- Com isso, a análise dos dados recuperados pela ferramenta OLAP fica mais fácil, pois o usuário não estará vendo informações iguais em formatos diferentes.

ETL: algumas dificuldades

- Várias origens para um mesmo dado: conflitos estruturais, de conteúdo e de formato nos dados;
- Dados “faltantes”, dados com “erros”;
- Não conformidade dos dados com as regras do negócio;
- Dados significativos em campos de entrada livre;
- Necessidade de normalização/desnormalização de dados;
- Necessidade de “juntar/separar” atributos;
- Diversos formatos de dados (xls, pdf, xml, etc.);
- Incompatibilidade entre ambientes operacionais diferentes.

Suite BI Open Pentaho: Data Integration

Exercícios da folha 6.

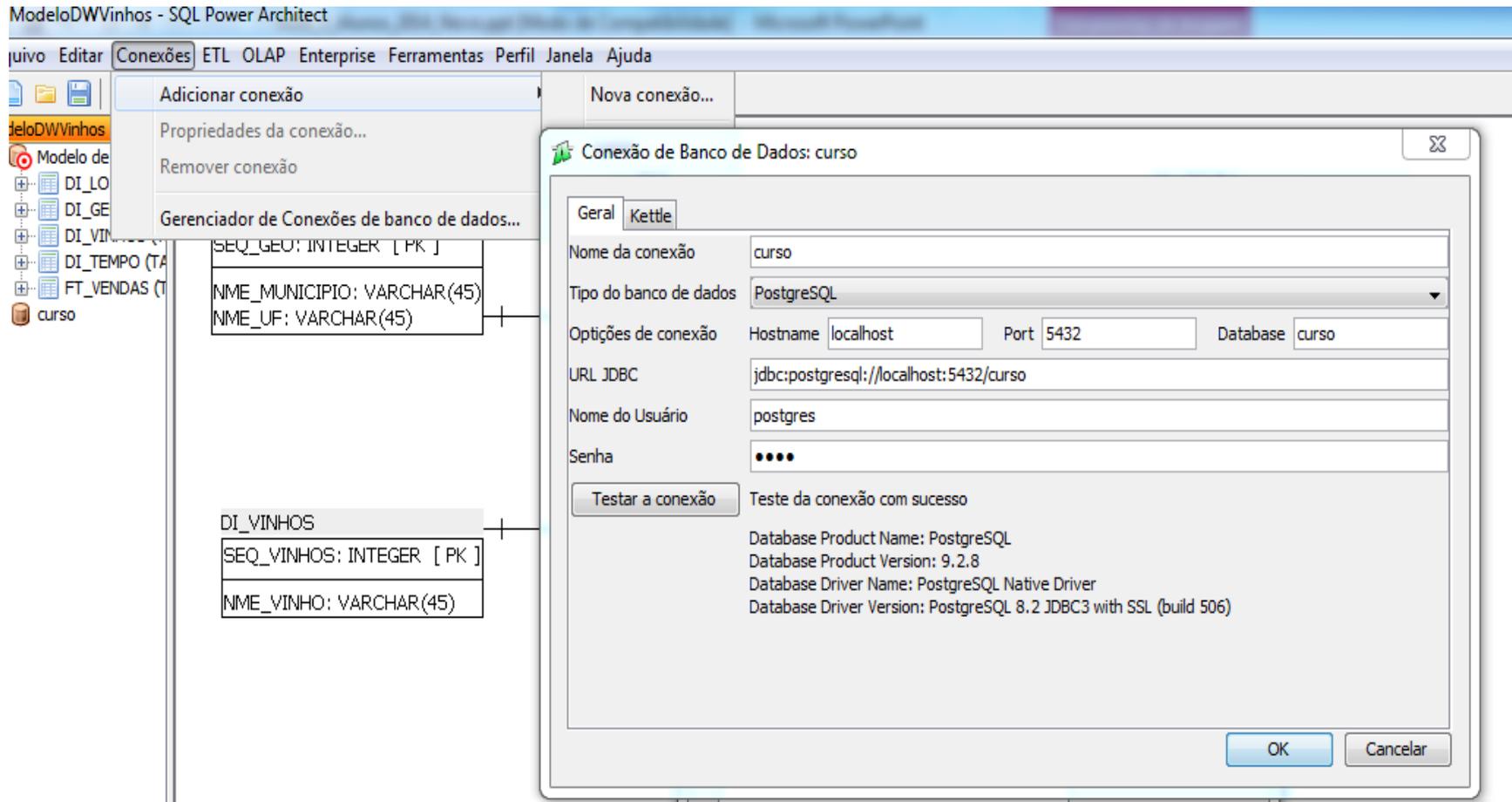
The screenshot displays the pgAdmin 4 interface. On the left, the 'Browser' pane shows a tree view of databases under 'PostgreSQL 9.6'. The 'Schemas (5)' folder is selected, and a context menu is open over it, with 'Create' and 'Schema...' options highlighted. The main pane shows 'Database sessions' and 'Tuples in' charts. The 'Database sessions' chart shows 1.00 for Active, 0.00 for Idle, and 0.00 for Total. The 'Tuples in' chart shows 1.00 for Active, 0.00 for Idle, and 0.00 for Deletes. On the right, a 'Create - Schema' dialog box is open, showing the 'General' tab with the following fields:

- Name:
- Owner:
- Comment:

1. Criamos o Schema *dwvinhos_pdi*

Suite BI Open Pentaho: Data Integration

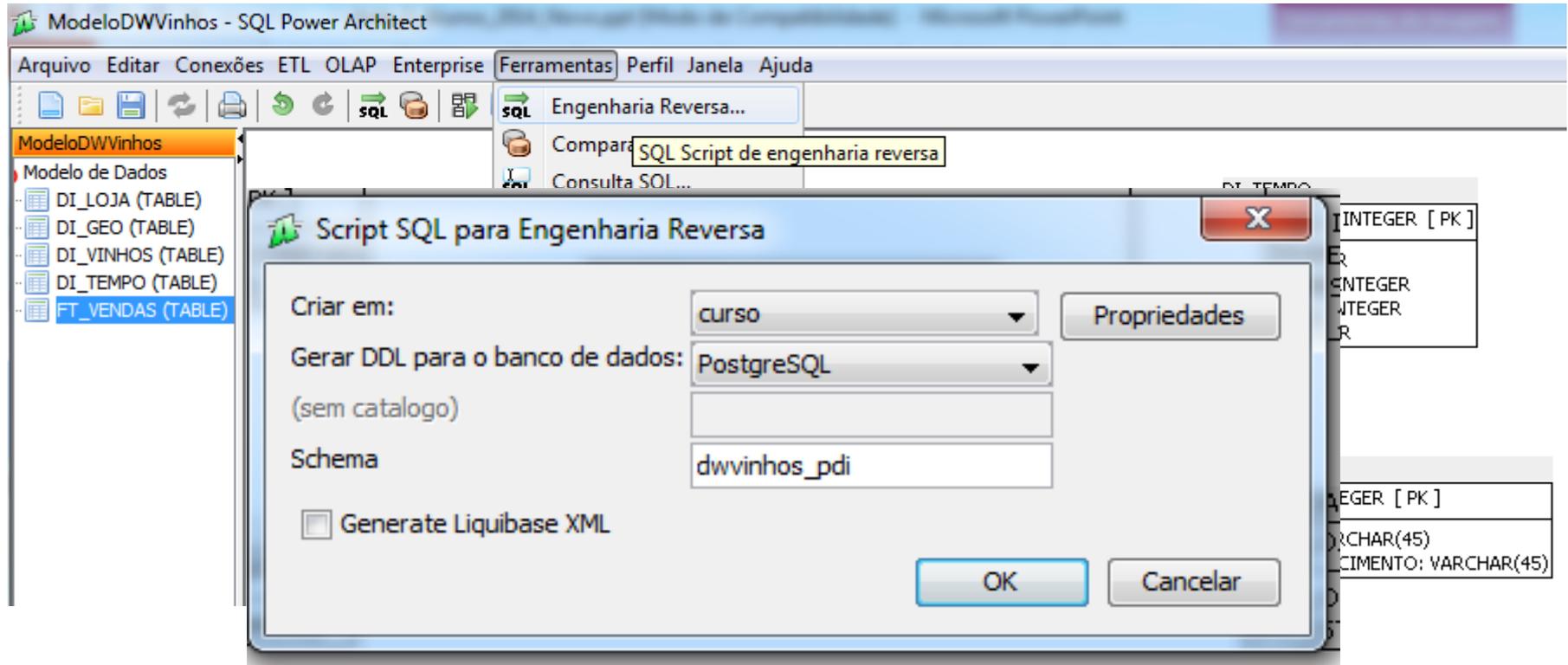
Exercícios da folha 6.



2. Criamos a conexão “*curso*” na opção:
Conexões – Adicionar conexão – Nova conexão

Suite BI Open Pentaho: Data Integration

Exercícios da folha 6.



3. Geramos o scripts das tabelas: *Ferramentas – Engenharia Reversa*. Escolhemos a base *curso*, o SGBD *PostgreSQL* e o schema *dwvinhos_pdi*, e damos *OK*.

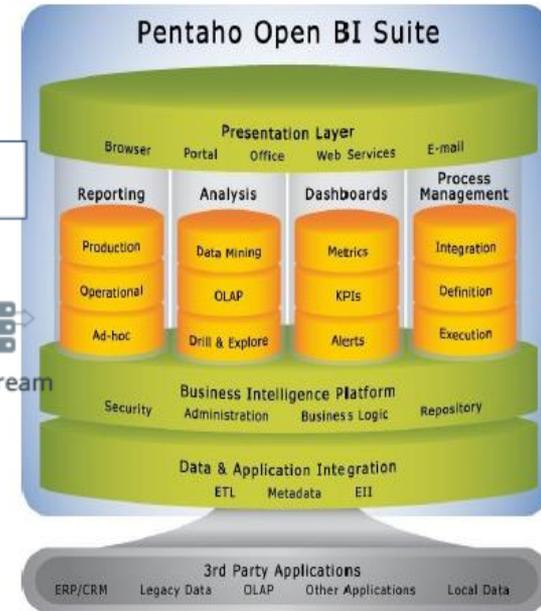
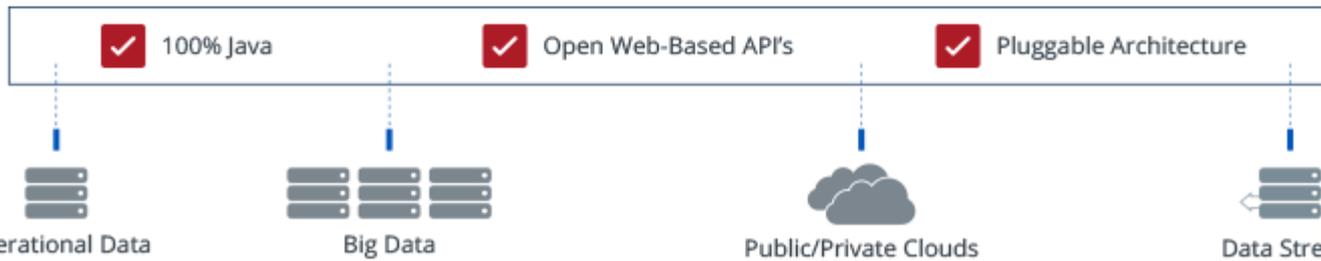
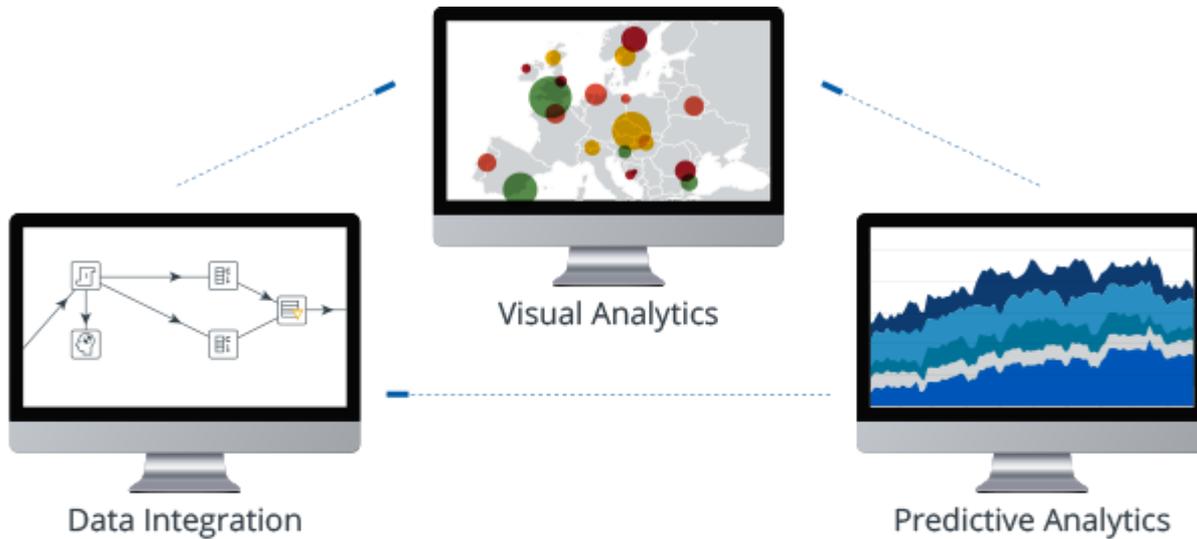
Suite BI Open Pentaho: Data Integration

The image displays a screenshot of a Pentaho Data Integration suite interface. It features several overlapping windows and data structure definitions:

- DI_GEO** table structure:
 - SEQ_GEO: INTEGER NOT NULL [PK]
 - NME_MUNICIPIO: VARCHAR(45) NOT NULL
 - NME_UF: VARCHAR(45) NOT NULL
 - COD_MUNICIPIO: INTEGER NOT NULL
 - COD_UF: VARCHAR NOT NULL
- DI_TEMPO** table structure:
 - SEQ_TEMPO: INTEGER NOT NULL [PK]
 - MES: INTEGER NOT NULL
 - TRIMESTRE: INTEGER NOT NULL
 - SEMESTRE: INTEGER NOT NULL
 - ANO: INTEGER NOT NULL
- DI_VINHOS** table structure:
 - SEQ_VINHOS: INTEGER NOT NULL [PK]
 - NME_VINHO: VARCHAR(45) NOT NULL
 - COD_VINHO: VARCHAR NOT NULL
- DI_ESTAB** table structure (partially visible):
 - ...A: INTEGER NOT NULL [PK]
 - ...A: VARCHAR(45) NOT NULL
 - ...ABELECIMENTO: VARCHAR(45) NOT NULL
 - ...A: INTEGER NOT NULL
 - ...COD_ESTABLECIMENTO: INTEGER NOT NULL
- Visualização script SQL** window:
 - Seu banco de dados destino é curso
 - CREATE SEQUENCE dwvinhos_pdi.di_tempo_seq_tempo_seq;
 - CREATE TABLE dwvinhos_pdi.DI_TEMPO (
 - SEQ_TEMPO INTEGER NOT NULL DEFAULT
 - nextval('dwvinhos_pdi.di_tempo_seq_tempo_seq'),
 - MES INTEGER NOT NULL,
- Relatório de erro** window:
 - Exceção inesperada
 - Exception type: org.postgresql...
 - Detail string: ERROR: colun...
- Mensagem** window:
 - 17 das 17 instruções executadas com sucesso.
- Bottom SQL window:
 - NME_VINHO VARCHAR(45) NOT NULL,
 - COD_VINHO VARCHAR NOT NULL,
 - CONSTRAINT di_vinhos_pk PRIMARY KEY
 - (SEQ_VINHOS)
 -);
 - ALTER SEQUENCE dwvinhos_pdi.di_vinhos_seq_vinhos_seq

4. Executamos o scripts.

Suite BI Open Pentaho



Suite BI Open Pentaho: Data Integration

sourceforge.net/projects/pentaho/files/

Apps Configurações Nova guia Importado Impressão Diário d... Comitê Gestor da I... Patrocinadores

Outros favoritos Lista de leitura

SOURCEFORGE

Open Source Software Business Software Resources

Help Create Join Login

Search for software or solutions

Home / Browse / Business & Enterprise / Enterprise / OLAP / Pentaho from Hitachi Vantara / Files

HITACHI
Inspire the Next

Pentaho from Hitachi Vantara

End to end data integration and analytics platform
Brought to you by: [beccany](#), [ecropper](#), [larrygrill](#), [lcheng-pentaho](#), and [3 others](#)

Summary **Files** Reviews Support Wiki News

Download Latest Version
pdi-ce-9.1.0.0-324.zip (1.8 GB)

Get Updates

Home

Name	Modified	Size	Downloads / Week
Pentaho 9.1	2020-09-25		6,702
White Papers	2020-04-20		37
Pentaho 9.0	2020-02-01		298
Pentaho 8.3	2019-07-10		426
Pentaho 8.2	2018-12-02		398

Recommended Projects

JasperReports Server

Get latest updates about Open Source Projects, Conferences and News.

Sign Up

No, Thank you

Suite BI Open Pentaho: Data Integration

Descompactar o arquivo `pdi-ce-9.X.X-stable.zip` e iniciar

Spoon.bat (Windows) ou **Spoon.sh** (Linux, MacOS).

PDI está formado por um conjunto de ferramentas:

Spoon: é a ferramenta gráfica para construir transformações e jobs.

Pan: é a ferramenta que nos permite executar as transformações do *spoon* de uma linha de comando.

Kitchen: similar ao Pan, mas para executar jobs.

Carte: é um servidor web que permite a execução remota de transformações e jobs.

Suite BI Open Pentaho: Data Integration

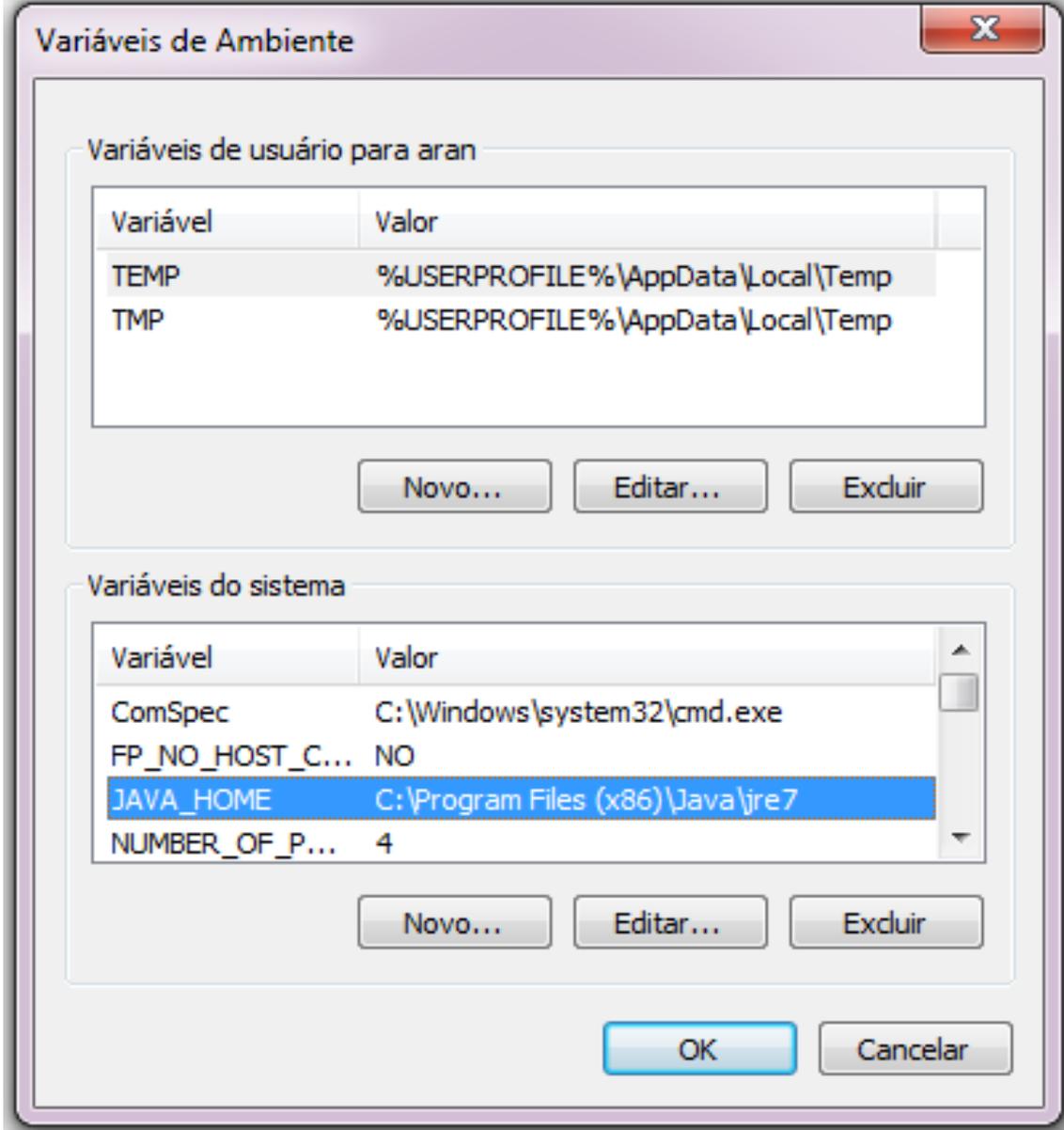
Configurando o JAVA.

Observação: utilizar JDK 8

Criando variável JAVA_HOME

1. Abrir Painel de Controle\Systema;
2. Selecione a aba Avançado e clique em Variáveis de Ambiente;
3. Em Variáveis do sistema clique no botão Nova;
4. Na janela Nova variável de sistema, digite JAVA_HOME no campo Nome da variável e no campo Valor da variável digite o caminho do diretório onde foi instalado o JDK e clique no botão OK;

Suite BI Open Pentaho: Data Integration



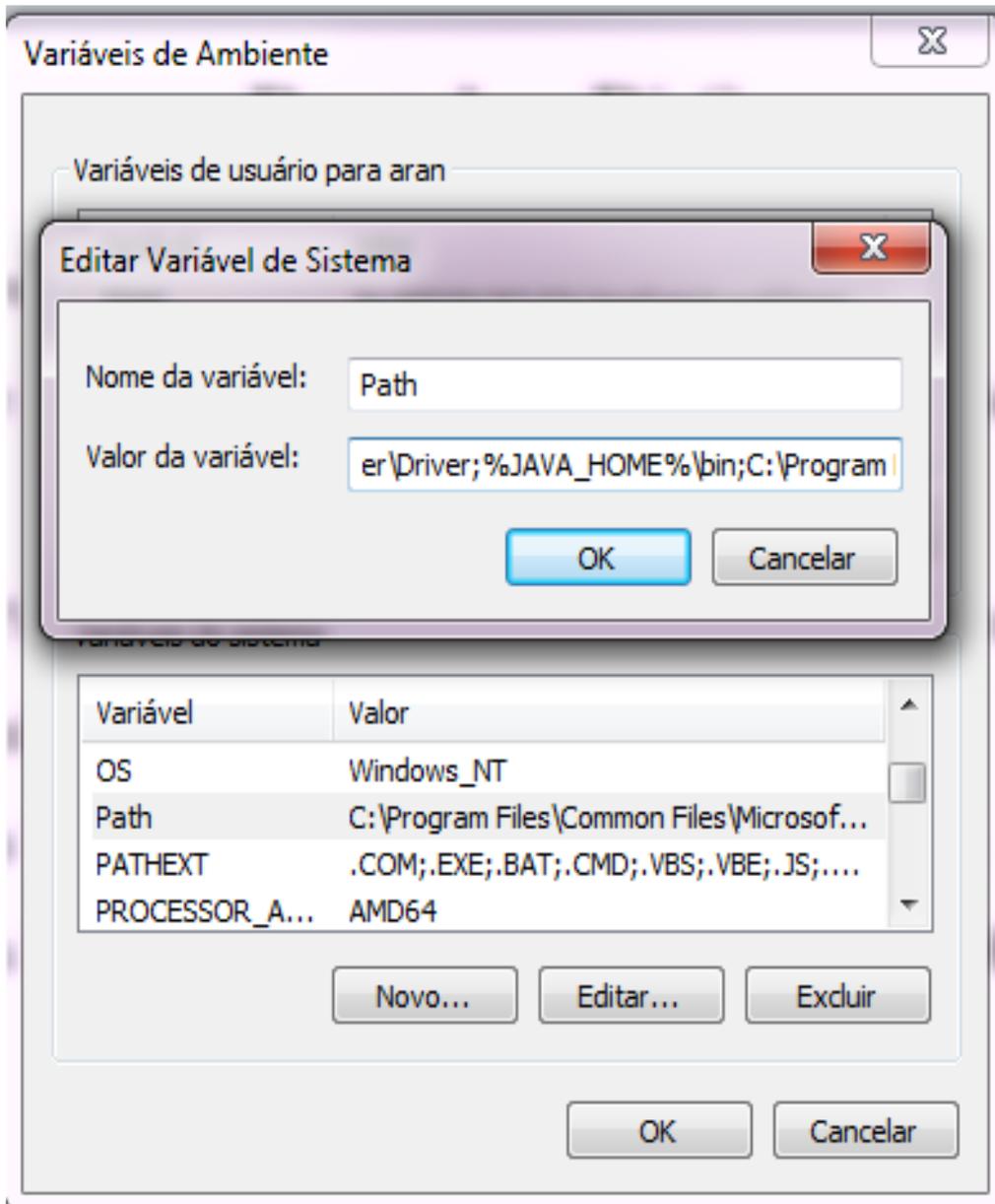
Suite BI Open Pentaho: Data Integration

Configurando o JAVA

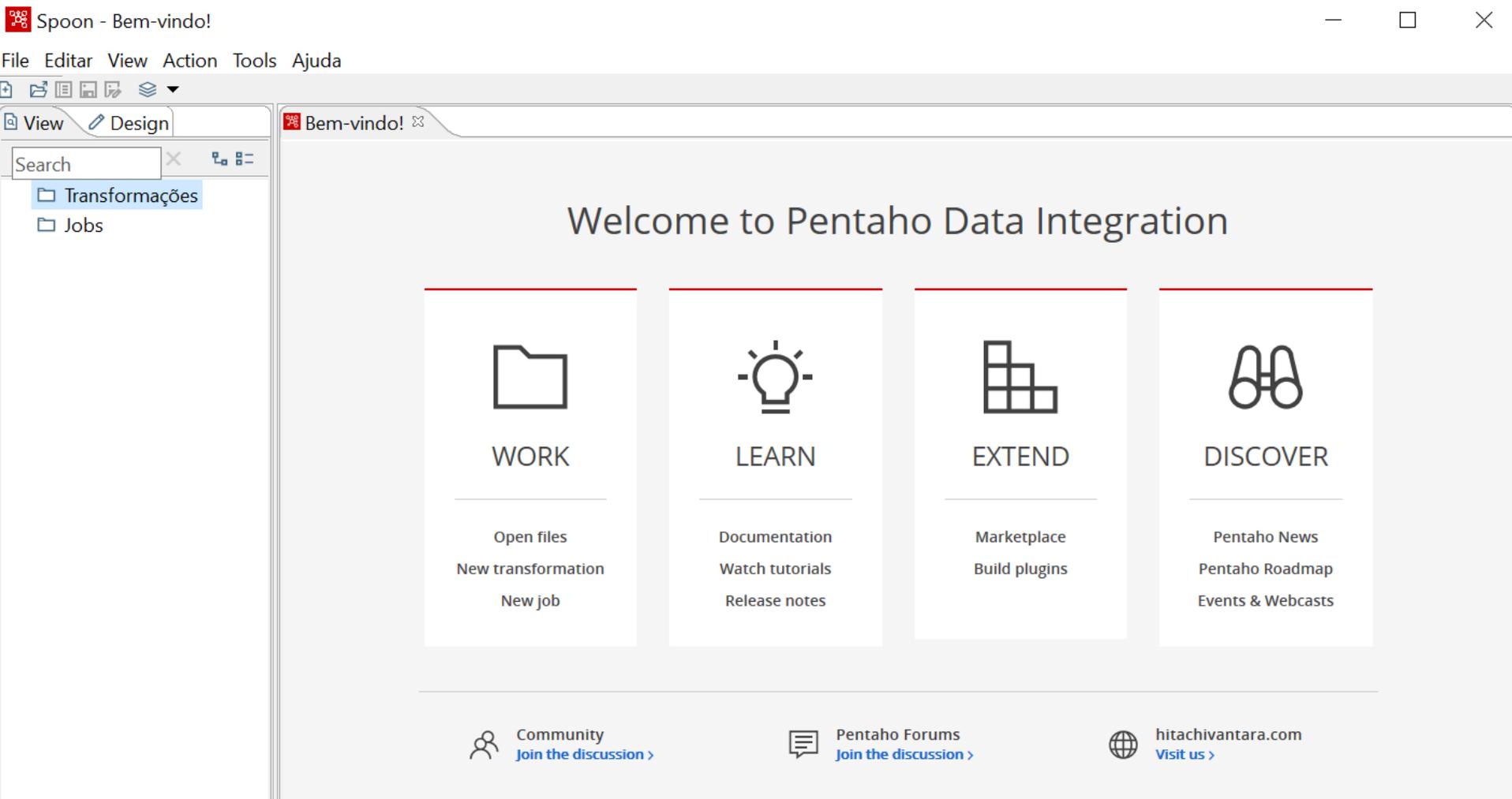
Editando variável PATH

1. Selecione a variável Path em Variáveis do sistema e clique no botão Editar;
2. Digite “;%JAVA_HOME%\bin” ao final do conteúdo do campo Valor da variável e em seguida clique no botão OK;
3. Na janela Variáveis de ambiente clique no botão OK;
4. Na janela Propriedades do sistema clique no botão OK.

Suite BI Open Pentaho: Data Integration



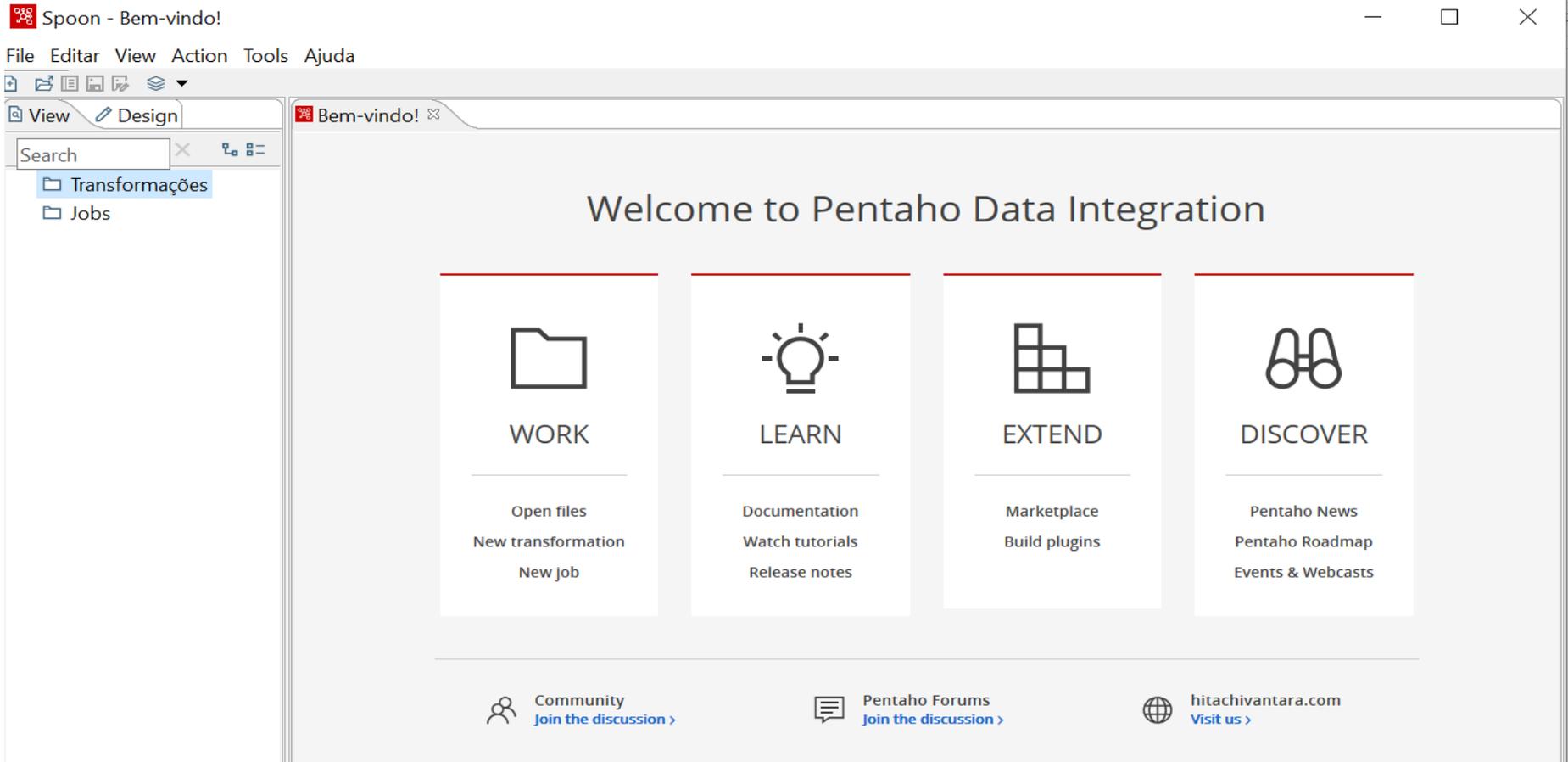
Suite BI Open Pentaho: Data Integration



Tela principal do aplicativo *data integration*.

Suite BI Open Pentaho: Data Integration

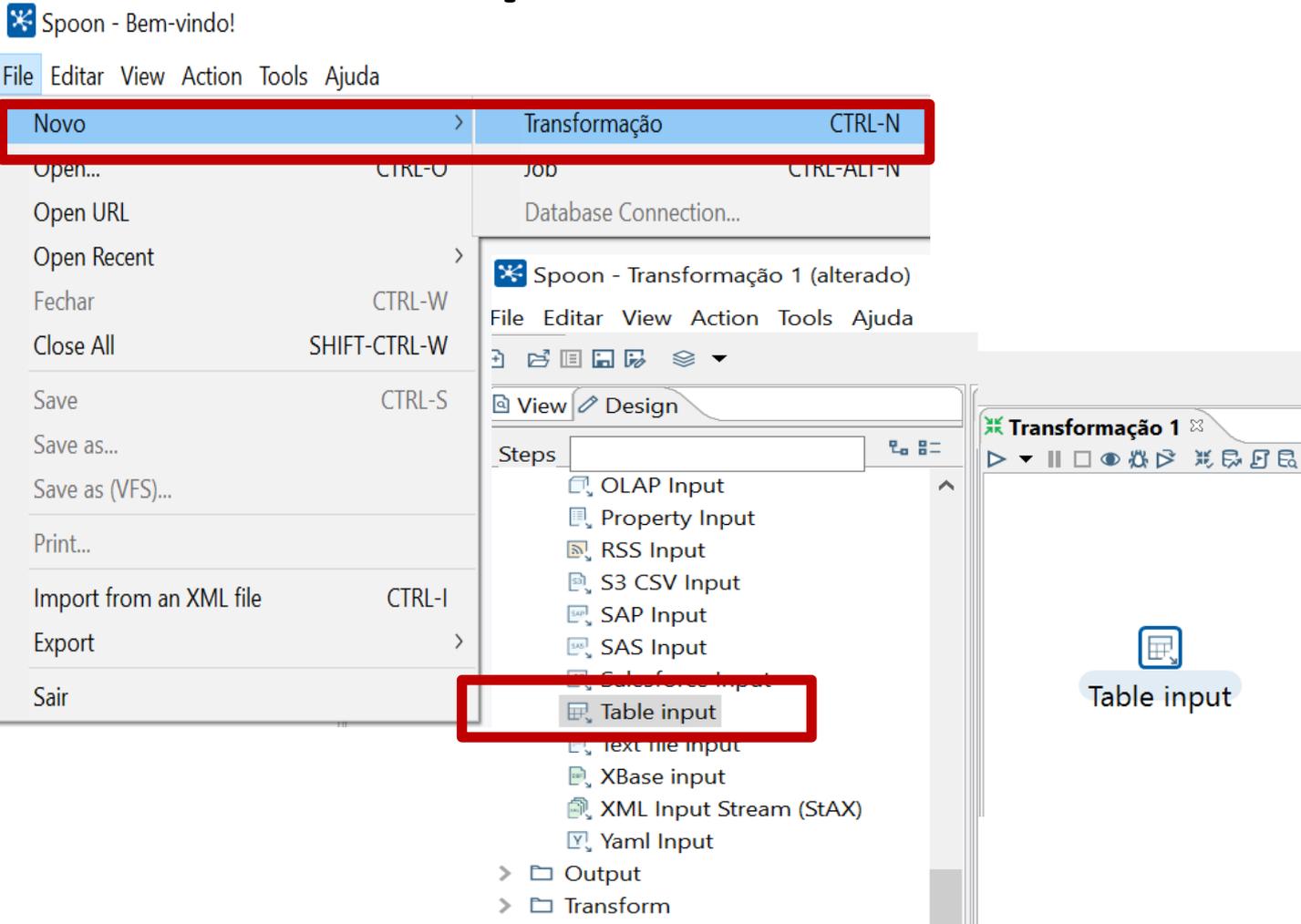
Continuação dos Exercícios da folha 6.



Vamos a carregar a dimensão vinhos.

5. Criamos uma nova transformação: *File – Novo - Transformação*

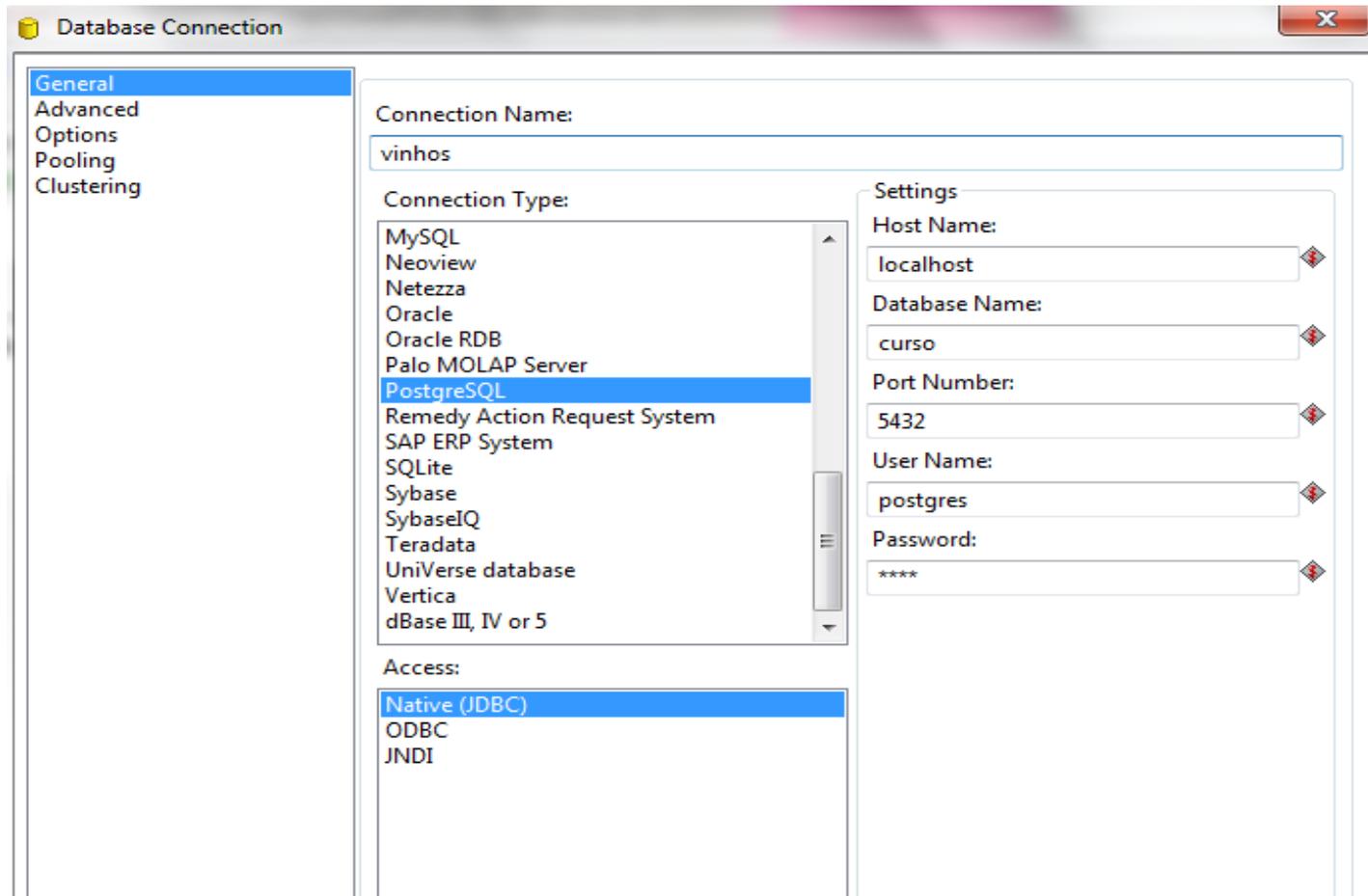
Suite BI Open Pentaho: Data Integration



6. Carregar a tabela DI_VINHOS. Criamos uma nova transformação. Incluímos o componente “*Table Input*” e configuramos o componente. Primeiro criamos uma nova conexão.

Suite BI Open Pentaho: Data Integration

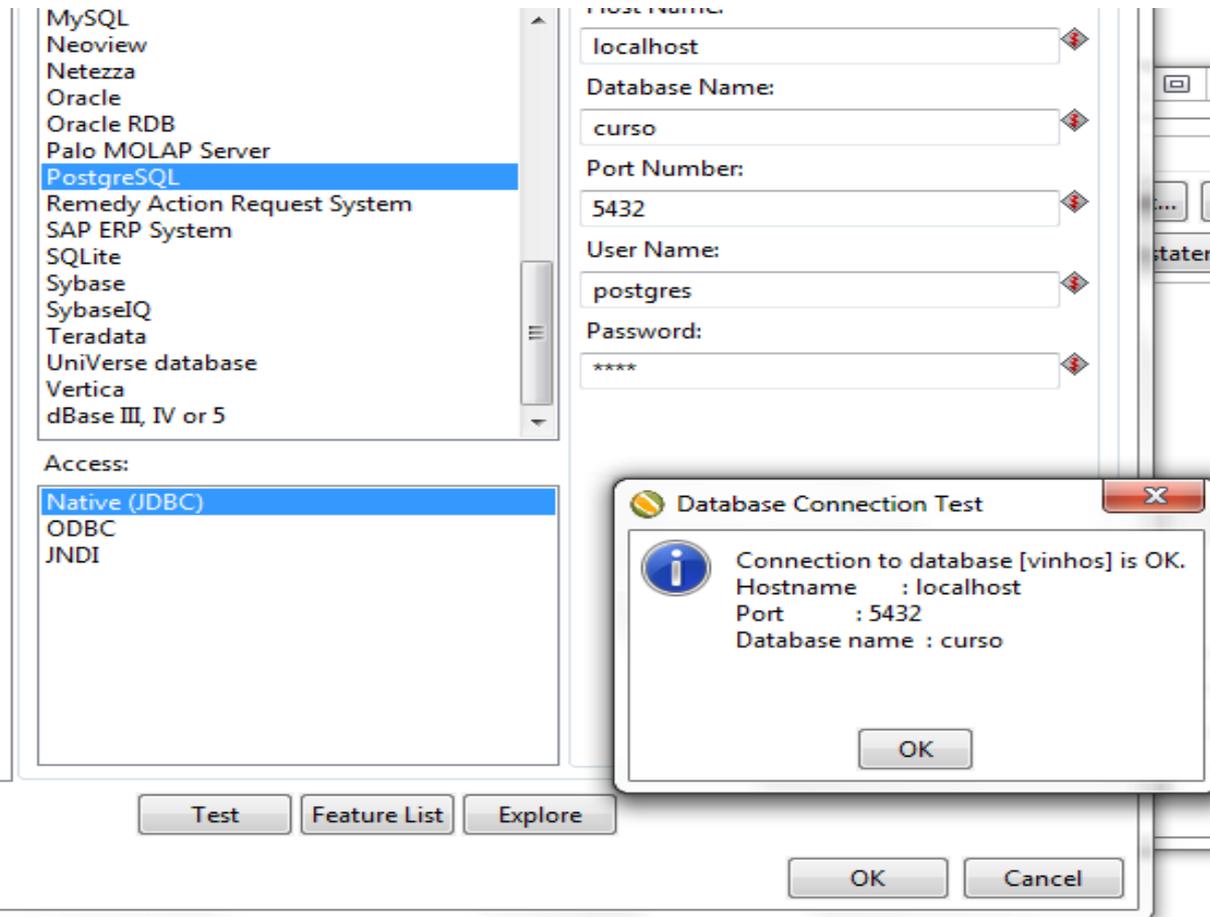
Exercícios da folha 6.



7. O servidor será “localhost”, a base de dados “curso”, login/senha do banco e o nome da conexão (nome fantasia) “vinhos”

Suite BI Open Pentaho: Data Integration

Exercícios da folha 6.



8. Testamos a conexão.

Suite BI Open Pentaho: Data Integration

Letura de Tabela

Nome do Step Table input

Connection **vinhos** os

SQL

```
SELECT * FROM RELACIONAL_VINHOS.EN_VINHOS
```

Linha 1 Coluna 41

Enable lazy conversion

Replace variables in script?

Insert data from step

Executar para cada linha?

Tamanho l mite 0

Help

OK

Preview

Cancela

Examine preview data

Rows of step: Table input (20 rows)

#	codigo_vinho	nme_vinho
1	1	Almad�n
2	2	Amadeu
3	3	Carta Vieja
4	4	Casa de Bento
5	5	Casa Valduga
6	6	Concha y Toro
7	7	Dal Pizzol
8	8	Do Lugar
9	9	Don Laurindo
1.	10	Luigi Bosca
1.	11	Miolo
1.	12	Navarro Correa
1.	13	San Conrado
1.	14	Santa Ana
1.	15	Santa Carolina
1.	16	Santa Helena
1.	17	Santa Rosa
1.	18	Terrazas
1.	19	Terrazas

Fecha

Show Log

9. Configuramos o SQL e efetuamos um *preview* para ver o resultado.

Suite BI Open Pentaho: Data Integration

Output

- Automatic Documentation Output
- Delete
- Insert / Update**
- JSON Output
- LDAP Output
- Microsoft Access Output
- Microsoft Excel Output

Table input → Insert / Update

Step name: Gravar_Vinhos

Connection: vinhos

Target schema: dwvinhos_pdi

Target table: dj_vinhos

Commit size: 100

Don't perform any updates:

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

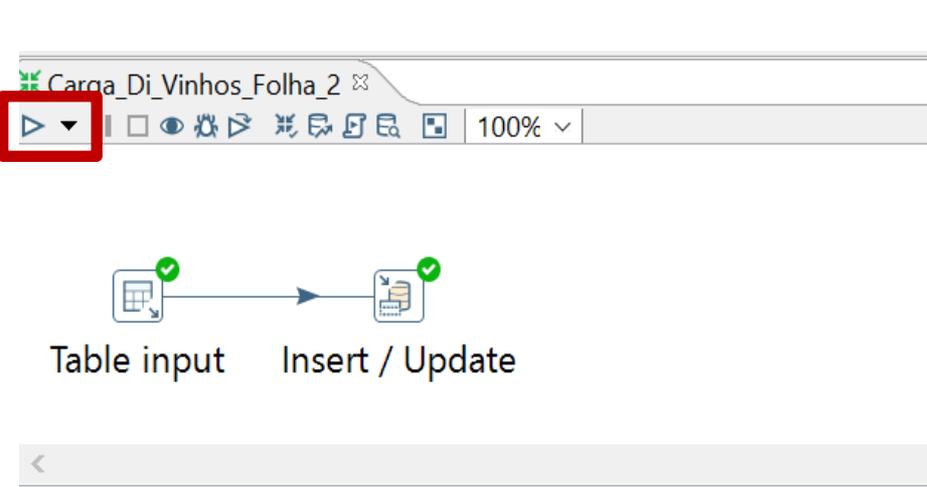
#	Table field	Comparator	Stream field1	Stream field2
1	cod_vinho	=	codigo_vinho	

Update fields:

#	Table field	Stream field	Update
1	cod_vinho	codigo_vinho	Y
2	nme_vinho	nme_vinho	Y

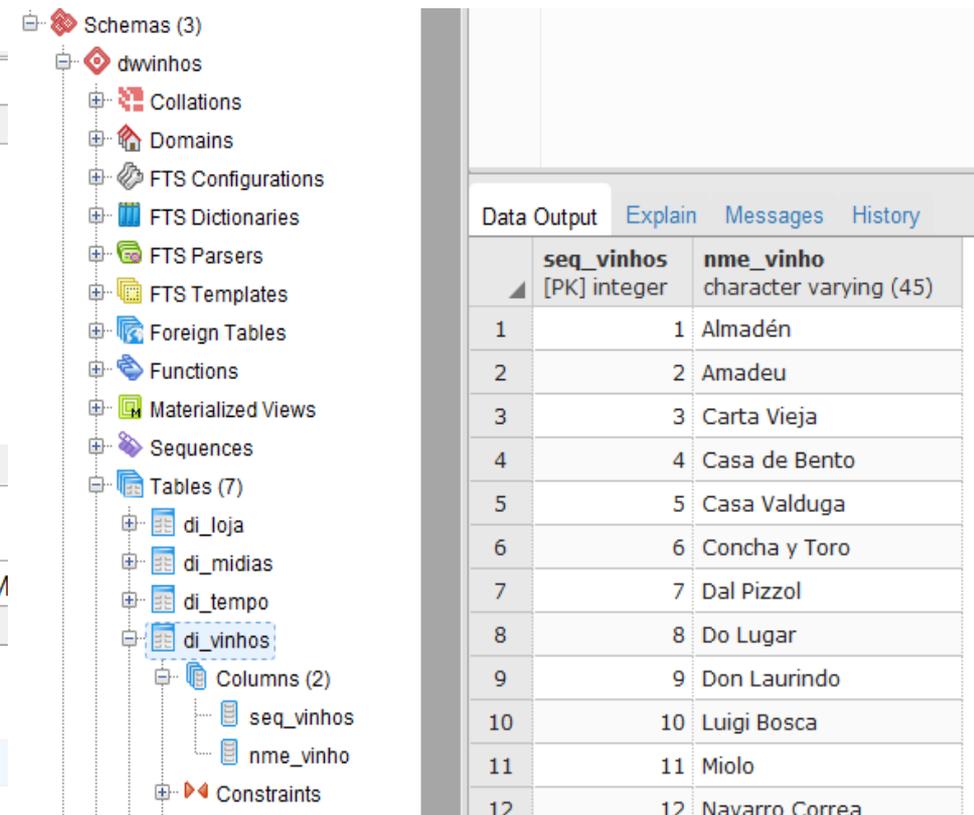
10. Incluímos na transformação o componente de “*Insert/Update*” e configuramos o componente com a tabela de saída.

Suite BI Open Pentaho: Data Integration



The screenshot shows the Pentaho Data Integration (PDI) interface. At the top, a taskbar displays the file name 'Carga_Di_Vinhos_Folha_2'. Below it, a toolbar contains various icons, with a red box highlighting the 'Run' button (a play icon). The main workspace shows a data flow diagram with two steps: 'Table input' and 'Insert / Update', both marked with green checkmarks. Below the diagram, the 'Execution Results' section is visible, showing a table with execution metrics for each step.

#	Nome do step	Copia nr	Lidos	escritos	Entrada	Saída
1	Table input	0	0	20	20	0
2	Insert / Update	0	20	20	20	20



The screenshot shows the Pentaho Data Warehouse (PDW) schema browser. The 'Schemas (3)' tree is expanded to show the 'dwinhos' schema. Under 'Tables (7)', the 'di_vinhos' table is selected. The 'Columns (2)' section shows the table structure:

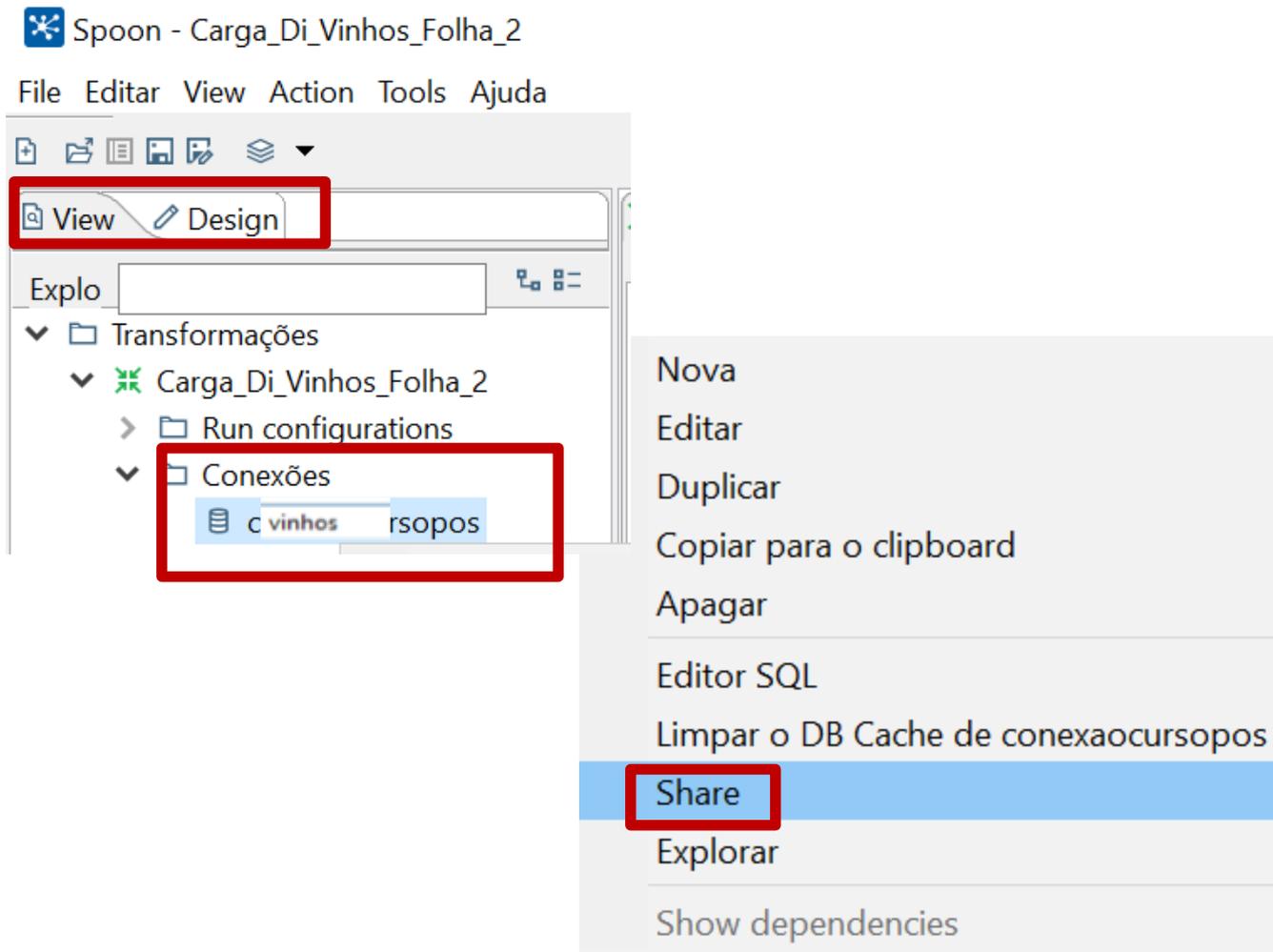
seq_vinhos	nme_vinho
[PK] integer	character varying (45)

Below the schema browser, a 'Data Output' window displays the data from the 'di_vinhos' table:

seq_vinhos	nme_vinho
1	Almadén
2	Amadeu
3	Carta Vieja
4	Casa de Bento
5	Casa Valduga
6	Concha y Toro
7	Dal Pizzol
8	Do Lugar
9	Don Laurindo
10	Luigi Bosca
11	Miolo
12	Navarro Correa

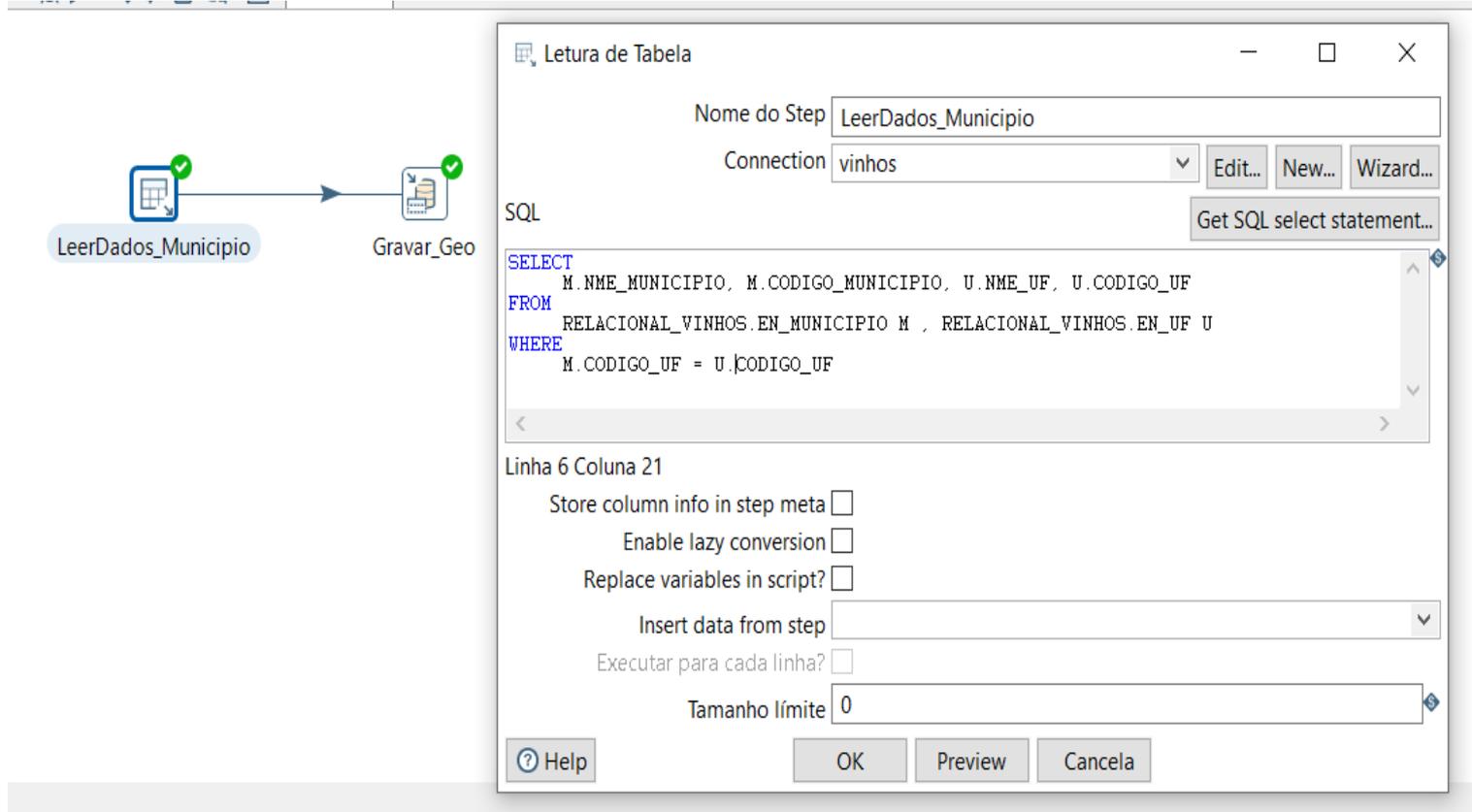
11. Salvamos, executamos a transformação e verificamos o resultado na base de dados.

Suite BI Open Pentaho: Data Integration



12. Na aba “View”, selecionamos a opção “Conexões” e em “vinhos”, fazemos um “share” para compartilhar a conexão com outras transformações.

Suite BI Open Pentaho: Data Integration



The screenshot displays the configuration for a 'Letura de Tabela' (Table Input) step in Pentaho Data Integration. The step is named 'LeerDados_Municipio' and is connected to a 'vinhos' connection. The SQL query is as follows:

```
SELECT
M.NME_MUNICIPIO, M.CODIGO_MUNICIPIO, U.NME_UF, U.CODIGO_UF
FROM
RELACIONAL_VINHOS.EN_MUNICIPIO M, RELACIONAL_VINHOS.EN_UF U
WHERE
M.CODIGO_UF = U.CODIGO_UF
```

The dialog also includes several options for step configuration:

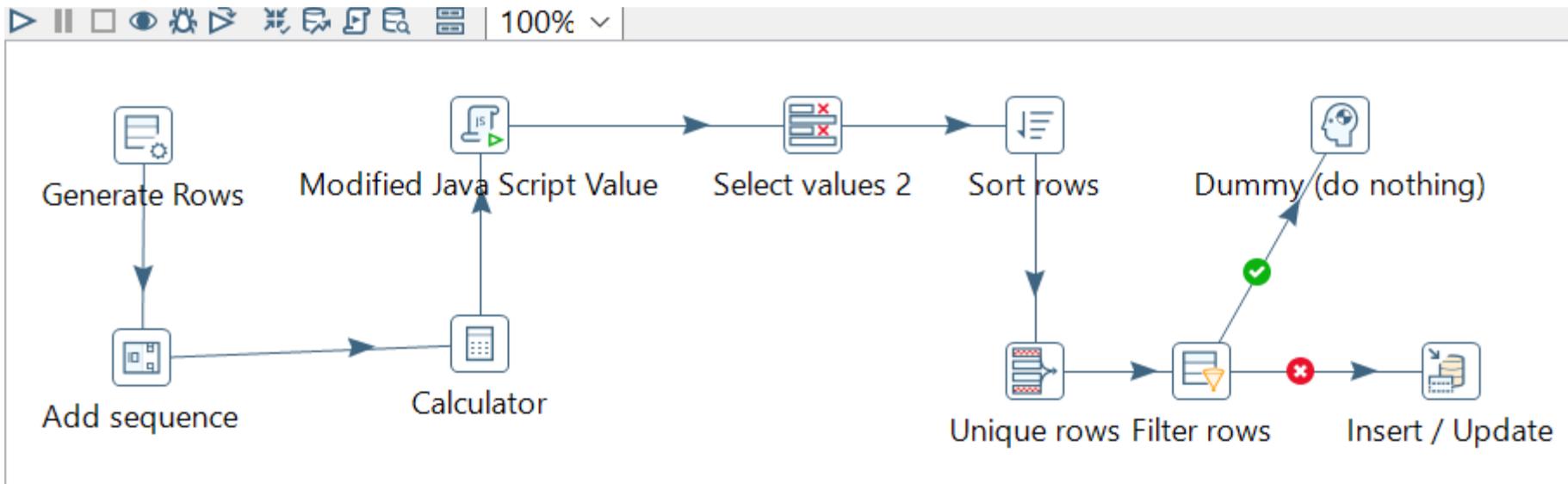
- Store column info in step meta:
- Enable lazy conversion:
- Replace variables in script?:
- Insert data from step:
- Executar para cada linha?:
- Tamanho limite: 0

The dialog also features a 'Get SQL select statement...' button and 'Help', 'OK', 'Preview', and 'Cancela' buttons at the bottom.

13. Criamos outra transformação para carregar a tabela DI_GEO e DI_LOJA. Incluímos o componente “*Table Input*”. Escolhemos a conexão e configuramos o SQL. Fazer o *preview* para verificar a saída.³¹

Suite BI Open Pentaho: Data Integration

Folha 6: DI_Tempo



1. Transformação a ser construída para carregar a dimensão Tempo.

Suite BI Open Pentaho: Data Integration

Generate Rows Modified Java Script Value Select values 2 Sort rows Dummy (do nothing)

Generate Rows

Nome do Step Generate Rows

Limit 5000

Never stop generating rows

Interval in ms (delay)

Current row time field name

Previous row time field name

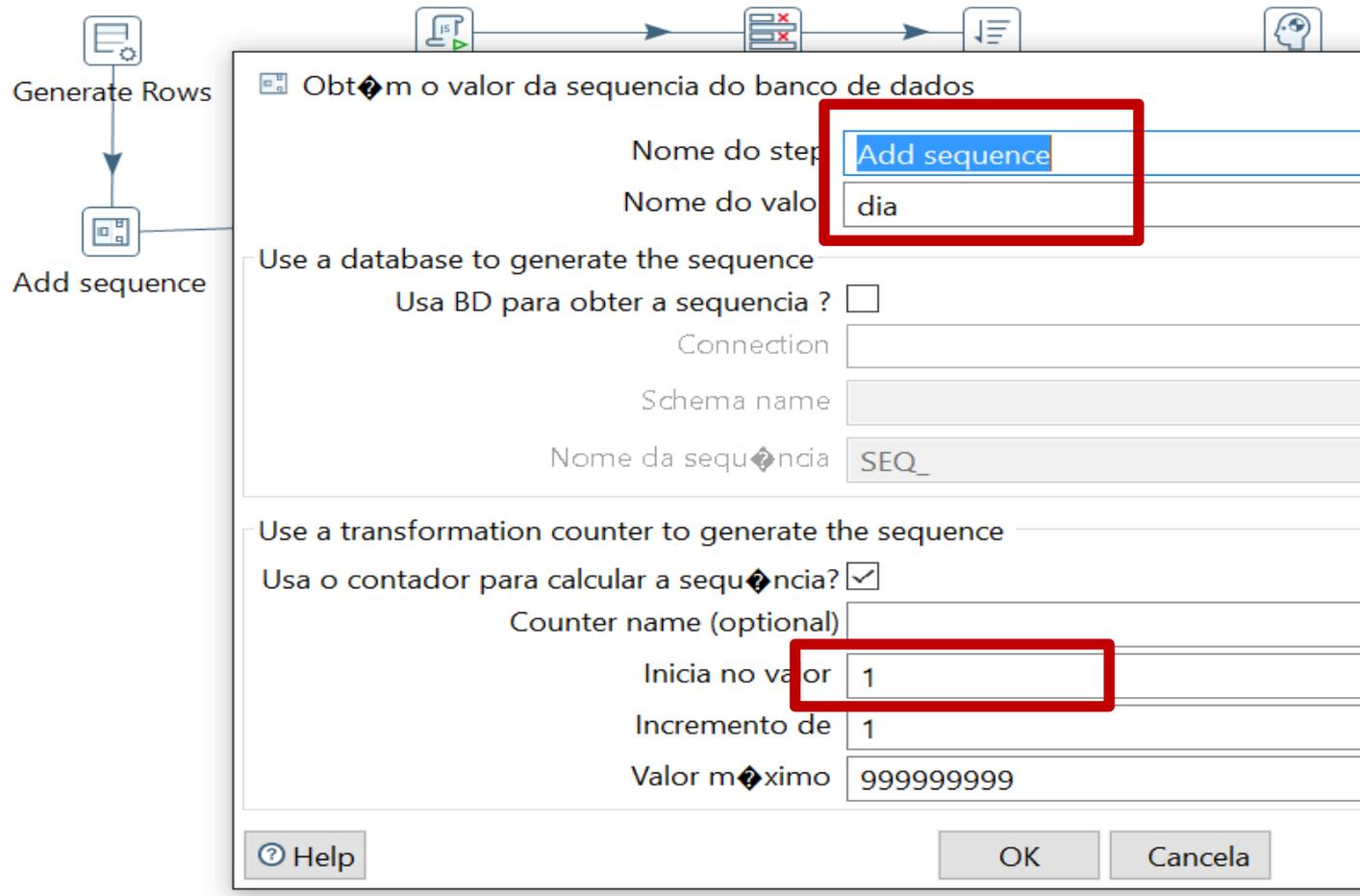
Fields :

#	Nome	Tipo	Formato	Tamanho	Precisão	Moeda	Decimal	Grupo	Valor	Set empty string?
1	Data	Date	yyyy-MM-dd						2009-01-01	N

Help OK Preview Cancela

2. Configuração do componente “*Generate rows*”. Definimos a variável **Data** do tipo *date* e formato *yyy-mm-dd* iniciando em 2009-01-01. Geramos 5.000 linhas.

Suite BI Open Pentaho: Data Integration



3. Configuração do componente “*Add sequence*”. Adicionamos a variável **dia**, iniciando em 1.

Suite BI Open Pentaho: Data Integration

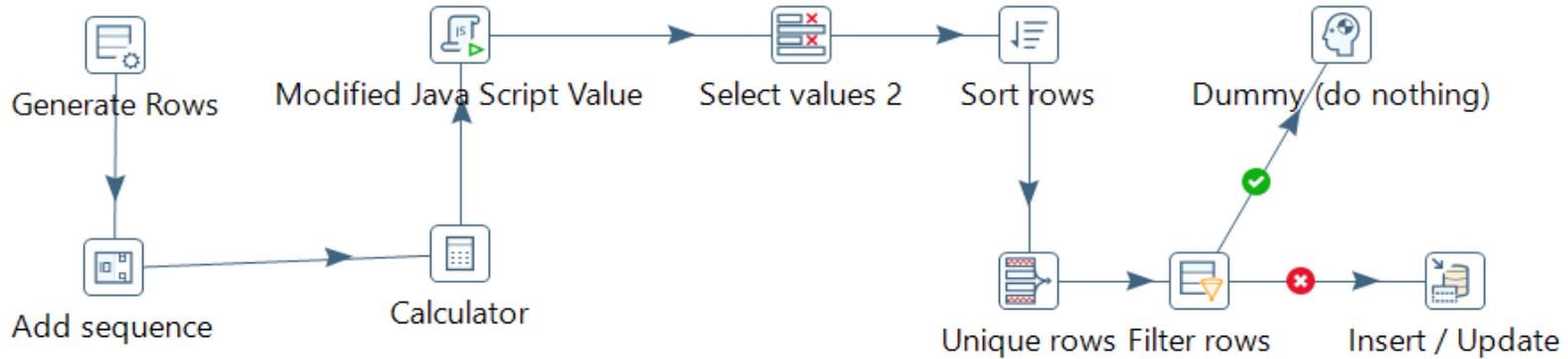
Examine preview data

Rows of step: Add sequence (1000 rows)

#	Data	dia
1	2009-01-01	1
2	2009-01-01	2
3	2009-01-01	3
4	2009-01-01	4
5	2009-01-01	5
6	2009-01-01	6
7	2009-01-01	7
8	2009-01-01	8
9	2009-01-01	9
10	2009-01-01	10
11	2009-01-01	11
12	2009-01-01	12
13	2009-01-01	13
14	2009-01-01	14

4. Fazemos um *preview* para visualizar a transformação.

Suite BI Open Pentaho: Data Integration



Calculadora

Nome do Step:

Campos:

#	Novo campo	Cálculo	Campo A	Campo B	Campo C	Tipo do valor	Tam
1	DataCalculo	Data A + B Dias	Data	dia		None	
2	Ano	Ano da data A	DataCalculo			None	
3	Mes	Mes da data A	DataCalculo			None	
4	Trimestre	Quarter of date A	DataCalculo			None	

5. Configuração do componente “*Calculadora*” para definir as variáveis **ano**, **mes** e **trimestre**. Para isso, necessitamos definir a variável auxiliar **datacalculo**.

Suite BI Open Pentaho: Data Integration

Generate Rows → Modified Java Script Value → Select values 2 → Sort rows → Dummy (do nothing)

Script Values / Mod

Step name: Modified Java Script Value

Java script functions :

- Transform Scripts
- Transform Constants
- Transform Functions
- Input fields
 - Data
 - dia
 - DataCalculo
 - Ano
 - Mes
 - Trimestre
- Output fields
 - Please use the 'Repla

Java script :

```
Script 1 x  
//Script here  
var semestre = (Mes<7)? 1:2;
```

Linern: 0
Compatibility mode? Optimization level | 9

#	Fieldname	Rename to	Type	Length	Precision	Replace value 'Fieldname' or 'Rename to'
1	semestre		Integer	16	2	N

6. Configuração do componente “*Java Script*” para definir a variável **semestre**.

Suite BI Open Pentaho: Data Integration

Examine preview data

Rows of step: Modified Java Script Value (1000 rows)

#	Data	dia	DataCalculo	Ano	Mes	Trimestre	semestre
1	2009-01-01	1	2009/01/02 00:00:00.000	2009	1	1	1
2	2009-01-01	2	2009/01/03 00:00:00.000	2009	1	1	1
3	2009-01-01	3	2009/01/04 00:00:00.000	2009	1	1	1
4	2009-01-01	4	2009/01/05 00:00:00.000	2009	1	1	1
5	2009-01-01	5	2009/01/06 00:00:00.000	2009	1	1	1
6	2009-01-01	6	2009/01/07 00:00:00.000	2009	1	1	1
7	2009-01-01	7	2009/01/08 00:00:00.000	2009	1	1	1
8	2009-01-01	8	2009/01/09 00:00:00.000	2009	1	1	1
9	2009-01-01	9	2009/01/10 00:00:00.000	2009	1	1	1
10	2009-01-01	10	2009/01/11 00:00:00.000	2009	1	1	1
11	2009-01-01	11	2009/01/12 00:00:00.000	2009	1	1	1
12	2009-01-01	12	2009/01/13 00:00:00.000	2009	1	1	1
13	2009-01-01	13	2009/01/14 00:00:00.000	2009	1	1	1
14	2009-01-01	14	2009/01/15 00:00:00.000	2009	1	1	1
15	2009-01-01	15	2009/01/16 00:00:00.000	2009	1	1	1
16	2009-01-01	16	2009/01/17 00:00:00.000	2009	1	1	1
17	2009-01-01	17	2009/01/18 00:00:00.000	2009	1	1	1

Close Stop Get more rows

7. Fazemos um *preview* para visualizar a transformação.

Suite BI Open Pentaho: Data Integration

 Select / Rename values

Select & Alter	Remove	Meta-data		
Fields :				
# ^	Fieldname	Rename to	Length	Precision
1	Data			
2	dia			
3	DataCalculo			
4	Ano			
5	Mes			
6	Trimestre			
7	semestre			

 Select / Rename values

Select & Alter	Remove	Meta-data
Fields to remove :		
# ^	Fieldname	
1	Data	
2	dia	
3	DataCalculo	

8. Configuração do componente “*Select Values*” para remover as variáveis que não serão carregadas na base de dados.

Suite BI Open Pentaho: Data Integration

Examine preview data

Rows of step: Select values 2 (1000 rows)

#	Ano	Mes	Trimestre	semestre
1	2009	1	1	1
2	2009	1	1	1
3	2009	1	1	1
4	2009	1	1	1
5	2009	1	1	1
6	2009	1	1	1
7	2009	1	1	1
8	2009	1	1	1
9	2009	1	1	1
10	2009	1	1	1
11	2009	1	1	1
12	2009	1	1	1
13	2009	1	1	1
14	2009	1	1	1
15	2009	1	1	1

9. Fazemos um *preview* para visualizar a transformação.

Suite BI Open Pentaho: Data Integration

The image shows two configuration windows from the Pentaho Data Integration suite. The 'Sort rows' window is on the left, and the 'linhas únicas' (Unique rows) window is on the right.

Sort rows configuration:

- Nome do Step: 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?

Fields to compare on (Sort rows):

#	Fieldname	Ascending	Case sensitive compare?	Presorted?
1	Ano	S	N	N
2	Mes	S	N	N

linhas únicas configuration:

- Nome do Step: Unique rows
- Add counter to output? Counter field:
- Redirect duplicate row Error description:

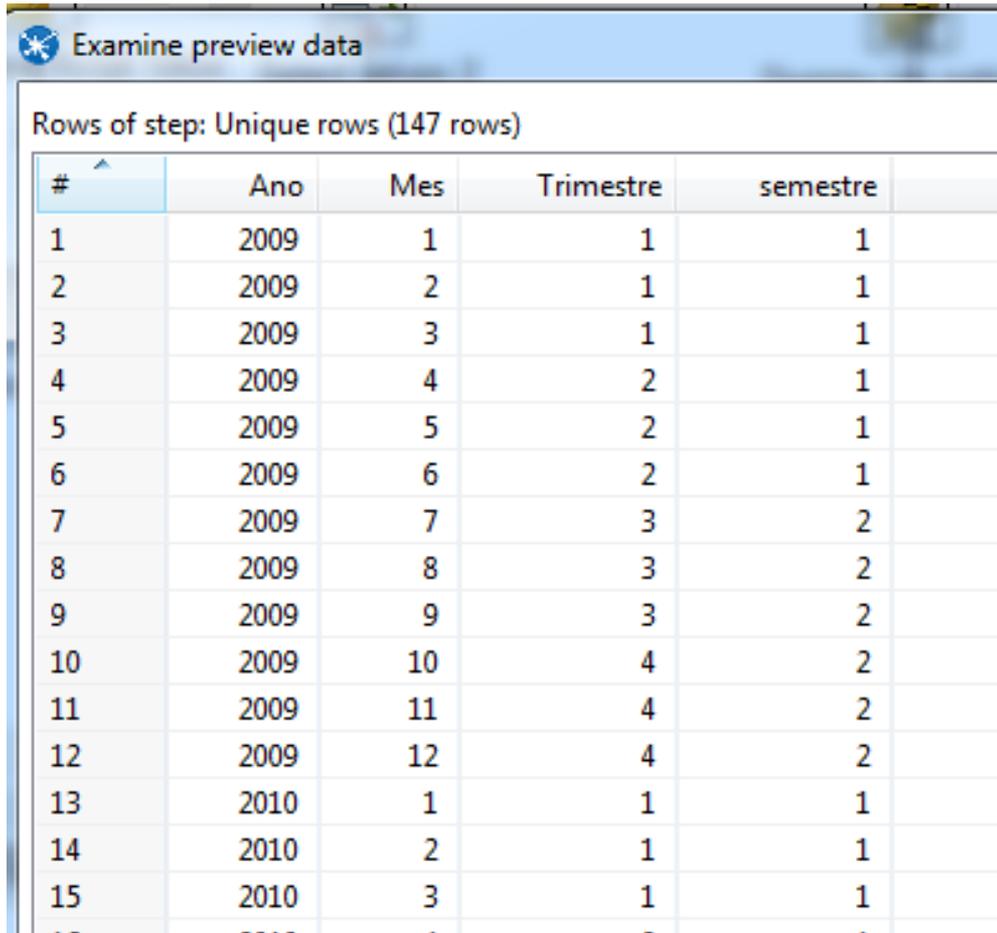
Fields to compare on (linhas únicas):

#	Fieldname	Ignore case
1	Ano	N
2	Mes	N

Buttons: Help, OK, Cancela, Get

10. Configuração dos componentes “Sort” e “Unique” para eliminar os registros duplicados.

Suite BI Open Pentaho: Data Integration



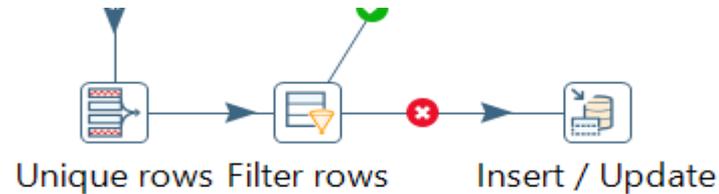
Examine preview data

Rows of step: Unique rows (147 rows)

#	Ano	Mes	Trimestre	semestre
1	2009	1	1	1
2	2009	2	1	1
3	2009	3	1	1
4	2009	4	2	1
5	2009	5	2	1
6	2009	6	2	1
7	2009	7	3	2
8	2009	8	3	2
9	2009	9	3	2
10	2009	10	4	2
11	2009	11	4	2
12	2009	12	4	2
13	2010	1	1	1
14	2010	2	1	1
15	2010	3	1	1

11. Fazemos um *preview* para visualizar a transformação.

Suite BI Open Pentaho: Data Integration



Filter rows

Step name: Filter rows

Send 'true' data to step: Dummy (do nothing)

Send 'false' data to step: Insert / Update

The condition:

Ano > 2020 (Integer)

Help OK Cancela

12. Configuração do componente “Filter” para eliminar as instâncias acima do ano 2020 e gravar as instâncias anteriores de 01/2009 a 12/2020.

Suite BI Open Pentaho: Data Integration

Insert / Update

Step name: Insert / Update

Connection: conexaocursopos [Edit... New... Wizard...]

Target schema: dwvinhos [Navega...]

Target table: di_tempo [Browse...]

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	ano	=	Ano		Get fields
2	mes	=	Mes		

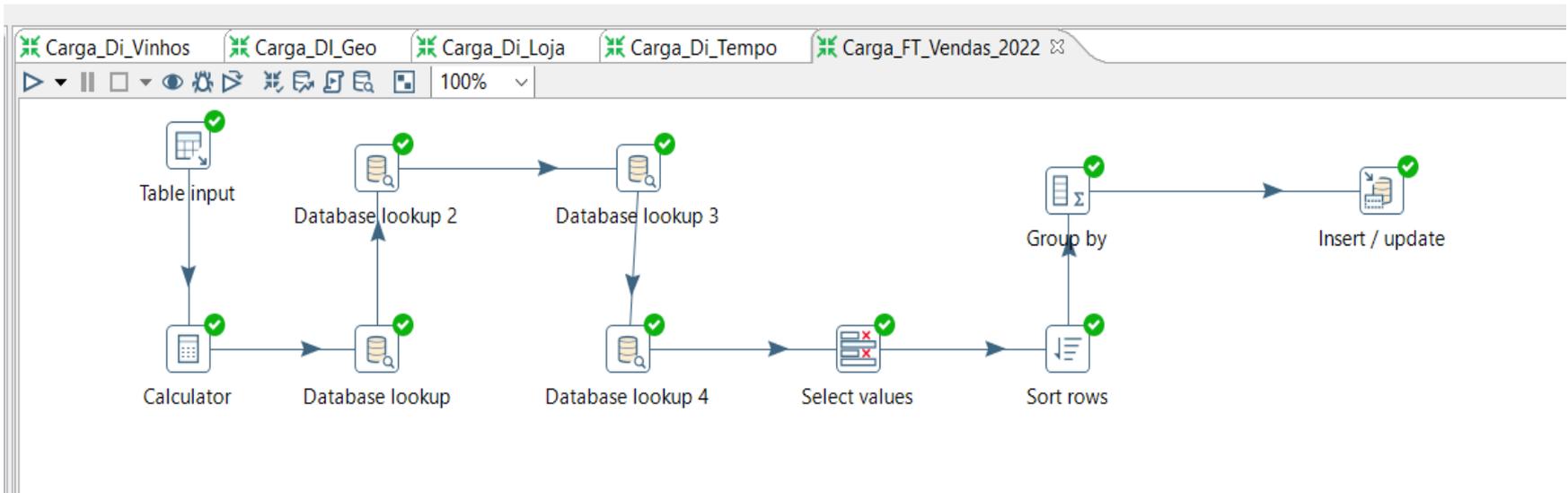
Update fields:

#	Table field	Stream field	Update	
1	ano	Ano	Y	Get update fields Edit mapping
2	mes	Mes	Y	
3	trimestre	Trimestre	Y	
4	semestre	Semestre	Y	

13. Configuração do componente "Insert/Update".

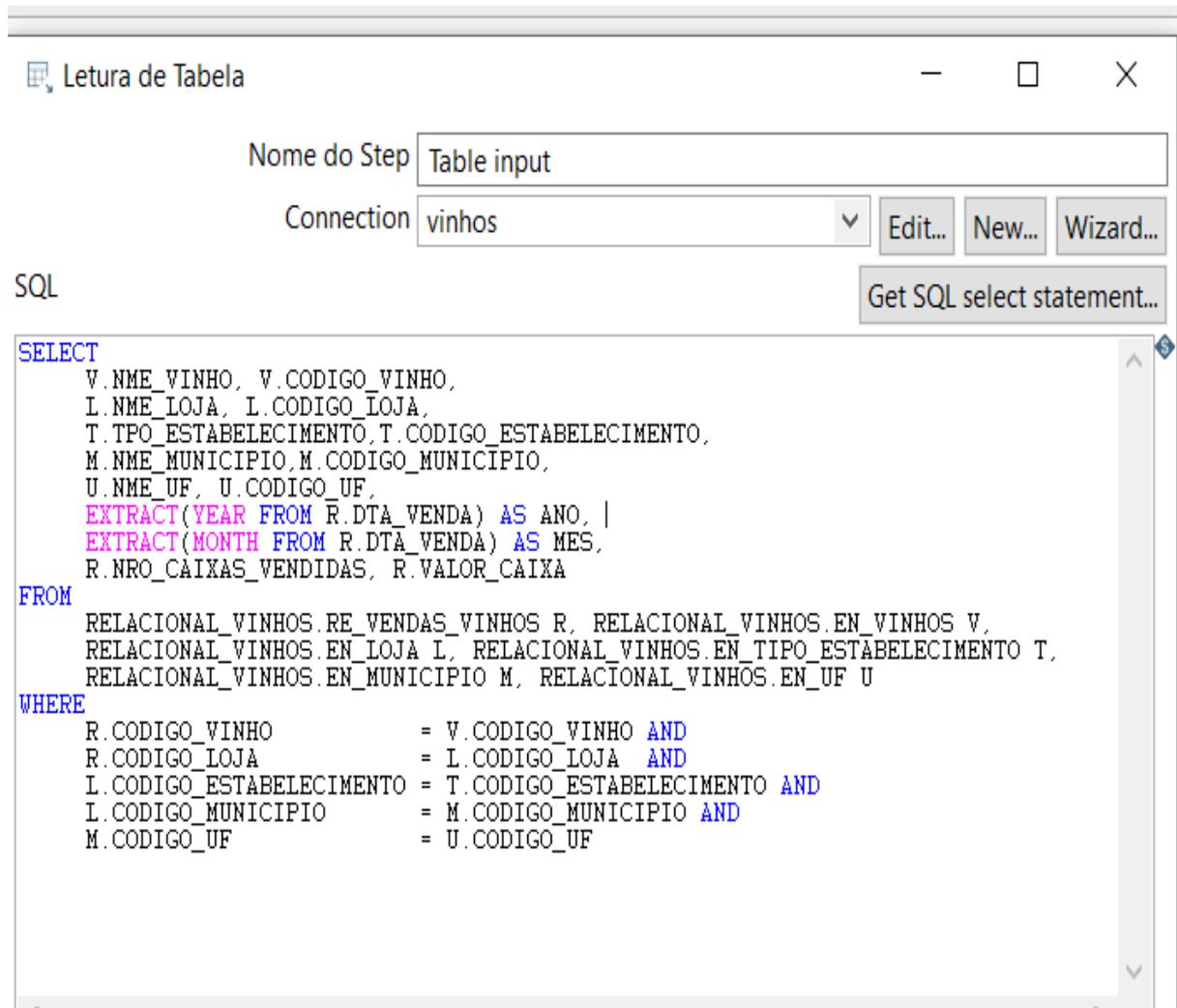
Suite BI Open Pentaho: Data Integration

Folha 6: Tabela Fato



1. Verificar se a conexão do componente “*Table Input*” está funcionando.

Suite BI Open Pentaho: Data Integration



Letura de Tabela

Nome do Step: Table input

Connection: vinos

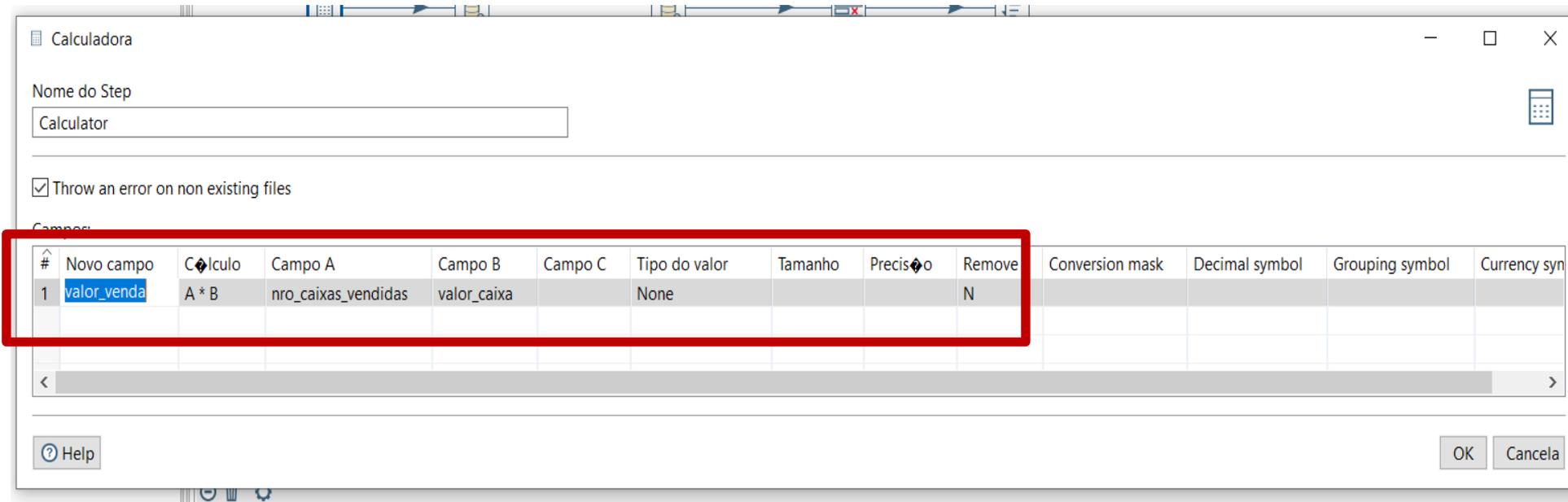
Buttons: Edit..., New..., Wizard...

SQL: Get SQL select statement...

```
SELECT
  V.NME_VINHO, V.CODIGO_VINHO,
  L.NME_LOJA, L.CODIGO_LOJA,
  T.TPO_ESTABELECIMENTO, T.CODIGO_ESTABELECIMENTO,
  M.NME_MUNICIPIO, M.CODIGO_MUNICIPIO,
  U.NME_UF, U.CODIGO_UF,
  EXTRACT(YEAR FROM R.DTA_VENDA) AS ANO, |
  EXTRACT(MONTH FROM R.DTA_VENDA) AS MES,
  R.NRO_CAIXAS_VENDIDAS, R.VALOR_CAIXA
FROM
  RELACIONAL_VINHOS.RE_VENDAS_VINHOS R, RELACIONAL_VINHOS.EN_VINHOS V,
  RELACIONAL_VINHOS.EN_LOJA L, RELACIONAL_VINHOS.EN_TIPO_ESTABELECIMENTO T,
  RELACIONAL_VINHOS.EN_MUNICIPIO M, RELACIONAL_VINHOS.EN_UF U
WHERE
  R.CODIGO_VINHO      = V.CODIGO_VINHO AND
  R.CODIGO_LOJA       = L.CODIGO_LOJA  AND
  L.CODIGO_ESTABELECIMENTO = T.CODIGO_ESTABELECIMENTO AND
  L.CODIGO_MUNICIPIO  = M.CODIGO_MUNICIPIO AND
  M.CODIGO_UF         = U.CODIGO_UF
```

2. No componente de “*Table input*”, estamos pegando todos os dados necessários do esquema relacional.

Suite BI Open Pentaho: Data Integration



3. Configuração do componente “Calculadora” para criar a medida que está faltando.

Suite BI Open Pentaho: Data Integration

Lookup de valor do banco de dados

Nome do Step: Database lookup

Connection: vinhos

Lookup schema: dwvinhos_pdi

Tabela Lookup: di_vinhos

Tamanho do cache em linhas (0=cache total): 0

Load all data from table:

A chave(s) para examinar o valor(s):

#	Campo da tabela	Comparador	Campo1	Campo2
1	cod_vinho	=	codigo_vinho	

Valores a serem retornados da tabela lookup:

#	campo	Novo nome	Default	Tipo
1	seq_vinhos			Integer

on Results

4. Os componentes “*Database lookup*” são necessários para obter as chaves (*seq_*) das dimensões.

Suite BI Open Pentaho: Data Integration

Select values

Select & Alter Remove Meta-data

Fields :

#	Fieldname
1	nme_vinho
2	codigo_vinho
3	nme_loja
4	codigo_loja
5	tpo_estabelecimento
6	codigo_estabelecimento
7	nme_municipio
8	codigo_municipio
9	nme_uf
10	codigo_uf
11	ano
12	mes
13	nro_caixas_vendidas
14	valor_caixa
15	valor_venda
16	seq_vinhos
17	seq_loja
18	seq_tempo
19	seq_geo

Select values

Step name Select values

Select & Alter Remove Meta-data

Fields to remove :

#	Fieldname
1	nme_vinho
2	codigo_vinho
3	nme_loja
4	codigo_loja
5	tpo_estabelecimento
6	codigo_estabelecimento
7	nme_municipio
8	codigo_municipio
9	nme_uf
10	codigo_uf
11	ano
12	mes

5. O componente “*Select values*” é utilizado para ficar somente com os atributos necessários para gravar na tabela fato. Incluir o atributo *desconto*.

Suite BI Open Pentaho: Data Integration

Sort rows

Nome do Step: Sort rows

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

TMP-file prefix: out

Sort size (rows in memory): 1000000

Free memory threshold (in %):

This is the free memory threshold in percent of the maximum JVM memory. We will increase the sort buffer until we reach this limit.

Only pass unid

Fields:

#	Fieldname	Ascending	Case sensitive compare?	Sort based on current locale?	Collator Strength	Presorte
1	seq_vinhos	S				
2	seq_loja	S				
3	seq_tempo	S				
4	seq_geo	S				

6. Configuração do componente “Sort rows”: ordena os registros pelos campos indicados.

Suite BI Open Pentaho: Data Integration

Step name

Include all rows?

Temporary files directory

TMP-file prefix

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	seq_vinhos
2	seq_loja
3	seq_tempo
4	seq_geo

Aggregates :

#	Name	Subject	Type
1	nro_caixas_vendidas	nro_caixas_vendidas	Sum
2	valor_caixa	valor_caixa	Average (Mean)
3	valor_venda	valor_venda	Sum

7. Configuração do componente “Group By”.

Suite BI Open Pentaho: Data Integration

Insert / update

Step name:

Connection:

Target schema:

Target table:

Commit size:

Don't perform any updates:

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

#	Table field	Comparator	Stream field1	Stream field2
1	seq_vinhos	=	seq_vinhos	
2	seq_loja	=	seq_loja	
3	seq_tempo	=	seq_tempo	
4	seq_geo	=	seq_geo	

Update fields:

#	Table field	Stream field	Update
1	seq_vinhos	seq_vinhos	Y
2	seq_loja	seq_loja	Y
3	seq_tempo	seq_tempo	Y
4	seq_geo	seq_geo	Y
5	nro_caixas_vendidas	nro_caixas_vendidas	Y
6	valor_caixa	valor_caixa	Y
7	valor_venda	valor_venda	Y

8. Configuração do componente “*Insert/Update*”. Execute a transformação

Suite BI Open Pentaho: Data Integration

Execution Results

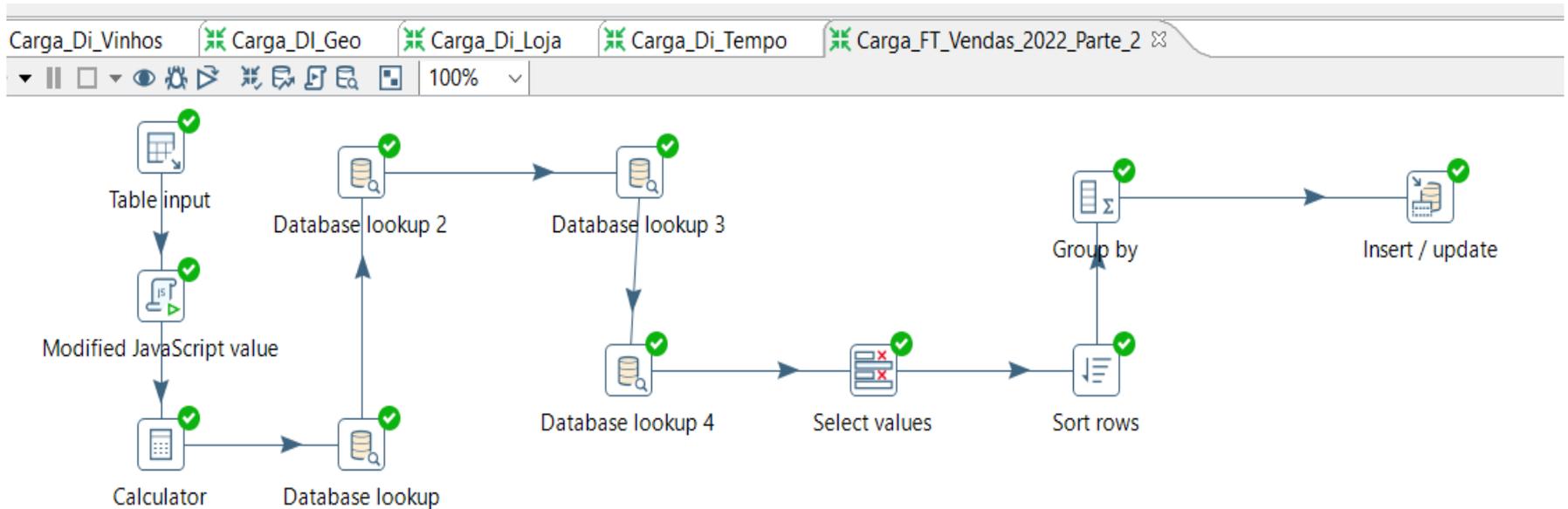
Execution History | Logging | Step Metrics | Performance Graph | Metrics | Preview data

#	Nome do step	Copia nr	Lidos	escritos	Entrada	Saída	Atualizados	Rejected	Erros	Ativo	Tempo	Velocidade (r/s)	Pri/ent/sai
1	Table input	0	0	2155	2155	0	0	0	0	Finished	0.1s	32.164	-
2	Database lookup	0	2155	2155	2155	0	0	0	0	Finished	1.2s	1.822	-
3	Database lookup 2	0	2155	2155	2155	0	0	0	0	Finished	1.6s	1.344	-
4	Database lookup 3	0	2155	2155	2155	0	0	0	0	Finished	1.7s	1.304	-
5	Select values	0	2155	2155	0	0	0	0	0	Finished	1.7s	1.300	-
6	Calculator	0	2155	2155	0	0	0	0	0	Finished	1.7s	1.297	-
7	Sort rows	0	2155	2155	0	0	0	0	0	Finished	1.7s	1.286	-
8	Group by	0	2155	667	0	0	0	0	0	Finished	1.7s	1.269	-
9	Select values 2	0	667	667	0	0	0	0	0	Finished	1.7s	386	-
10	Insert / Update	0	667	667	667	667	0	0	0	Finished	2.7s	247	-

9. Transformação executada.

Suite BI Open Pentaho: Data Integration

Exercícios da folha 6: item d



Alteração da carga da tabela fato para incluir uma nova medida.

Suite BI Open Pentaho: Data Integration

Script Values / Mod

Step name Modified Java Script Value

Java script functions :

- > Transform Scripts
- > Transform Constants
- > Transform Functions
- ▼ Input fields
 - ▶ nme_vinho
 - ▶ nme_loja
 - ▶ tpo_estabelecimento
 - ▶ nme_municipio
 - ▶ nme_uf
 - ▶ ano
 - ▶ mes
 - ▶ nro_caixas_vendidas
 - ▶ valor_caixa
- ▼ Output fields
 - Please use the 'Repl

Java script :

```
Script 1 ⌵  
//Script here  
  
var desconto = 0;  
  
if (nro_caixas_vendidas >=40) desconto = 25;  
else if (nro_caixas_vendidas >=20) desconto = 10;
```

Position: 6, 25

Compatibility mode? Optimization level 9

Fields

#	Fieldname	Rename to	Type	Length	Precision	Replace value 'Fieldname' or 'Rename to'
1	desconto		Number	16	2	N

1. Configuração do componente *Java Script* para calculo da porcentagem de desconto.

Suite BI Open Pentaho: Data Integration

Select / Rename values

Select & Alter Remove Meta-data

Fields :

#	Fieldname	Rename to
1	nme_vinho	
2	nme_loja	
3	tpo_estabelecimento	
4	nme_municipio	
5	nme_uf	
6	ano	
7	mes	
8	nro_caixas_vendidas	
9	valor_caixa	
10	seq_vinhos	
11	seq_loja	
12	seq_tempo	
13	desconto	

Calculadora

Nome do Step Calculator

Campos:

#	Novo campo	Cálculo	Campo A	Campo B	Camp
1	valor_venda	$A * B$	nro_caixas_vendidas	valor_caixa	
2	valor_venda_desconto	$A - (A * B / 100)$	valor_venda	desconto	

2. Incluir o campo de **desconto** no componente *Select*, e calcular a nova medida (**Valor_venda_desconto**) no componente *Calculadora*. ⁵⁶

Suite BI Open Pentaho: Data Integration

Step name

Include all rows?

Temporary files directory

TMP-file prefix

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	seq_vinhos	
2	seq_loja	
3	seq_tempo	
4	seq_geo	

Aggregates :

#	Name	Subject	Type
1	nro_caixas_vendidas	nro_caixas_vendidas	Sum
2	valor_caixa	valor_caixa	Average (Mean)
3	valor_venda	valor_venda	Sum
4	valor_Venda_desconto	valor_Venda_desconto	Sum

3. Incluir a nova medida no componente *Group by*.

Suite BI Open Pentaho: Data Integration

Insert / update

Step name:

Connection:

Target schema:

Target table:

Commit size:

Don't perform any updates:

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

#	Table field	Comparator	Stream field1	Stream field2
1	seq_vinhos	=	seq_vinhos	
2	seq_loja	=	seq_loja	
3	seq_tempo	=	seq_tempo	
4	seq_geo	=	seq_geo	

Update fields:

#	Table field	Stream field	Update
1	seq_vinhos	seq_vinhos	Y
2	seq_loja	seq_loja	Y
3	seq_tempo	seq_tempo	Y
4	seq_geo	seq_geo	Y
5	nro_caixas_vendidas	nro_caixas_vendidas	Y
6	valor_caixa	valor_caixa	Y
7	valor_venda	valor_venda	Y
8	valor_Venda_desconto	valor_Venda_desconto	Y

4. Incluir a nova medida no componente *Insert/Update*.

Suite BI Open Pentaho: Data Integration

Exercícios da folha 6 – item e - Apagar dados

Execute SQL script

Step name

Connection

SQL script to execute. (statements separated by ;) Question marks will be replaced by arguments.

```
DELETE FROM DWVINHOS_PDI.FT_VENDAS;  
DELETE FROM DWVINHOS_PDI.DI_VINHOS;  
DELETE FROM DWVINHOS_PDI.DI_LOJA;  
DELETE FROM DWVINHOS_PDI.DI_GEO;  
DELETE FROM DWVINHOS_PDI.DI_TEMPO;
```

Line 4 Column 32

Execute for each row?

Execute as a single statement?

Variable substitution?

Bind parameters?

Quote Strings?

Parameters :

#	Field name to be used as argument
1	

Field to contain insert stats

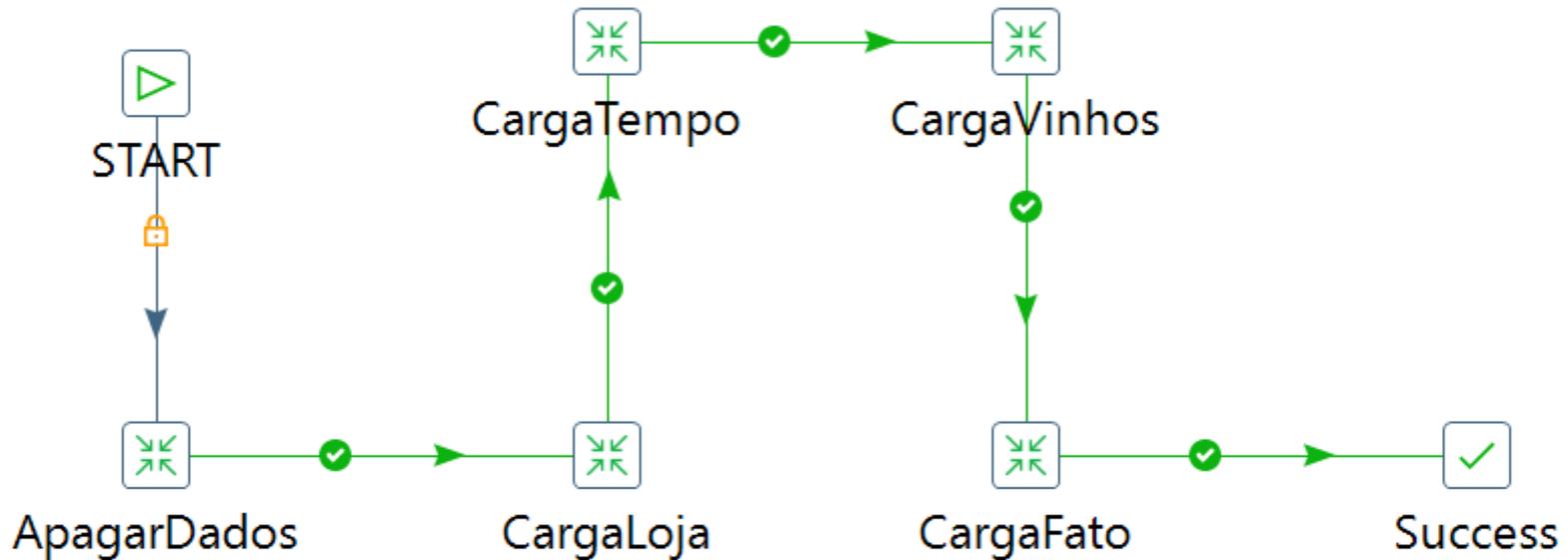
Field to contain Update stats

Field to contain Delete stats

Field to contain Read stats

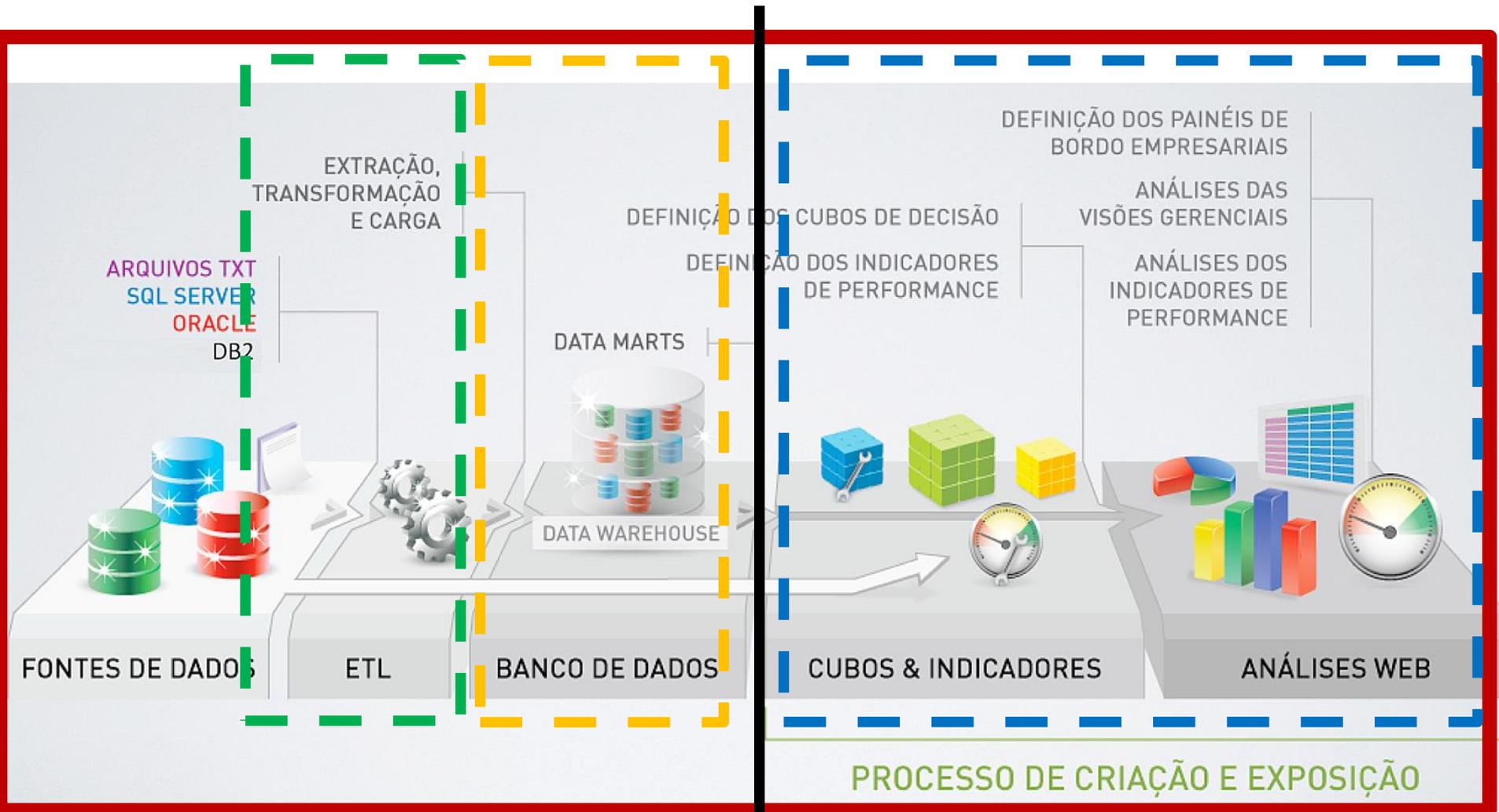
Suite BI Open Pentaho: Data Integration

Exercícios da folha 6 – item f - Criação da JOB.



Criação da *job*.

Suite BI Open Pentaho: Data Integration



Equipe técnica que desenvolve e oferece suporte ao BI

Analistas de negócio

Fundamentação da disciplina

