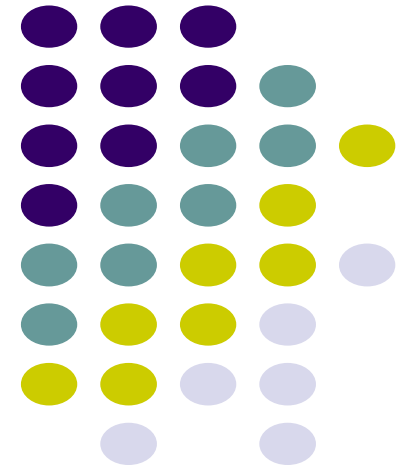


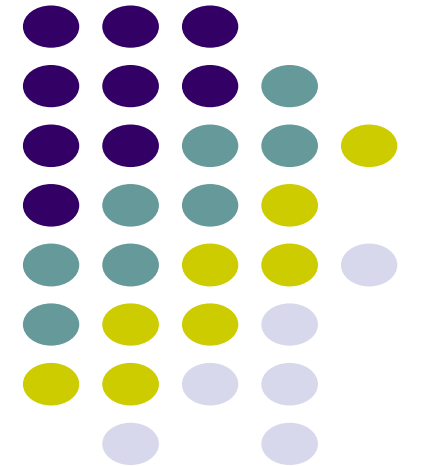


# Business Intelligence: O desenho do Data Warehouse

Vitor Valerio de Souza Campos



# Processo para criar o Data Stage e o Data Warehouse





# DS – Dimensão Produto no Data Stage

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

D\_Produto

Cod_Produto (varchar(20))
Nome (varchar(50))
Tamanho (varchar(05))
Linha (varchar(05))
Cor (varchar(20))
Ativo (varchar(01))
LinData (date)
LinOrig (varchar(50))



# DS – Dimensão Produto no Data Stage

- Script da tabela chamada D\_Produto.

```
CREATE TABLE D_Produto(  
    Cod_Produto varchar(20) NOT NULL,  
    Nome varchar(50) NOT NULL,  
    Tamanho varchar(5) NOT NULL,  
    Cor varchar(20) NOT NULL,  
    LinData date NOT NULL,  
    LinOrig varchar(50) NOT NULL  
);
```

```
create index IX_ProdutoNM on D_Produto (nome);
```

```
create index IX_ProdutoCod on D_Produto (Cod_Produto);
```



# DW – Dimensão Produto

- Script da tabela chamada D\_Produto.

```
CREATE TABLE D_Produto(  
    Id_Produto int NOT NULL,  
    Cod_Produto varchar(20) NULL,  
    Nome varchar(50) NULL,  
    Tamanho varchar(5) NULL,  
    Cor varchar(20) NULL,  
    LinData date NULL,  
    LinOrig varchar(50) NULL,  
    version int NULL,  
    date_from timestamp NULL,  
    date_to timestamp NULL,
```



# DW – Dimensão Produto

- Script da tabela chamada D\_Produto.

```
CONSTRAINT PK_D_Produto PRIMARY KEY  
(  
    Id_Produto  
)  
);
```

```
CREATE INDEX idx_D_Produto_lookup ON D_Produto  
(Cod_Produto, Nome, Tamanho, Cor, Ativo, LinData, LinOrig)  
;  
CREATE INDEX idx_D_Produto_tk ON D_Produto (Id_Produto)  
;
```



# DS – Dimensão Produto no Data Stage

- Algumas considerações:
  - Essa dimensão é a que pode sofrer alterações ao longo do tempo.
  - Primeiramente, precisamos entender qual alteração que levaremos em consideração. No nosso exemplo, o que queremos é que o nome do produto seja tratado como alteração. Isto é, se um mesmo produto receber um novo nome, teremos de estar preparados para capturar essa alteração. Como vimos no desenho do DW, tem-se mais de uma forma de se resolver essa questão, mas partimos para a chamada 2SCD.



# DS – Dimensão Produto no Data Stage

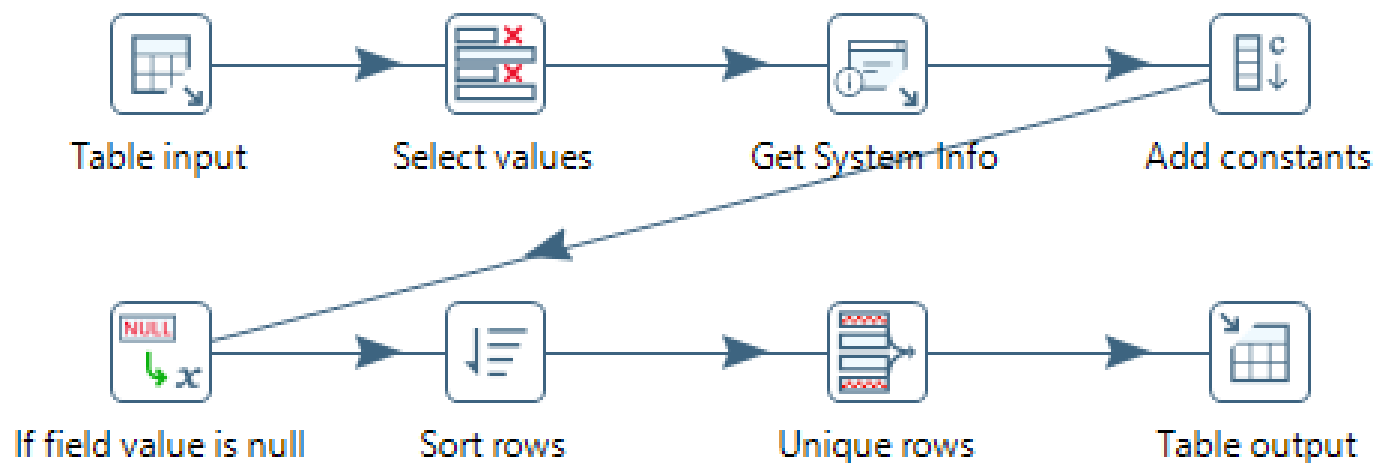
- As transformações no pentaho para fazer carga dos dados deverá responder pelos seguintes critérios:
  - Teremos de carregar os dados de novos produtos normalmente, fazendo com que as colunas da nova versão e validade dos dados seja o produto ATIVO.
  - Precisaremos fazer um update na coluna NOME se ocorrer de um produto receber um novo nome para um mesmo Cod\_Produto.
  - 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 Produto no Data Stage

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





# DS – Dimensão Produto no Data Stage

- Algumas observações:
  - Passo Table Input da transformação D\_Produto da dimensão Produto possui os mesmos dados de entrada que os da dimensão Tempo.
  - O Passo Select Values da transformação D\_Produto da dimensão Produto é similar ao passo Select Values BI da dimensão Tempo, só que nele tem-se o atributo de Produto 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\_Produto.



# DS – Dimensão Produto no Data Stage

- Passo Select values.

Nome do  
passo

Select values

Step name:

Select & Alter Remove Meta-data

Fields to alter the meta-data for :

#	Fieldname	Rename to	Type	Length
1	Cod_Produto		None	20
2	Produto	Nome	None	50
3	Tamanho		None	50
4	Cor		None	50

Get fields to change

OK Cancela Help

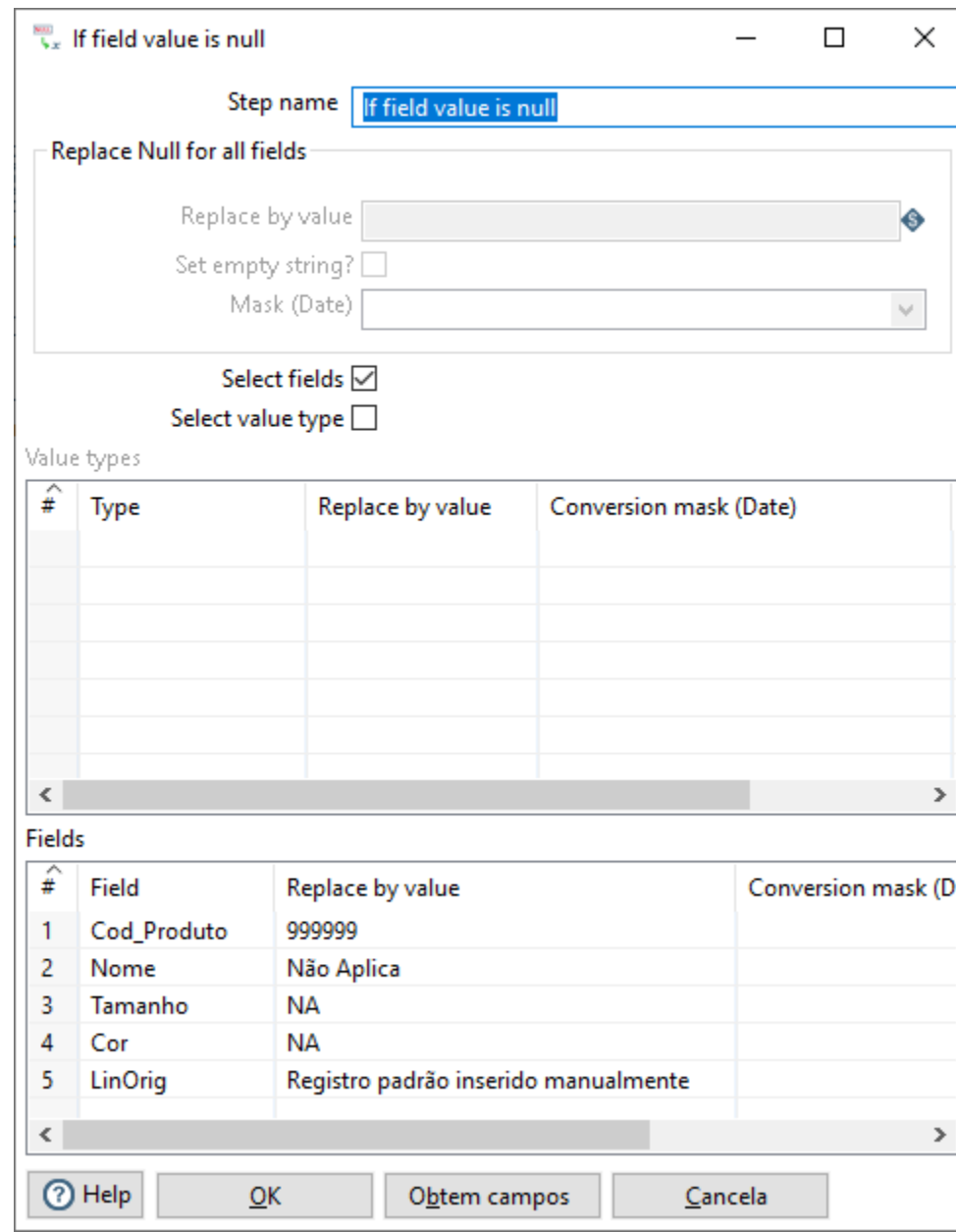
Na aba Meta-data são selecionados 4 campos, conforme mostrado na figura. Para obter este campo foi pressionado o botão Get fields to change e removidos os campos não pertencente a tabela D\_Produto da dimensão Geografia.



# DS – Dimensão Produto no Data Stage

- Passo If field value is null

Na parte inferior, no item Fields utilizamos o botão Obtêm campos para selecionar os campos Cod\_Produto, Nome, Tamanho, Cor, e LinOrig e na coluna Replace by value adicionar respectivamente 999999, Não Aplica, NA, NA e Registro padrão inserido manualmente.



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 mask (D)
1	Cod_Produto	999999	
2	Nome	Não Aplica	
3	Tamanho	NA	
4	Cor	NA	
5	LinOrig	Registro padrão inserido manualmente	

? Help OK Obtêm campos Cancela



# DS – Dimensão Produto no Data Stage

- Passo Sort rows

Na parte inferior, no item Fields utilizamos o botão Obtem campos para selecionar os campos Cod\_Produto e Nome e ordenamos de forma ascendente.

Nome do passo

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	Cod_Produto	S	N	N
2	Nome	S	N	N

< >

Help OK Cancela Obtem campos



# DS – Dimensão Produto no Data Stage

- Passo Unique rows

Na parte inferior, no item Fields to compare on utilizamos o botão Get para selecionar os campos Cod\_Produto e Nome.  
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	Cod_Produto	N
2	Nome	N

Help OK Cancela Get



# DS – Dimensão Produto no Data Stage

- Passo Table output.

Nome do Step: Table output

Connection: DS

Target schema:

Target table: d\_produto

Commit size: 1000

Truncate table: ☐

Ignore insert errors: ☐

Specify database fields: ☒

Main options Database fields

Colunas a inserir:

#	Table field	Stream field
1	Cod_Produto	Cod_Produto
2	Nome	Nome
3	Tamanho	Tamanho
4	Cor	Cor
5	LinData	LinData
6	LinOrig	LinOrig

Buttons: 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\_Produto.

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



# DS – Dimensão Produto no Data Stage

- Algumas observações:
  - Sempre apagamos a tabela D\_Produto do DS para iniciar uma carga sem resquícios de cargas anteriores.
  - A transformação e (quando houver) validação dos dados ocorrem na inserção na tabela D\_Produto no DS. Quando os dados forem ser inseridos no DW, já deverão estar ok.
  - A execução da mesma transformação no Pentaho não acrescenta dados já existentes na tabela D\_Produto no Data Warehouse. Só vai ser inserido dados se os mesmos não existirem nele.





# DS – Dimensão Produto no Data Warehouse

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





# DS – Dimensão Produto no Data Warehouse

- Passo Table input.

Nome do passo

Letura de Tabela

Nome do Step: Table input

Connection: DS

SQL:

```
SELECT  
  Cod_Produto  
  , Nome  
  , Tamanho  
  , Cor  
  , LinData  
  , LinOrig  
FROM D_Produto
```

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\_produto no banco de dados DS



# DS – Dimensão Produto no Data Warehouse

- Passo Dimension 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. Observe que tem a opção Update the dimension que se ativa atualiza a dimensão com base nas informações no fluxo de entrada; se não estiver ativado, a dimensão realiza apenas pesquisas e adiciona o campo de chave técnica aos fluxos.

Dimension lookup/update

Step name: Dimension lookup/update

Update the dimension? ☒

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

Target schema: public [Navega...]

Target table: d\_produto [Browse...]

Commit size: 100

Enable the cache? ☐

Pre-load the cache? ☐

Cache size in rows (0 = cache all):

Keys Fields

Key fields (to look up row in dimension):

#	Dimension field	Field in stream
1	cod_produto	cod_produto

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



# DS – Dimensão Produto no Data Warehouse

- Passo Dimension lookup/update.

Step name: Dimension lookup/update

Update the dimension? ☒

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

Target schema: public [Navega...]

Target table: d\_produto [Browse...]

Commit size: 100

Enable the cache? ☐

Pre-load the cache? ☐

Cache size in rows (0 = cache all):

Keys Fields

Key fields (to look up row in dimension):

#	Dimension field	Field in stream
1	cod_produto	cod_produto

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\_Produto no DW nada é inserido caso contrário novos valores são inseridos.

Eles são obtidos a partir do botão Obtem Campos.



# DS – Dimensão Produto no Data Warehouse

- Passo Dimension lookup/update.

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

O item Version field recebe no nome de version e tem uma faixa de datas válidas, que no caso foi escolhido 1900 a 2199.

Technical key field: id\_produto (dropdown) New name:

Creation of technical key:  
☒ Use table maximum + 1  
☐ Use sequence   
☐ Use auto increment field

Version field: version (dropdown)

Stream Datefield:

Date range start field: date\_from (dropdown) Min. year: 1900

Use an alternative start date? ☐ <Select Option> (dropdown)

Table date range end: date\_to (dropdown) Max. year: 2199

Buttons: OK, Cancela, Get Fields, SQL

Help: Help



# DS – Dimensão Produto no Data Stage

- Algumas considerações:
  - Ao carregarmos os produtos — por ser a primeira carga —, não teremos nenhuma alteração de nome, então serão 60 produtos carregados. Quando formos efetuar a carga dos próximos arquivos (dos demais anos que existem na nossa origem), faremos uma alteração no nome de um dos produtos a fim de testar a 2SCD.