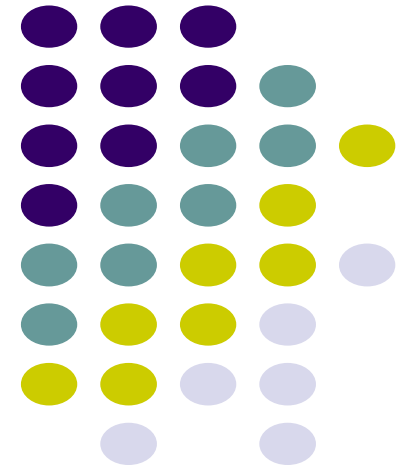


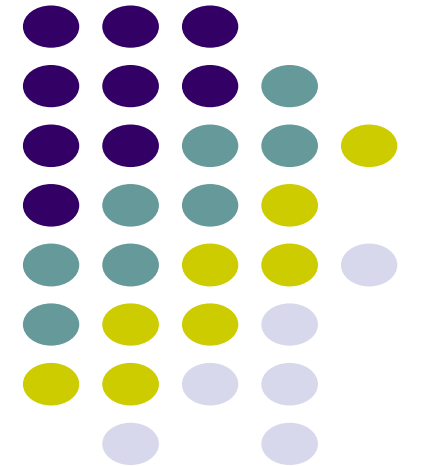


Business Intelligence: O desenho do Data Warehouse

Vitor Valerio de Souza Campos



Processo para criar o Data Stage e o Data Warehouse





DS – Fato Vendas

- Fato Vendas é formada por grupo F_Venda e F_VendaDetalhe.

F_Venda

Data (date)
Nr_NF (varchar(10))
Id_Cliente (int)
Id_Funcionario (int)
Id_RegiaoVendas (int)
Vlr_Imposto (decimal(18,2))
Vlr_Frete (decimal(18,2))
LinData (date)
LinOrig (varchar(50))

F_VendaDetalhe

Data (date)
Nr_NF (varchar(10))
Id_Cliente (int)
Id_Funcionario (int)
Id_RegiaoVendas (int)
Id_Produto (int)
Vlr_Unitario (decimal(18,2))
Qtd_Vendida (int)
LinData (date)
LinOrig (varchar(50))



DS – Fato Vendas

- Algumas considerações:
 - Finalizada a carga das dimensões, vamos carregar nossa Fato, a primeira de duas que temos em nosso DW. Teremos que começar pela Fato de Venda, dado que os registros dela serão “pais” da tabela detalhada de Vendas, o que nos impossibilitaria de carregarmos a Fato Venda Detalhes primeiramente.



DS – Fato Vendas

- Algumas considerações:
 - vamos recordar o primeiro dos conceitos do DW. Nenhuma coluna pode aceitar null. Se um dado veio nulo, o processo de carga deve entender se é algo esperado ou um erro. Deve-se preencher o campo com uma informação genérica para que se saiba que aquele dado não foi carregado. Essa prática resolve muitos problemas de origens de dados incoerentes. Isso quer dizer que temos de impedir que um dado da Fato venha com um *nulo* ou que, pior ainda, deixe de ser carregado (tornando os números incorretos).



DS – Fato Vendas

- Algumas considerações:
 - Para darmos esse tratamento, foi inserido um registro genérico em cada uma das dimensões. Ou seja, se tiver ocorrido qualquer problema na carga de uma dimensão e não houver uma relação válida entre o Fato e essa Dimensão, um registro genérico será atribuído, a fim de se manter os valores coerentes.



DS – Fato Vendas

- Algumas considerações:
 - Coloque como código dos dados genéricos apenas informações que, **com certeza**, não existam na realidade. Por exemplo, usar o dia 01/01/2020 como genérica para a D_Data é um erro, pois existirão vendas reais nesse dia, e elas seriam misturadas com as vendas cujas datas não foram estipuladas. Use sempre informações bastante diferentes das possíveis (por exemplo, 01/01/1900 para a Data!)
 - Vamos então criar a estrutura no nosso DS. E como a Fato no DS não receberá uma surrogate própria, ela será **exatamente** igual à estrutura do DW.



DS – Fato Vendas

- Script da tabela chamada F_Venda.

```
CREATE TABLE F_Venda(  
    Data date NOT NULL,  
    Nr_NF varchar (10) NOT NULL,  
    Id_Cliente int NOT NULL,  
    Id_Funcionario int NOT NULL,  
    Id_RegiaoVendas int NOT NULL,  
    Vlr_Imposto decimal(18, 2) NOT NULL,  
    Vlr_Frete decimal(18, 2) NOT NULL,  
    LinData date NOT NULL,  
    LinOrig varchar(50) NOT NULL,
```




DS – Fato Vendas

- Script da tabela chamada F_Venda.

```
CONSTRAINT PK_F_Venda PRIMARY KEY  
(  
    Data ,  
    Nr_NF ,  
    Id_Cliente ,  
    Id_Funcionario ,  
    Id_RegiaoVendas  
)  
);
```



Data Warehouse – Fato Vendas

- Script da tabela chamada F_Venda.

```
CREATE TABLE F_Venda(  
    Id_Fato int NOT NULL,  
    Data date NOT NULL,  
    Nr_NF varchar (10) NOT NULL,  
    Id_Cliente int NOT NULL,  
    Id_Funcionario int NOT NULL,  
    Id_RegiaoVendas int NOT NULL,  
    Vlr_Imposto decimal(18, 2) NOT NULL,  
    Vlr_Frete decimal(18, 2) NOT NULL,  
    LinData date NOT NULL,  
    LinOrig varchar(50) NOT NULL,
```



Data Warehouse – Fato Vendas

- Script da tabela chamada F_Venda.

```
CONSTRAINT PK_F_Venda PRIMARY KEY  
(  
    Data ,  
    Nr_NF ,  
    Id_Cliente ,  
    Id_Funcionario ,  
    Id_RegiaoVendas  
)  
);
```



Data Warehouse – Fato Vendas

- Script da tabela chamada F_Venda.

```
ALTER TABLE F_Venda ADD CONSTRAINT FK_F_Venda_D_Cliente FOREIGN  
KEY(Id_Cliente) REFERENCES D_Cliente (Id_Cliente); ON DELETE CASCADE;
```

```
ALTER TABLE F_Venda ADD CONSTRAINT FK_F_Venda_D_Data FOREIGN  
KEY(Data) REFERENCES D_Data (Data); ON DELETE CASCADE;
```

```
ALTER TABLE F_Venda ADD CONSTRAINT FK_F_Venda_D_Funcionario  
FOREIGN KEY(Id_Funcionario) REFERENCES D_Funcionario (Id_Funcionario);  
ON DELETE CASCADE;
```

```
ALTER TABLE F_Venda ADD CONSTRAINT FK_F_Venda_D_RegiaoVendas  
FOREIGN KEY(Id_RegiaoVendas) REFERENCES D_RegiaoVendas  
(Id_RegiaoVendas); ON DELETE CASCADE;
```



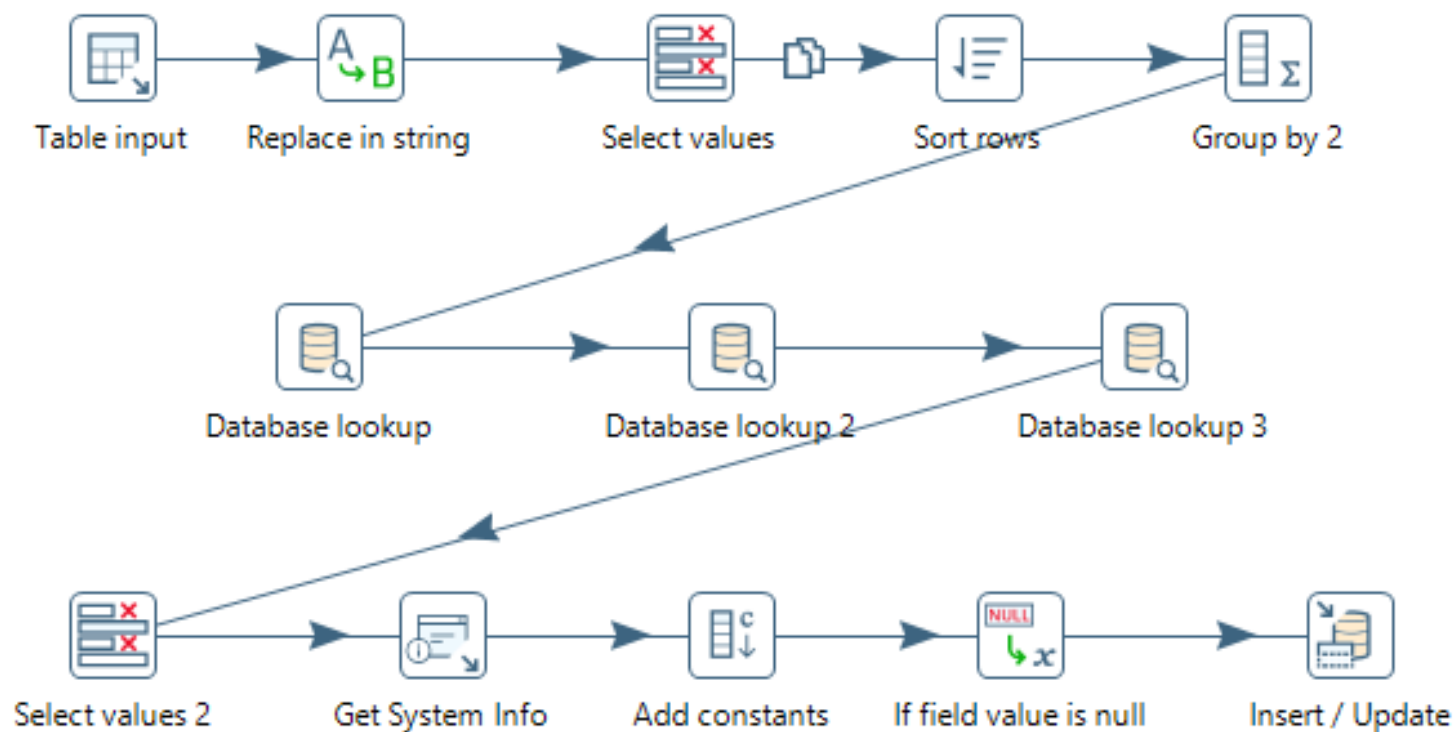
DS – Fato Vendas no Data Stage

- As transformações no pentaho para fazer carga dos dados deverá responder pelos seguintes critérios: Selecionar todos os dados de venda;
 - Atribuir a cada dimensão, sua surrogate, mesmo que seja a dos registros genéricos.
 - Avaliar se o dado já existe ou não na Fato do DW e carregar apenas os novos registros.
 - 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 – Fatos Vendas no Data Stage

- A transformação seguinte mostra os passos para carga da F_Vendas no Data Stage.





DS – Fatos Vendas no Data Stage

- Algumas observações:
 - Passo Table Input da transformação Fatos Venda possui os mesmos dados de entrada que os da dimensão Tempo.
 - O passo Replace in string substitui . (ponto) por – (traço) da mesma forma que na dimensão Tempo.
 - Get System Info tem a mesma informação da transformação cliente e ela é armazenada na tabela F_Venda.
 - Add constants tem a mesma informação da transformação cliente e ela é armazenada na tabela F_Venda.



DS – Fatos Vendas 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	NrNf		None	50
2	DataVenda		String	10
3	CodCliente		None	50
4	RegiaoVendas		None	50
5	VendedorLogin		None	50
6	ImpTotal		None	50
7	Frete		None	50

Get fields to change

Help OK Cancela

Na aba Meta-data são selecionados 7 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 F_Vendas da Fato Vendas.



DS – Fatos Vendas no Data Stage

- Passo Sort rows

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

Nome do passo

Sort rows

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	RegiaoVendas	S	N	N

< >

Help OK Cancela Obtem campos



DS – Fatos Vendas no Data Stage

- Passo Group by 2

Na parte inferior, no item The fields that make up the group utilizamos o botão Get Fields para selecionarmos os campos listados na figura.

Na parte inferior, no item Aggregates utilizamos o botão Get lookup fields para selecionarmos os campos listados na figura.

Nome do passo

Group by

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	NrNf
2	DataVenda
3	CodCliente
4	RegiaoVendas
5	VendedorLogin

Aggregates :

#	Name	Subject	Type	Va
1	ImpTotal	ImpTotal	Sum	
2	Frete	Frete	Sum	

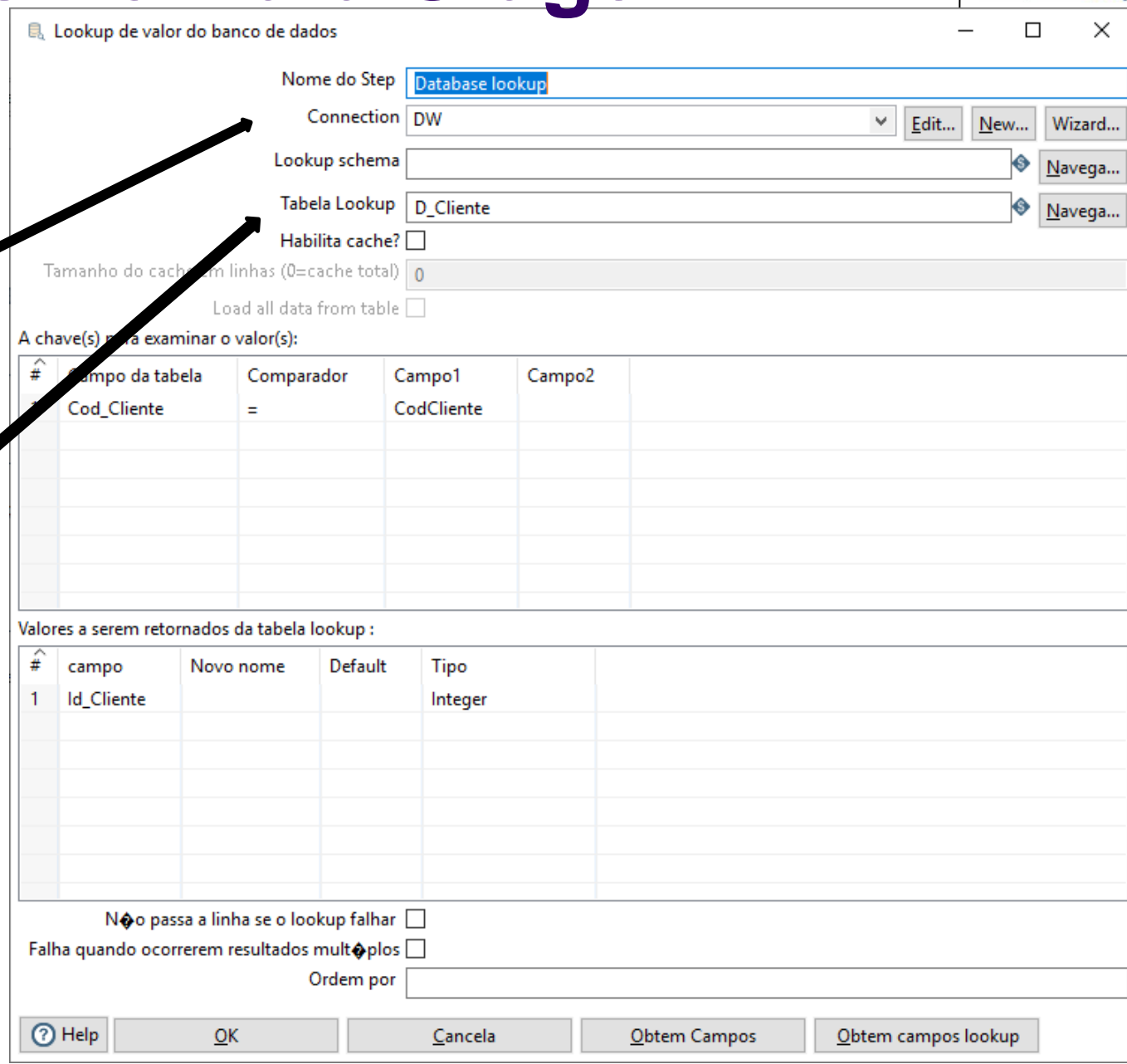


DS – Fatos Vendas no Data Stage

- 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_Cliente que é selecionada pelo botão de navegação



Lookup de valor do banco de dados

Nome do Step: Database lookup

Connection: DW

Lookup schema:

Tabela Lookup: D_Cliente

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	Cod_Cliente	=	CodCliente	

Valores a serem retornados da tabela lookup:

#	campo	Novo nome	Default	Tipo
1	Id_Cliente			Integer

☐ Não passa a linha se o lookup falhar

☐ Falha quando ocorrerem resultados múltiplos

Ordem por:

? Help OK Cancela Obtem Campos Obtem campos lookup



DS – Fatos Vendas no Data Stage

- Passo Database lookup.

Nesta área o item Campo da tabela tem atribuído o “Cod_Cliente” pertencente a tabela D_Cliente obtido pelo campo Obtem campos lookup. A coluna Campo1 obtido a partir do botão Obtem Campos que vem pelo fluxo e é CodCliente. Os valores deles são comparados e caso sejam iguais o Id_Cliente é recuperado e segue pelo fluxo.



Lookup de valor do banco de dados

Nome do Step: Database lookup

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

Lookup schema: [Navega...]

Tabela Lookup: D_Cliente [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	Cod_Cliente	=	CodCliente	

Valores a serem retornados da tabela lookup :

#	campo	Novo nome	Default	Tipo
1	Id_Cliente			Integer

Não passa a linha se o lookup falhar ☐

Falha quando ocorrerem resultados múltiplos ☐

Ordem por: []

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



DS – Fatos Vendas no Data Stage

- Passo Database lookup 2.

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 [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	login	=	VendedorLogin	

Valores a serem retornados da tabela lookup :

#	campo	Novo nome	Default	Tipo
1	Id_Funcionario			Integer

Não passa a linha se o lookup falhar ☐

Falha quando ocorrerem resultados múltiplos ☐

Ordem por: []

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



DS – Fatos Vendas no Data Stage

- Passo Database lookup 2.

Nesta área o item Campo da tabela tem atribuído o “login” pertencente a tabela D_Funcionario obtido pelo campo Obtem campos lookup. A coluna Campo1 obtido a partir do botão Obtem Campos que vem pelo fluxo e é VendedorLogin. 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	login	=	VendedorLogin	

Valores a serem retornados da tabela lookup :

#	campo	Novo nome	Default	Tipo
1	Id_Funcionario			Integer

Não passa a linha se o lookup falhar ☐

Falha quando ocorrerem resultados múltiplos ☐

Ordem por: []

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



DS – Fatos Vendas no Data Stage

- Passo Database lookup 3.

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_RegiaoVendas que é selecionada pelo botão de navegação

Lookup de valor do banco de dados

Nome do Step: Database lookup 3

Connection: DW

Lookup schema:

Tabela Lookup: D_RegiaoVendas

Habilita cache?: ☐

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

Load all data from table: ☐

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

#	Nome da tabela	Comparador	Campo1	Campo2
1	Nome	=	RegiaoVendas	

Valores a serem retornados da tabela lookup:

#	campo	Novo nome	Default	Tipo
1	Id_RegiaoVendas			Integer

☐ Não passa a linha se o lookup falhar

☐ Falha quando ocorrerem resultados múltiplos

Ordem por:

Help OK Cancela Obtem Campos Obtem campos lookup



DS – Fatos Vendas no Data Stage

- Passo Database lookup 3.

Nesta área o item Campo da tabela tem atribuído o “nome” pertencente a tabela D_RegiaoVendas obtido pelo campo Obtem campos lookup. A coluna Campo1 obtido a partir do botão Obtem Campos que vem pelo fluxo e é RegiaoVendas. Os valores deles são comparados e caso sejam iguais o Id_RegiaoVendas é recuperado e segue pelo fluxo.



Lookup de valor do banco de dados

Nome do Step: Database lookup 3

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

Lookup schema: [Navega...]

Tabela Lookup: D_RegiaoVendas [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	=	RegiaoVendas	

Valores a serem retornados da tabela lookup :

#	campo	Novo nome	Default	Tipo
1	Id_RegiaoVendas			Integer

Não passa a linha se o lookup falhar ☐

Falha quando ocorrerem resultados múltiplos ☐

Ordem por: []

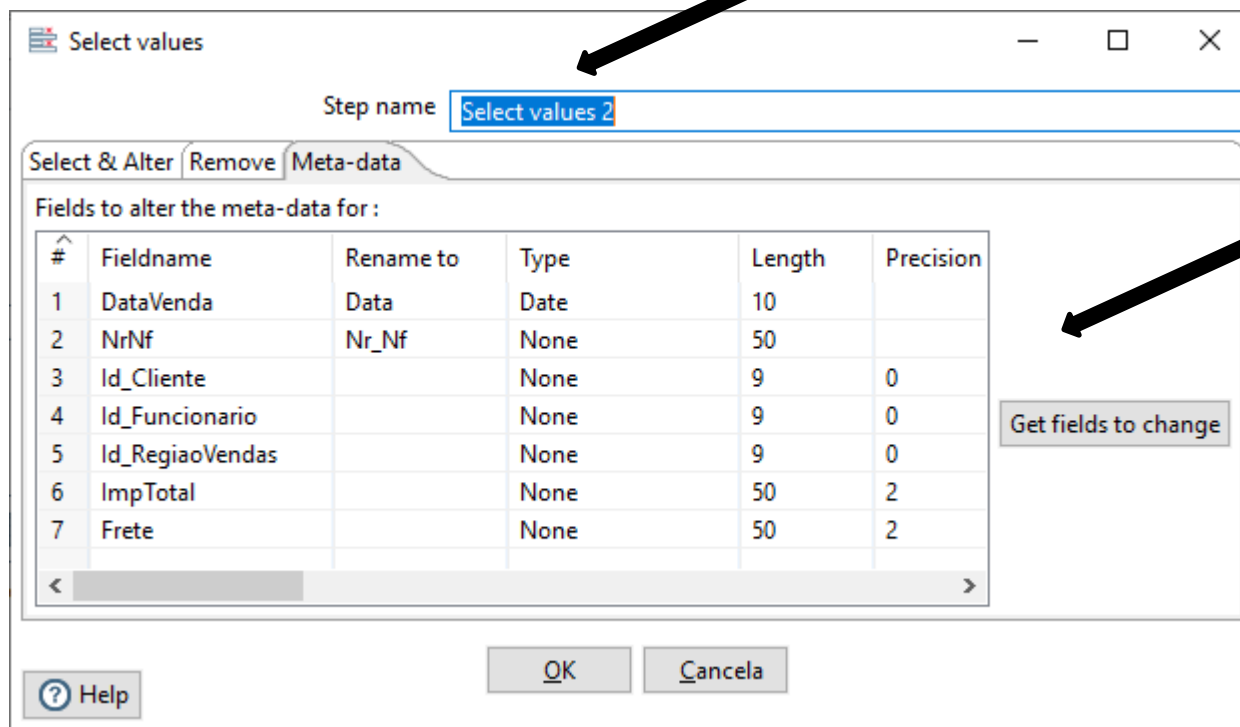
[?] Help [OK] [Cancela] [Obtem Campos] [Obtem campos lookup]



DS – Fatos Vendas no Data Stage

- Passo Select values 2

Nome do passo



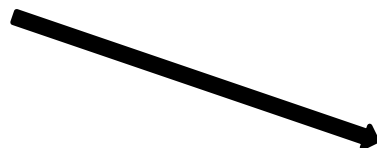
Na aba Meta-data são selecionados 7 campos, conforme mostrado na figura. Para obter estes campos foi pressionado o botão Get fields to change e removidos os campos não pertencente a tabela F_Vendas da Fato Vendas.



DS – Fatos Vendas no Data Stage

- Passo If field value is null

Na parte inferior, no item Fields utilizamos o botão Obtêm campos para selecionar os campos mostrados na figura e na coluna Replace by value adicionar respectivamente os valores mostrados na figura.



Step name: **If field value is null**

Replace Null for all fields

Replace by value:

Set empty string? ☐

Mask (Date):

Select fields ☒

Select value type ☐

Value types

#	Type	Replace by value	Conversion mask (Date)

Fields

#	Field	Replace by value	Conversion ma
1	Data	1900/01/01	
2	ImpTotal	0	
3	Frete	0	
4	Id_Cliente	1	
5	Id_Funcionario	1	
6	Id_RegiaoVendas	1	
7	LinOrig	Registro padrão inserido manualmente	
8	Nr_Nf	0	



DS – Fatos Vendas no Data Stage

- Passo Insert / Update.

Step name: Insert / Update

Connection: DS

Target schema:

Target table: F_Venda

Commit size: 100

Don't perform any updates: ☐

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

#	Table field	Comparator	Stream field1	Stream field
1	data	=	Data	
2	Nr_NF	=	Nr_Nf	
3	Id_Cliente	=	Id_Cliente	
4	Id_Funcionario	=	Id_Funcionario	
5	Id_RegiaoVendas	=	Id_RegiaoVendas	

Update fields:

#	Table field	Stream field	Update
1	data	Data	Y
2	Nr_Nf	Nr_Nf	Y

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 F_Venda.

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



DS – Fatos Vendas no Data Stage

- Passo Insert / Update.

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

#	Table field	Comparator	Stream field1	Stream fiel	Get fields
1	data	=	Data		
2	Nr_NF	=	Nr_Nf		
3	Id_Cliente	=	Id_Cliente		
4	Id_Funcionario	=	Id_Funcionario		
5	Id_RegiaoVendas	=	Id_RegiaoVendas		

< >

Update fields:

#	Table field	Stream field	Update	Get update fields	Edit mapping
1	data	Data	Y		
2	Nr_Nf	Nr_Nf	Y		
3	Id_Cliente	Id_Cliente	Y		
4	Id_Funcionario	Id_Funcionario	Y		
5	Id_RegiaoVendas	Id_RegiaoVendas	Y		
6	Vlr_Imposto	ImpTotal	Y		
7	Vlr_Frete	Frete	Y		
8	LinData	LinData	Y		
9	LinOrig	LinOrig	Y		

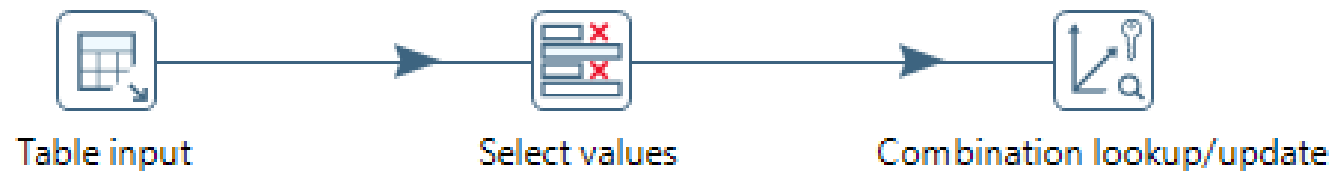
? Help OK Cancela SQL

Usado para obter os campos que vem do passo anterior e associar com os campos da Tabela alvo F_Venda que vão ser inseridos na tabela alvo.



DS – Fatos Vendas no Data Warehouse

- A transformação seguinte mostra os passos para carga da F_Vendas no Data Warehouse.

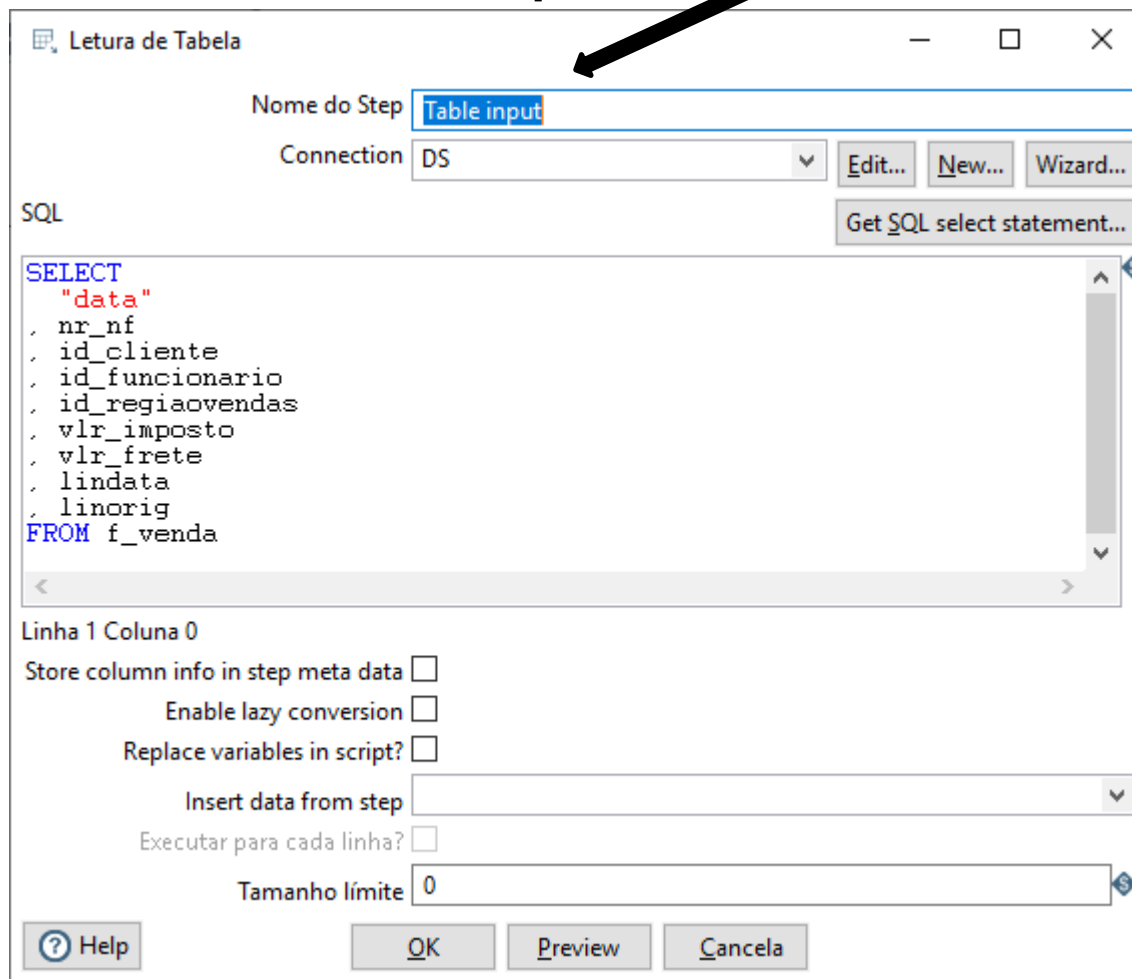




DS – Fatos Vendas no Data Warehouse

- Passo Table input.

Nome do passo



Letura de Tabela

Nome do Step: Table input

Connection: DS

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

SQL:

```
SELECT
  "data"
, nr_nf
, id_cliente
, id_funcionario
, id_regiaoventas
, vlr_imposto
, vlr_frete
, lndata
, linorig
FROM f_venda
```

Buttons: Get SQL select statement...

Linha 1 Coluna 0

Store column info in step meta data ☐

Enable lazy conversion ☐

Replace variables in script? ☐

Insert data from step

Executar para cada linha? ☐

Tamanho limite: 0

Buttons: ? Help, OK, Preview, Cancela

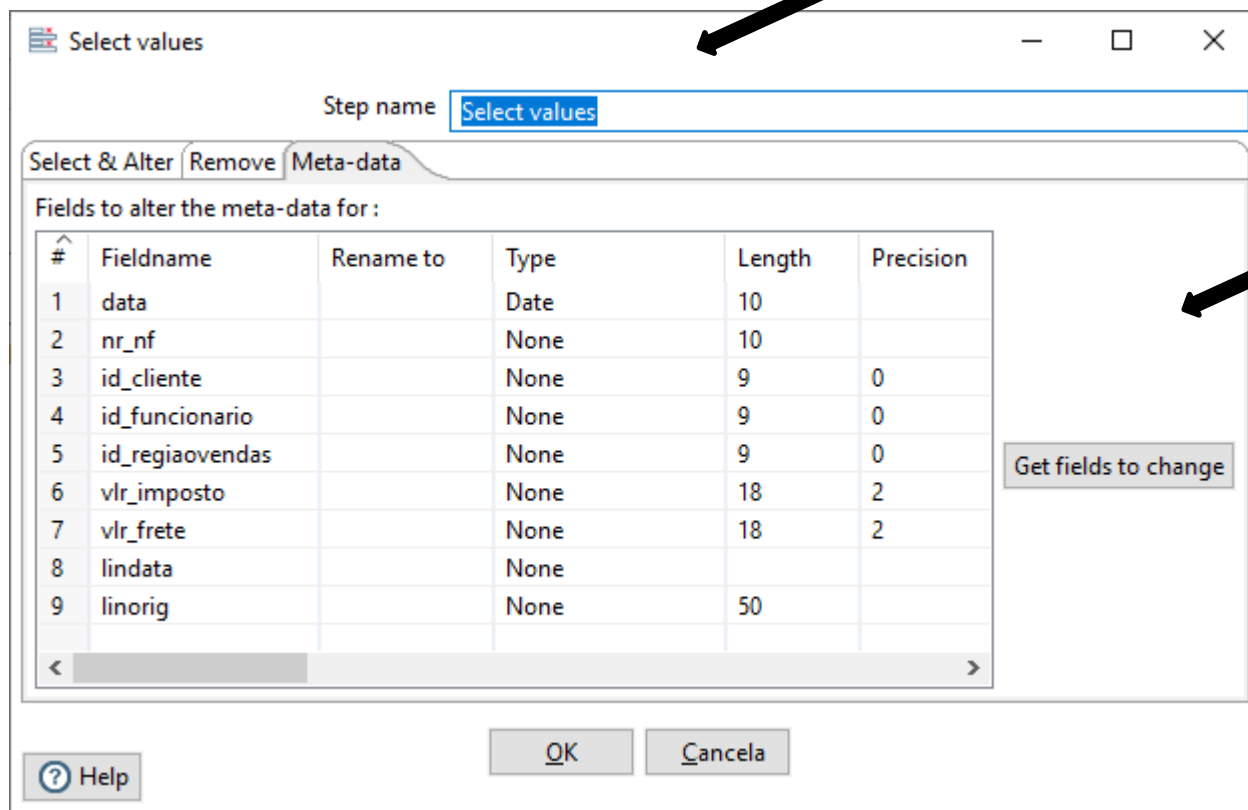
Obtém os nomes dos campos da tabela f_venda no banco de dados DS



DS – Fatos Vendas no Data Stage

- 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	Precision
1	data		Date	10	
2	nr_nf		None	10	
3	id_cliente		None	9	0
4	id_funcionario		None	9	0
5	id_regiaoavendas		None	9	0
6	vlr_imposto		None	18	2
7	vlr_frete		None	18	2
8	lindata		None		
9	linorig		None	50	

Get fields to change

OK Cancela Help

Na aba Meta-data são selecionados 9 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 F_Vendas da Fato Vendas.



DS – Fatos Vendas 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 F_Venda que é selecionada pelo botão de navegação

Combinação lookup / update

Nome do Step: Combination lookup/update

Connection: DW

Target schema:

Tabela de destino: F_Venda

Confirma tamanho: 100

Tamanho do Cache: 9999

Pre-load the cache? ☐

Campos Chave (para verificar linha na tabela):

#	Campo Dimensão	Campo no fluxo
1	data	data
2	nr_nf	nr_nf
3	id_cliente	id_cliente
4	id_funcionario	id_funcionario
5	id_regiao vendas	id_regiao vendas
6	Vlr_Imposto	vlr_imposto
7	Vlr_frete	Vlr_frete
8	lindata	lindata
9	linorig	linorig

Campo chave técnica: Id_Fato

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 – Fatos Vendas 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 F_Venda 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_Fato a partir de 1 e é incrementado de uma unidade.



Nome do Step: Combination lookup/update

Connection: DW

Target schema:

Tabela de destino: F_Venda

Confirma tamanho: 100

Tamanho do Cache: 9999

Pre-load the cache? ☐

Campos Chave (para verificar linha na tabela):

#	Campo Dimensão	Campo no fluxo
1	data	data
2	nr_nf	nr_nf
3	id_cliente	id_cliente
4	id_funcionario	id_funcionario
5	id_regiao vendas	id_regiao vendas
6	Vlr_Imposto	vlr_imposto
7	Vlr_frete	Vlr_frete
8	lindata	lindata
9	linorig	linorig

Campo chave técnica: Id_Fato

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