

## DIMENSÃO PERDIDA NO DATA WAREHOUSE

Hoje trouxe uma velha dúvida de diversos profissionais da Área de Business Intelligence, recentemente um aluno me intrigou sobre a pergunta: por que adicionar o registro: "INFORMAÇÃO NÃO IDENTIFICADA" na nossa dimensão, e a **sk** deverá ter um valor (Alguns colocam 0 outros -1 e -2).

Quantas vezes não nos pegamos com a seguinte situação: Em um sistema transacional temos um elemento órfão por exemplo um código, uma data digitada errada, ou até mesmo uma informação da tabela que não tem dados para uma determinada linha.

### Vamos aos exemplos:

Num_nf	Cod_produto	Cod_vendedor	Data_venda	Val_venda
1	2	1	12/06/2014	550,00
2	1		18/06/2014	450,00
3	2	2	31/12/2014	550,00

STG\_VENDA  
DIMENSAO\_PRODUTO

Sk_produto	Cod_produto	Des_produto	Dtc_inicio	Dtc_fim
-1	-1	Produto Não Informado	01/01/1900	
1	1	Monitor	01/01/1900	
2	2	Memória RAM	01/01/1900	

Neste primeiro exemplo, é possível verificar que na tabela de origem que iremos carregar nossa fato (STG\_VENDA), temos todos os códigos de produto com a mesma referência na nossa dimensão produto, porém o mesmo não acontece para o nosso campo Cod\_vendedor, onde é possível verificar que uma das linha temos o valor nulo para o campo.

DIMENSÃO VENDEDOR

Sk_vendedor	Cod_vendedor	Des_vendedor	Dtc_inicio	Dtc_fim
-1	-1	Vendedor Não Informado	01/01/1900	
1	1	Vinicius Oliveira	01/01/1900	
2	2	Fulano de Tal	01/01/1900	

Na nossa dimensão é possível ver que existe uma sk **-1**, essa sk indica que a informação de vendedor não foi localizada, assim na nossa fato será adicionado a sk correspondente ao valor nulo a "sk **-1**", isso permite que a informação não seja desprezadas e que as análises possam ser feitas com os valores reais.

Imagine se o gestor criasse um relatório de vendas por produtos e essa linha não fosse carregada por falta do código de vendedor, logo esse relatório estaria inconsistente. Para não correr esse risco vamos criar essa sk nas nossas dimensões.

Num_nf	Sk_produto	Sk_vendedor	Sk_tempo	Val_venda
1	2	1	x	550,00
2	1	-1	x	450,00
3	2	2	x	550,00

FATO VENDA

Na segunda linha percebemos que na coluna sk\_vendedor temos o valor -1 que representa uma informação na encontrada na dimensão. Agora você pode está se perguntando e se fosse um valor inconsistente ao invés de nulo, por exemplo:

Num_nf	Cod_produto	Cod_vendedor	Data_venda	Val_venda
1	2	1	12/06/2014	550,00
2	1	5	18/06/2014	450,00
3	2	2	31/12/2104	550,00

#### STG\_VENDA

Bom nesse exemplo gera uma dúvida, a informação existe na tabela porém não temos a mesma na nossa dimensão. Nesse caso colocamos -1?

Sim na maioria dos projetos que trabalho colocamos esse registro como -1, porém em outros é possível fazer da seguinte forma:

Sk_vendedor	Cod_vendedor	Des_vendedor	Dtc_inicio	Dtc_fim
-1		Vendedor Não Informado	01/01/1900	
-2		Informação inconsistente	01/01/1900	
1	1	Vinicius Oliveira	01/01/1900	
2	2	Fulano de Tal	01/01/1900	

#### DIMENSÃO VENDEDOR

Veja que no exemplo criamos a sk -2 e nela adicionamos a informação que esse elemento é inconsistente, logo nossa fato ficaria da seguinte forma:

Num_nf	Sk_produto	Sk_vendedor	Sk_tempo	Val_venda
1	2	1	x	550,00
2	1	-2	x	450,00
3	2	2	x	550,00

#### FATO VENDA

Veja que agora temos uma sk -2 assim percebemos que para essa fato a informação existe, porém a mesma é inconsistente.

#### Conclusão

O Tratamento de informações no Data Warehouse é muito importante para o seu desenvolvimento e confiabilidade, esse tipo de informação parece meio utópico principalmente por achar que essa situação não retrata seu ambiente operacional. Por experiência própria no desenvolvimento de diversos projetos, posso afirmar que esse é um problema comum e que deveremos está sempre atentos para essa situação.

Ficou com dúvida ou gostaria de realizar um elogio, crítica ou sugestão? Nos envie sua mensagem através de nosso site: [www.analiseerestado.com.br](http://www.analiseerestado.com.br)

[Clique aqui](#) e faça parte do nosso grupo no linkedin  
[Clique aqui](#) e curta nossa página no facebook  
[Clique aqui](#) e cadastre-se para receber novos artigos.

Realizado por:

Vinicius Oliveira – Ciências da Computação e MBA em Business Intelligence (BI).