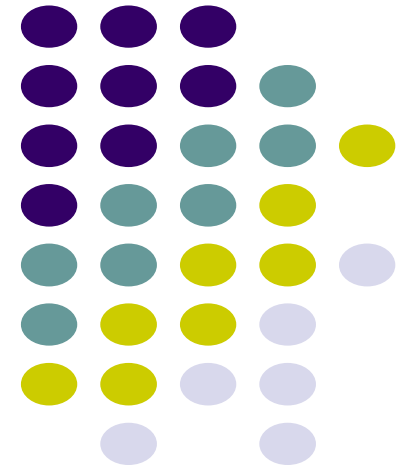




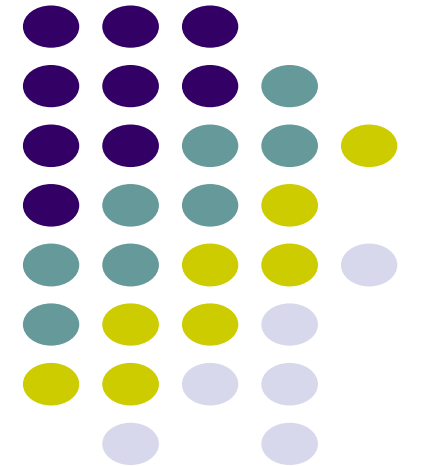
# Business Intelligence: O desenho do Data Warehouse

Vitor Valerio de Souza Campos



# Processo para criar o Data Stage e o Data Warehouse

---





# DS – Dimensão Vendedor no Data Stage

- Composta por uma chave primária (*Surrogate Key*) e as propriedades do vendedor.

| D_Funcionario |                       |
|---------------|-----------------------|
|               | Id_Funcionario (int)  |
|               | Nome (varchar(50))    |
|               | Login (varchar(50))   |
|               | LinData (date)        |
|               | LinOrig (varchar(50)) |



# DS – Dimensão Vendedor no Data Stage

- Script da tabela chamada D\_Funcionario.

```
CREATE TABLE D_Funcionario(  
    Nome varchar(50) NOT NULL,  
    Login varchar(50) NOT NULL,  
    LinData date NOT NULL,  
    LinOrig varchar(50) NOT NULL  
);  
create index IX_FuncionarioLogin on D_Funcionario(Login);
```

- Script da tabela chamada D\_Funcionario.

```
CREATE TABLE D_Funcionario(  
    Id_Funcionario int NOT NULL,  
    Nome varchar(50) NOT NULL,  
    Login varchar(50) NOT NULL,  
    Id_Chefe int NULL,  
    LinData date NOT NULL,  
    LinOrig varchar(50) NOT NULL,  
    CONSTRAINT PK_D_Funcionario PRIMARY KEY  
(  
        Id_Funcionario  
    )  
);
```



# DS – Dimensão Vendedor no Data Stage

- Algumas considerações:
  - Como vimos no projeto do DW, os nossos vendedores serão postos em uma dimensão do tipo parent-child. Esse conceito pode ser usado em todas as dimensões que são organizadas de forma hierárquica.
  - O interessante desse conceito é que ele pode formar uma árvore hierárquica infinita e complexa de forma extremamente simples: basta informar no registro, qual o Id do registro superior. E a carga é tão simples quanto o conceito!



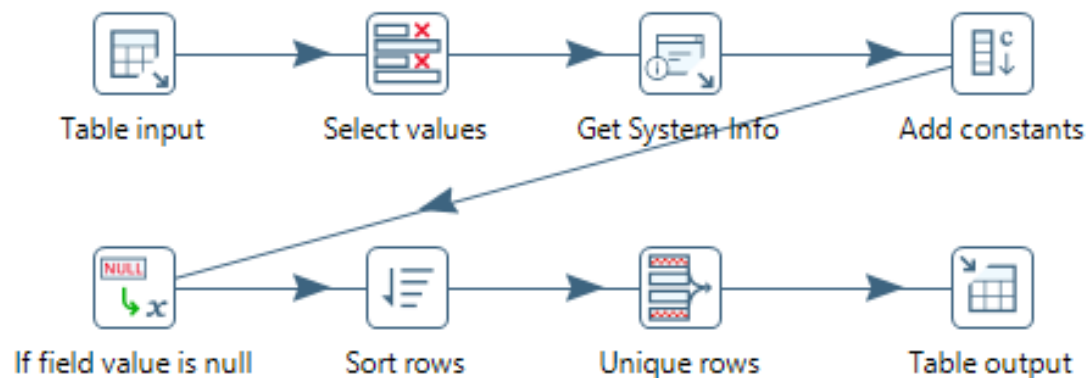
# DS – Dimensão Vendedor no Data Stage

- As transformações no pentaho para fazer carga dos dados deverá responder pelos seguintes critérios:
  - Teremos de carregar os dados dos funcionários, deixando a informação de Chefe em branco. Isso porque os “chefes” deverão já estar carregados e ter assim recebido sua surrogate, a fim de que possamos atribuí-la a cada funcionário.
  - Precisamos fazer um “update” na coluna Id\_Chefe que está em branco para vincular com o respectivo chefe para cada funcionário, deixando null se ele não tiver chefe.
  - Teremos de colocar a fonte do dado e a data em que ele entrou para nossa base, como recurso de Lineage para cada registro.



# DS – Dimensão Vendedor no Data Stage

- A transformação seguinte mostra os passos para carga da D\_Funcionario no Data Stage.







# DS – Dimensão Vendedor no Data Stage

- Algumas observações:
  - Passo Table Input da transformação Funcionário da dimensão Vendedor possui os mesmos dados de entrada que os da dimensão Tempo.
  - O Passo Select Values da transformação Funcionário da dimensão Vendedor é similar ao passo Select Values BI da dimensão Tempo, só que nele tem-se o atributo de VendedorLogin e VendedorNome que vem pelo fluxo.
  - Os passos Get System Info e Add constants tem a mesma informação da transformação cliente e ela é armazenada na tabela D\_Funcionario.



# DS – Dimensão Vendedor no Data Stage

- Passo Select values.

Nome do passo

Select values

Step name: Select values

Select & Alter Remove Meta-data

Fields to alter the meta-data for :

| # | Fieldname     | Rename to | Type | Length |
|---|---------------|-----------|------|--------|
| 1 | VendedorLogin |           | None | 50     |
| 2 | VendedorNome  |           | None | 50     |

Get fields to change

Help OK Cancela

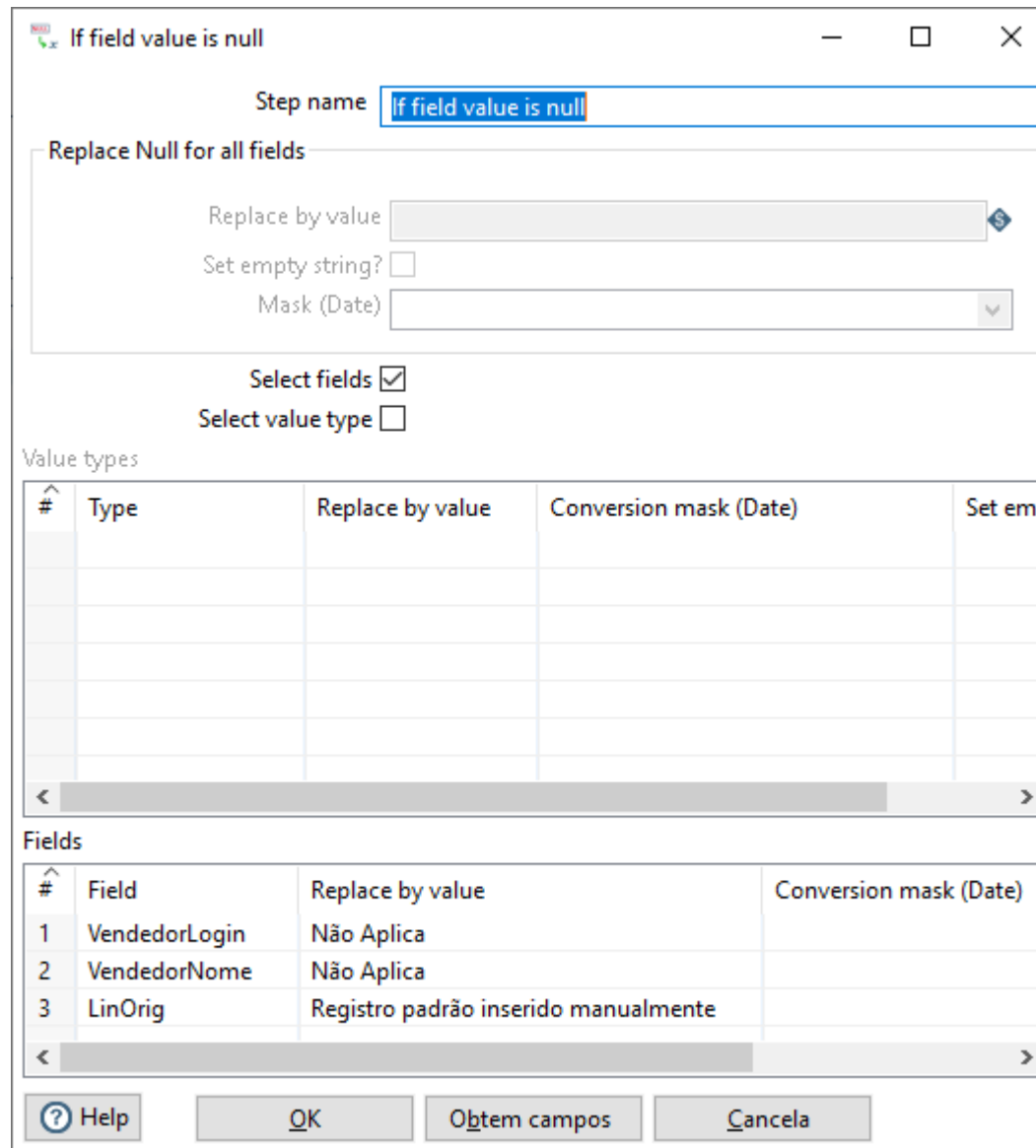
Na aba Meta-data são selecionados 4 campos, conforme mostrado na figura. Para obter estes campos foi pressionado o botão Get fields to change e removidos os campos não pertencente a tabela D\_Funcionario da dimensão Vendedor.



# DS – Dimensão Vendedor no Data Stage

- Passo If field value is null

Na parte inferior, no item Fields utilizamos o botão Obtêm campos para selecionar os campos VendedorLogin, VendedorNome e LinOrig e na coluna Replace by value adicionar respectivamente Não Aplica, Não Aplica e Registro padrão inserido manualmente.



The screenshot shows the 'If field value is null' dialog box. The 'Step name' is 'If field value is null'. The 'Replace Null for all fields' section is expanded, showing 'Replace by value' (empty), 'Set empty string?' (unchecked), and 'Mask (Date)' (empty). Below this, 'Select fields' is checked and 'Select value type' is unchecked. The 'Value types' table is empty. The 'Fields' table at the bottom contains three rows:

| # | Field         | Replace by value                     | Conversion mask (Date) |
|---|---------------|--------------------------------------|------------------------|
| 1 | VendedorLogin | Não Aplica                           |                        |
| 2 | VendedorNome  | Não Aplica                           |                        |
| 3 | LinOrig       | Registro padrão inserido manualmente |                        |

At the bottom of the dialog are buttons for 'Help', 'OK', 'Obtêm campos', and 'Cancela'. A black arrow points from the text in the previous block to the 'Obtêm campos' button.



# DS – Dimensão Vendedor no Data Stage

- Passo Sort rows

Na parte inferior, no item Fields utilizamos o botão Obtem campos para selecionar o campo VendedorLogin e ordenar de forma ascendente.

Nome do passo

Nome do Step: Sort rows

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

TMP-file prefix: out

Sort size (rows in memory): 1000000

Free memory threshold (in %):

Compress TMP Files? ☐

Only pass unique rows? (verifies) ☐

Fields :

| # | Fieldname     | Ascending | Case sensitive compare? | Sort based on current locale? |
|---|---------------|-----------|-------------------------|-------------------------------|
| 1 | VendedorLogin | S         | N                       | N                             |

< >

Help OK Cancela Obtem campos



# DS – Dimensão Vendedor no Data Stage

- Passo Unique rows

Na parte inferior, no item Fields to compare on utilizamos o botão Get para selecionar o campo VendedorLogin.

A função deste passo é eliminar as linhas duplicadas.

Nome do passo

linhas únicas

Nome do Step Unique rows

Settings

Add counter to output? ☐ Counter field

Redirect duplicate row ☐ Error description

Fields to compare on (no entries means: compare complete row)

| # | Fieldname     | Ignore case |
|---|---------------|-------------|
| 1 | VendedorLogin | N           |
|   |               |             |
|   |               |             |
|   |               |             |

Help OK Cancela Get



# DS – Dimensão Vendedor no Data Stage

- Passo Table output.

Nome do Step: Table output

Connection: DS

Target schema:

Target table: D\_Funcionario

Commit size: 1000

Truncate table: ☒

Ignore insert errors: ☐

Specify database fields: ☒

Main options Database fields

Colunas a inserir:

| # | Table field | Stream field   |
|---|-------------|----------------|
| 1 | login       | VendedorLog... |
| 2 | nome        | VendedorNo...  |
| 3 | LinData     | LinData        |
| 4 | LinOrig     | LinOrig        |

Get fields

Enter field mapping

Help OK Cancela SQL

Nome do passo

Usado para criar uma conexão ao banco de dados DS para o Postgres.

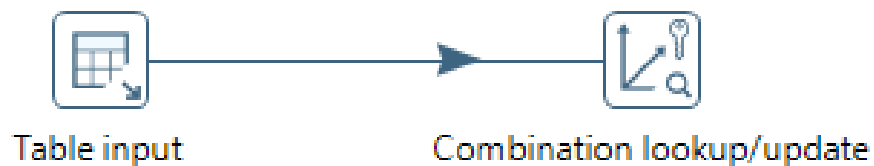
Use o botão Navega para escolher a tabela alvo que no caso é D\_Funcionario.

Usado para obter os campos que vem do passo anterior e associar com os campos da Tabela alvo D\_Funcionario.



# DS – Dimensão Vendedor no Data Warehouse

- A transformação seguinte mostra os passos para carga da D\_Funcionario no Data Warehouse.

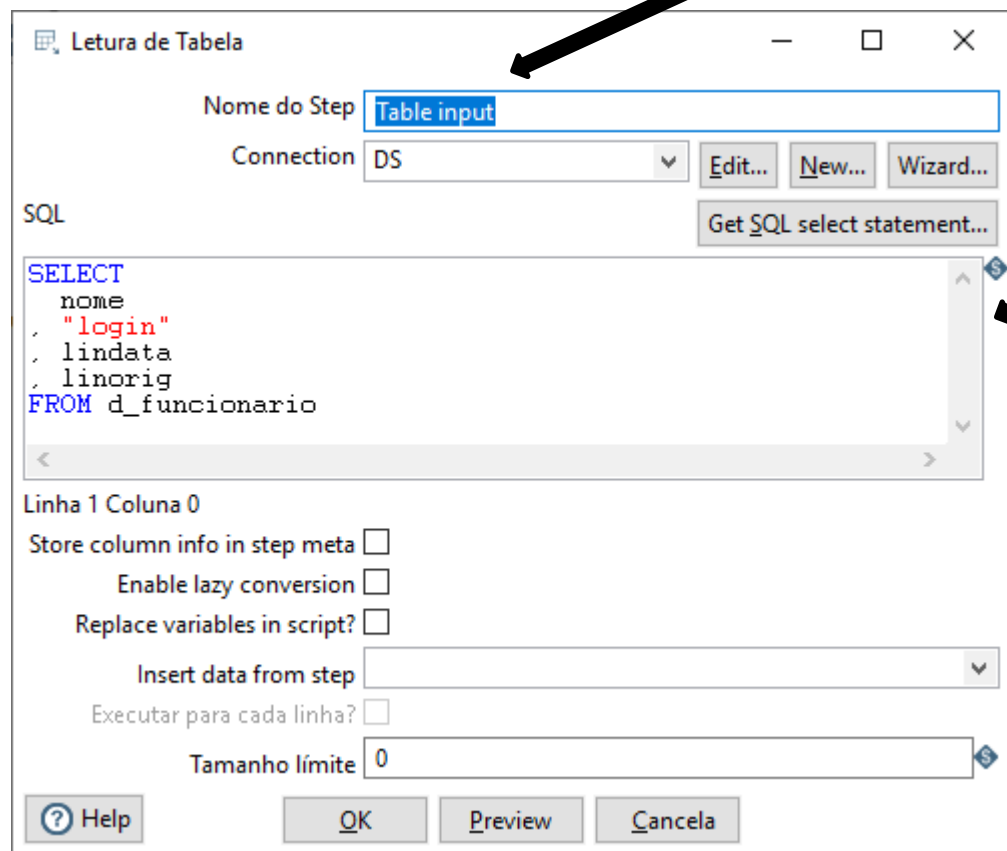




# DS – Dimensão Vendedor no Data Warehouse

- Passo Table input.

Nome do passo



Letura de Tabela

Nome do Step: Table input

Connection: DS

SQL:

```
SELECT
  nome
, "login"
, lindata
, linorig
FROM d_funcionario
```

Linha 1 Coluna 0

Store column info in step meta ☐

Enable lazy conversion ☐

Replace variables in script? ☐

Insert data from step

Executar para cada linha? ☐

Tamanho limite: 0

Buttons: Help, OK, Preview, Cancela

Obtém os nomes dos campos da tabela d\_Funcionario no banco de dados DS





# DS – Dimensão Vendedor no Data Warehouse

- Passo Combination lookup/update.

Nome do passo e o nome da conexão ao banco de dados, no caso DW, que é criada através do botão New conforme já explicado para conexão ao banco de dados DS

A tabela de destino deste passo é a tabela D\_Funcionario que é selecionada pelo botão de navegação

Combinação lookup / update

Nome do Step: Combination lookup/update

Connection: DW

Target schema:

Tabela de destino: D\_Funcionario

Confirma tamanho: 100

Tamanho do Cache: 9999

Pre-load the cache? ☐

Campos Chave (para verificar linha na tabela):

| # | Campo Dimensão | Campo no fluxo |
|---|----------------|----------------|
| 1 | Nome           | Nome           |
| 2 | login          | Login          |
| 3 | LinData        | LinData        |
| 4 | LinOrig        | LinOrig        |

Campo chave técnica: Id\_Funcionario

Criação de chave técnica:

- ☒ Usa tabela máxima + 1
- ☐ Usa sequência
- ☐ Usa o campo de auto incremento

Remove campos lookup? ☐

Usa hashcode? ☐

Campo Hashcode na tabela:

Date of last update field:

? Help OK Cancela Obtem Campos SQL



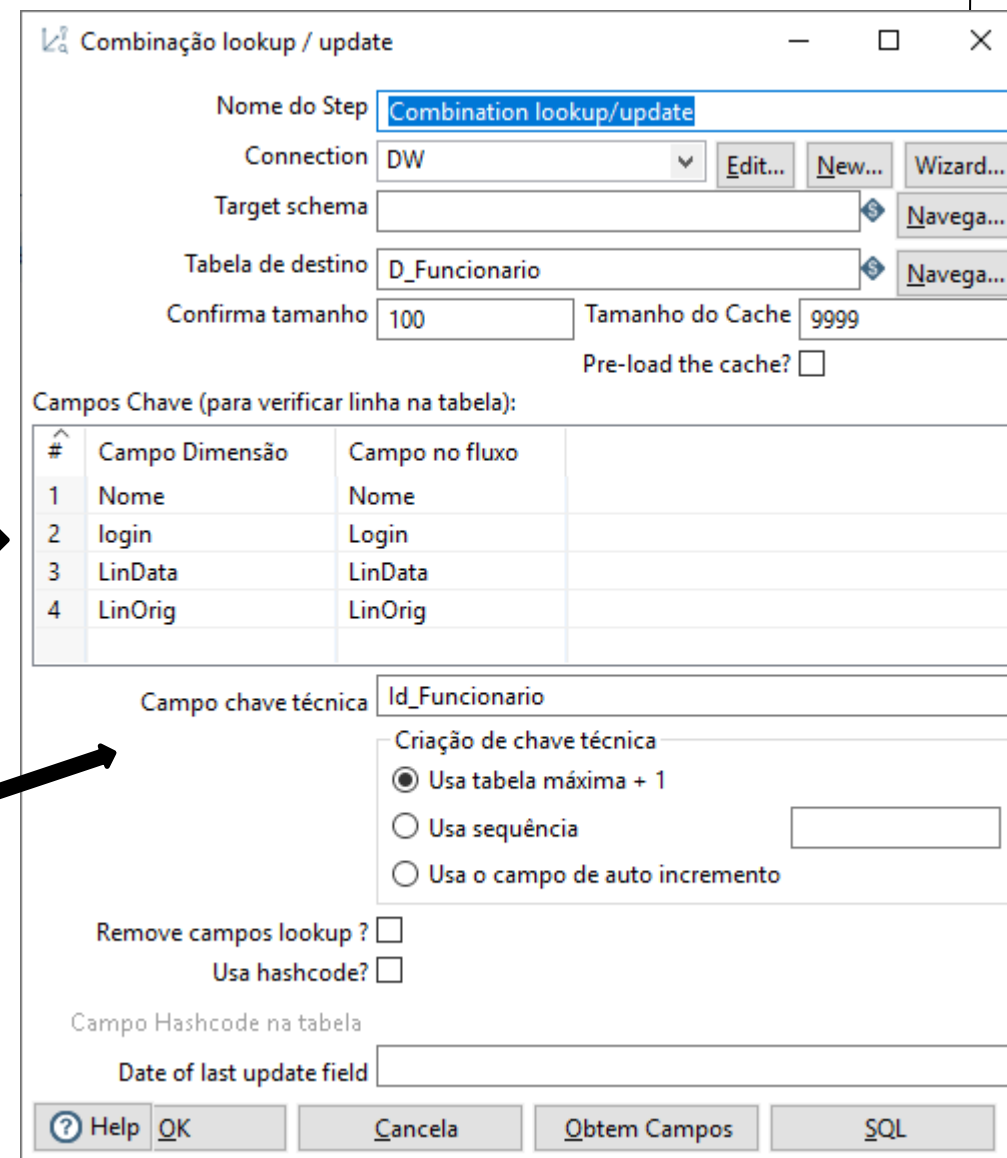
# DS – Dimensão Vendedor no Data Warehouse

- Passo Combination lookup/update.

O item Campos chave permite fazer a comparação com os campos do fluxo e relação aos campos da dimensão caso os valores já existam na tabela D\_Funcionario no DW nada é inserido caso contrário novos valores são inseridos.

Eles são obtidos a partir do botão Obtem Campos.

Campo chave técnica que é usada para criar os valores do campo Id\_Funcionario a partir de 1 e é incrementado de uma unidade.



Combinação lookup / update

Nome do Step: Combination lookup/update

Connection: DW

Target schema:

Tabela de destino: D\_Funcionario

Confirma tamanho: 100

Tamanho do Cache: 9999

Pre-load the cache? ☐

Campos Chave (para verificar linha na tabela):

| # | Campo Dimensão | Campo no fluxo |
|---|----------------|----------------|
| 1 | Nome           | Nome           |
| 2 | login          | Login          |
| 3 | LinData        | LinData        |
| 4 | LinOrig        | LinOrig        |

Campo chave técnica: Id\_Funcionario

Criação de chave técnica

☒ Usa tabela máxima + 1

☐ Usa sequência

☐ Usa o campo de auto incremento

Remove campos lookup? ☐

Usa hashcode? ☐

Campo Hashcode na tabela:

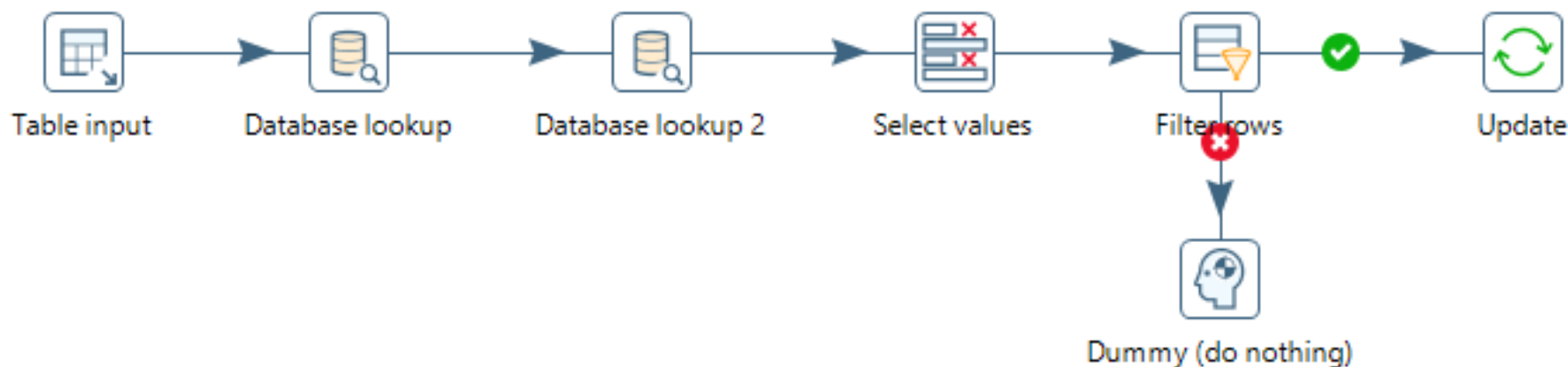
Date of last update field:

Buttons: Help, OK, Cancela, Obtem Campos, SQL



# DS – Dimensão Vendedor no Data Warehouse

- A transformação seguinte mostra os passos para carga da D\_Funcionario DW UPDATE ID\_CHEFE no Data Warehouse.





# DS – Dimensão Vendedor no Data Warehouse



- Algumas observações:
  - O Passo Select Values da transformação D\_Funcionario UPDATE ID CHEFE da dimensão Vendedor é similar ao passo Select Values BI da dimensão Tempo, só que nele tem-se o atributos de funcionário que vem pelo fluxo.



# DS – Dimensão Vendedor no Data Warehouse

- Passo Table input.

Nome do passo

Letura de Tabela

Nome do Step:

Connection:

SQL:

```
SELECT
  Id_Funcionario
, Nome
, Login
, Id_Chefe
, LinData
, LinOrig
FROM D_Funcionario
```

Linha 1 Coluna 0

Store column info in step meta ☐

Enable lazy conversion ☐

Replace variables in script? ☐

Insert data from step

Executar para cada linha? ☐

Tamanho limite:

Obtém os nomes dos campos da tabela d\_Funcionario no banco de dados DW



# DS – Dimensão Vendedor no Data Warehouse

- Passo Database lookup.

Nome do passo e o nome da conexão ao banco de dados, no caso DS.

A tabela de Lookup deste passo é a tabela Tblmp\_vendas que é selecionada pelo botão de navegação

Lookup de valor do banco de dados

Nome do Step: Database lookup

Connection: DS

Lookup schema:

Tabela Lookup: Tblmp\_Vendas

Habilita cache? ☐

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 | VendedorLogin   | =          | Login  |        |

Valores a serem retornados da tabela lookup :

| # | campo             | Novo nome | Default | Tipo   |
|---|-------------------|-----------|---------|--------|
| 1 | VendedorChefeNome |           |         | String |
| 2 | VendedorLogin     |           |         | String |

Não passa a linha se o lookup falhar ☐

Falha quando ocorrerem resultados múltiplos ☐

Ordem por:

Buttons: Help, OK, Cancela, Obtem Campos, Obtem campos lookup



# DS – Dimensão Vendedor no Data Warehouse

- Passo Database lookup.

Nesta área o item Campo da tabela tem atribuído o “VendedorLogin” pertencente a tabela TblImp\_Vendas obtido pelo campo Obtem campos lookup. A coluna Campo1 seleciona-se a partir do botão Obtem Campos o campo Login que vem pelo fluxo. Os valores deles são comparados e caso sejam iguais o VendedorChefeNome é recuperado e segue pelo fluxo.



Lookup de valor do banco de dados

Nome do Step: Database lookup

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

Lookup schema: [Navega...]

Tabela Lookup: TblImp\_Vendas [Navega...]

Habilita cache? ☐

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 | VendedorLogin   | =          | Login  |        |
|   |                 |            |        |        |
|   |                 |            |        |        |
|   |                 |            |        |        |
|   |                 |            |        |        |
|   |                 |            |        |        |
|   |                 |            |        |        |
|   |                 |            |        |        |
|   |                 |            |        |        |
|   |                 |            |        |        |

Valores a serem retornados da tabela lookup:

| # | campo             | Novo nome | Default | Tipo   |
|---|-------------------|-----------|---------|--------|
| 1 | VendedorChefeNome |           |         | String |
| 2 | VendedorLogin     |           |         | String |
|   |                   |           |         |        |
|   |                   |           |         |        |
|   |                   |           |         |        |
|   |                   |           |         |        |
|   |                   |           |         |        |
|   |                   |           |         |        |
|   |                   |           |         |        |
|   |                   |           |         |        |

Não passa a linha se o lookup falhar ☐

Falha quando ocorrerem resultados múltiplos ☐

Ordem por: [ ]

[?] Help [OK] [Cancela] [Obtem Campos] [Obtem campos lookup]



# DS – Dimensão Vendedor no Data Warehouse

- Passo Database lookup.

Nome do passo e o nome da conexão ao banco de dados, no caso DW, que é criada através do botão New conforme já explicado para conexão ao banco de dados DS

A tabela de Lookup deste passo é a tabela D\_Funcionario que é selecionada pelo botão de navegação

Lookup de valor do banco de dados

Nome do Step: Database lookup 2

Connection: DW

Lookup schema:

Tabela Lookup: D\_Funcionario

Habilita cache?: ☐

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 | Nome            | =          | VendedorChefeNome |        |
|   |                 |            |                   |        |
|   |                 |            |                   |        |
|   |                 |            |                   |        |
|   |                 |            |                   |        |
|   |                 |            |                   |        |
|   |                 |            |                   |        |
|   |                 |            |                   |        |
|   |                 |            |                   |        |

Valores a serem retornados da tabela lookup:

| # | campo          | Novo nome | Default | Tipo    |
|---|----------------|-----------|---------|---------|
| 1 | Id_Funcionario |           |         | Integer |
| 2 | Nome           |           |         | String  |
|   |                |           |         |         |
|   |                |           |         |         |
|   |                |           |         |         |
|   |                |           |         |         |
|   |                |           |         |         |
|   |                |           |         |         |
|   |                |           |         |         |

Não passa a linha se o lookup falhar: ☐

Falha quando ocorrerem resultados múltiplos: ☐

Ordem por:

? Help OK Cancela Obtem Campos Obtem campos lookup





# DS – Dimensão Vendedor no Data Warehouse

- Passo Database lookup.

Nesta área o item Campo da tabela tem atribuído o “Nome” pertencente a tabela D\_Funcionario obtido pelo campo Obtem campos lookup. A coluna Campo1 seleciona-se a partir do botão Obtem Campos o campo VendedorChefeNome que vem pelo fluxo. Os valores deles são comparados e caso sejam iguais o id\_Funcionario é recuperado e segue pelo fluxo.



Lookup de valor do banco de dados

Nome do Step: Database lookup 2

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

Lookup schema: [Navega...]

Tabela Lookup: D\_Funcionario [Navega...]

Habilita cache? ☐

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 | Nome            | =          | VendedorChefeNome |        |

Valores a serem retornados da tabela lookup :

| # | campo          | Novo nome | Default | Tipo    |
|---|----------------|-----------|---------|---------|
| 1 | Id_Funcionario |           |         | Integer |
| 2 | Nome           |           |         | String  |

Não passa a linha se o lookup falhar ☐

Falha quando ocorrerem resultados múltiplos ☐

Ordem por: [ ]

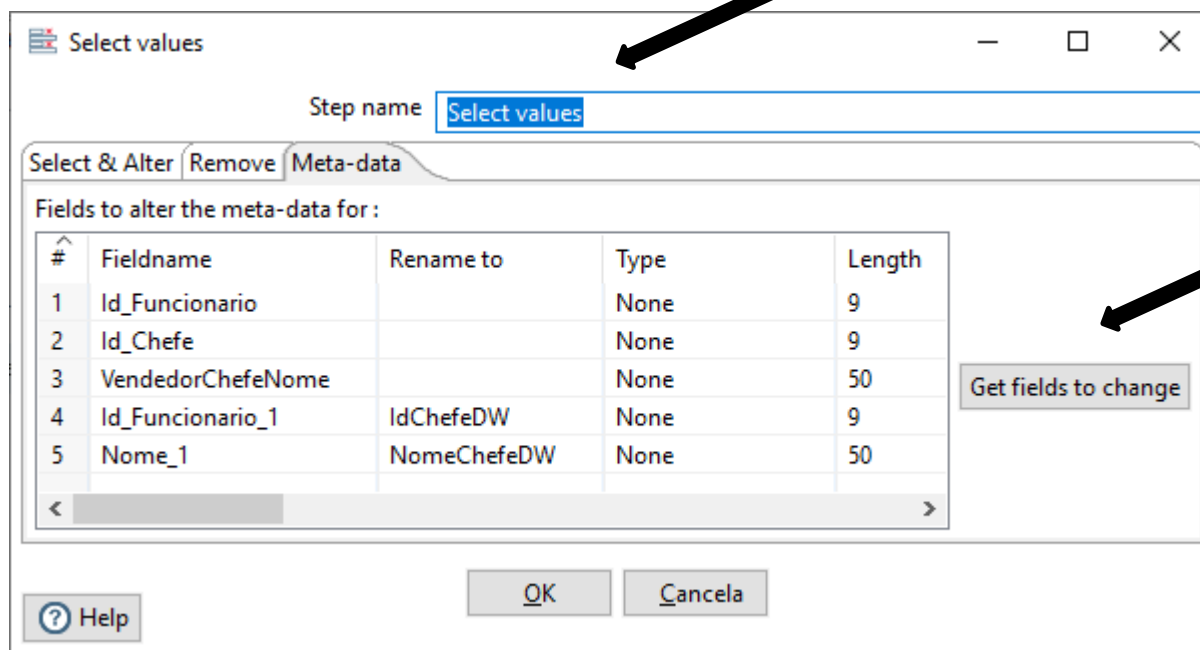
[?] Help [OK] [Cancela] [Obtem Campos] [Obtem campos lookup]



# DS – Dimensão Vendedor no Data Warehouse

- Passo Select values.

Nome do passo



Step name:

Select & Alter Remove Meta-data

Fields to alter the meta-data for:

| # | Fieldname         | Rename to   | Type | Length |
|---|-------------------|-------------|------|--------|
| 1 | Id_Funcionario    |             | None | 9      |
| 2 | Id_Chefe          |             | None | 9      |
| 3 | VendedorChefeNome |             | None | 50     |
| 4 | Id_Funcionario_1  | IdChefeDW   | None | 9      |
| 5 | Nome_1            | NomeChefeDW | None | 50     |

Get fields to change

Help OK Cancela

Na aba Meta-data são selecionados 5 campos, conforme mostrado na figura. Para obter estes campos foi pressionado o botão Get fields to change e foram selecionados os campos mostrados na figura que estão vindo pelo fluxo e dois campos foram renomeados para IdChefeDW e NomeChefeDW.



# DS – Dimensão Vendedor no Data Warehouse

- Passo Filter rows

Na parte inferior, no item The condition comparamos o campo VendedorChefeNome com NomeChefeDW caso seus valores sejam iguais os campos que vem pelo fluxo seguem para o passo Update, caso contrário seguem para o passo Dummy, que é um passo que não faz nada.

Nome do passo

Filter rows

Step name: Filter rows

Send 'true' data to step: Update

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

The condition:

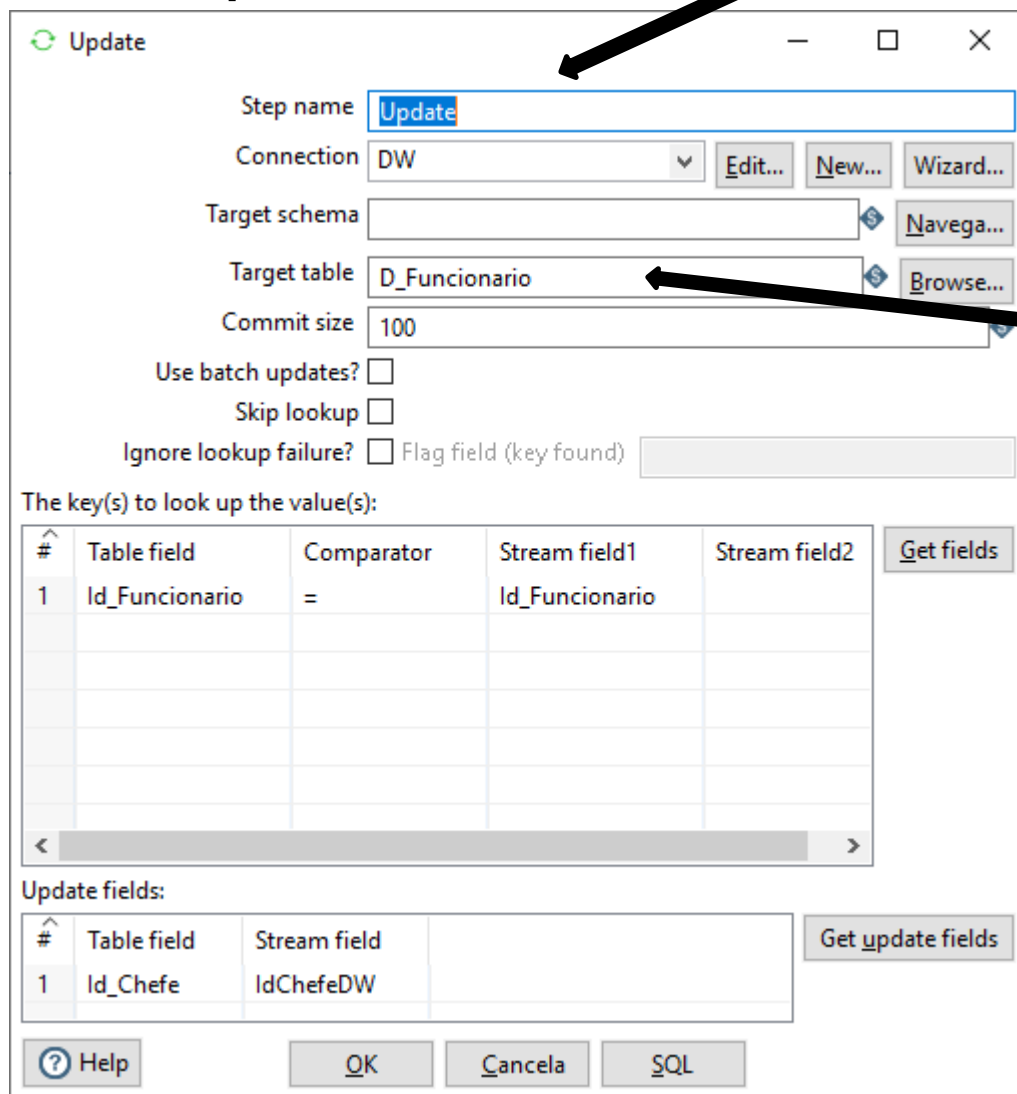
VendedorChefeNome = NomeChefeDW

Help OK Cancela



# DS – Dimensão Vendedor no Data Warehouse

- Passo Update.



Update

Step name: Update

Connection: DW

Target schema:

Target table: D\_Funcionario

Commit size: 100

Use batch updates? ☐

Skip lookup? ☐

Ignore lookup failure? ☐ Flag field (key found):

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

| # | Table field    | Comparator | Stream field1  | Stream field2 |
|---|----------------|------------|----------------|---------------|
| 1 | Id_Funcionario | =          | Id_Funcionario |               |

Update fields:

| # | Table field | Stream field |
|---|-------------|--------------|
| 1 | Id_Chefe    | IdChefeDW    |

Nome do passo

Usado para criar uma conexão ao banco de dados DS para o Postgres.

Use o botão Navega para escolher a tabela alvo que no caso é D\_Funcionário.

Usado para obter os campos que vem do passo anterior e associar com os campos da Tabela alvo D\_Pais.

Usado para obter o campo Id\_Chefe da tabela D\_Funcionario e atualizar o seu valor com o valor de IdChefeDW que vem do passo anterior.