

Slowly Changing Dimensions Tutorial

Pretende-se criar um tutorial para simular Slowly Changing Dimensions (SCD), usando um caso plausível para criar o mesmo. Para isso, será implementada uma **SCD do tipo 2**.

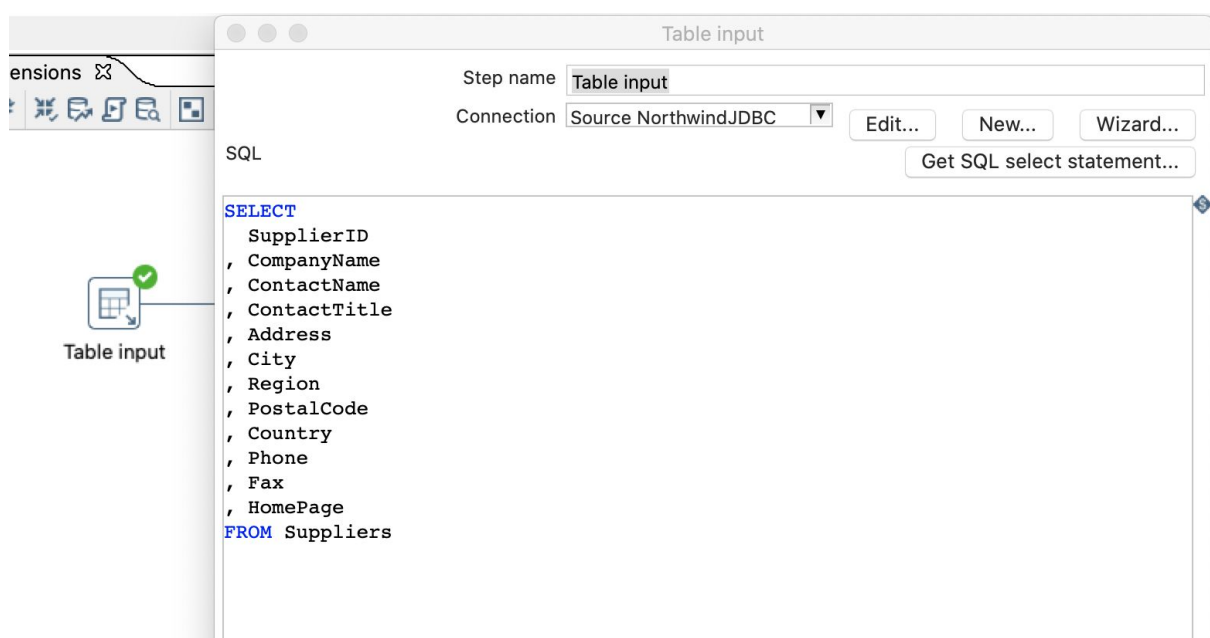
As **atualizações do Tipo 2** são também conhecidas pela **inserção da versão das linhas nas tabelas de dimensão**. Isto é, uma mudança é capturada ao inserir-se uma nova linha com os valores do atributo atualizador, deixando a linha existente sem modificação.

Este tipo é utilizado para **mudanças que se registam ao longo do tempo**, como por exemplo, para updates de cidades de utilizadores. Um utilizador pode mudar o seu endereço, mas para efeitos de análise de dados, talvez seja **relevante manter o histórico** todo do utilizador.

Como as tabelas de dimensão, normalmente, utilizam **menos espaço** que as tabelas de factos para armazenar dados, aumentar a tabela de dimensão em questão para conter as várias versões acaba por não ser um problema.

Segue-se um pequeno tutorial de como simular uma atualização deste tipo. Para isso, recorreu-se à dimensão *Supplier* da base de dados *Northwind*, e criou-se uma nova tabela na base de dados *target* na máquina local, réplica da tabela *Suppliers* da base de dados *source*. A ideia passa por analisar a tabela do *source*, e quando ocorrerem mudanças, estas serão inseridas na *target*.

Step 1: Criar um *Table Input* que irá seleccionar os dados da tabela *Suppliers* da base de dados *source*, neste caso *Northwind*. Será a tabela que estará a ser visualizada para detectar novos *updates*.

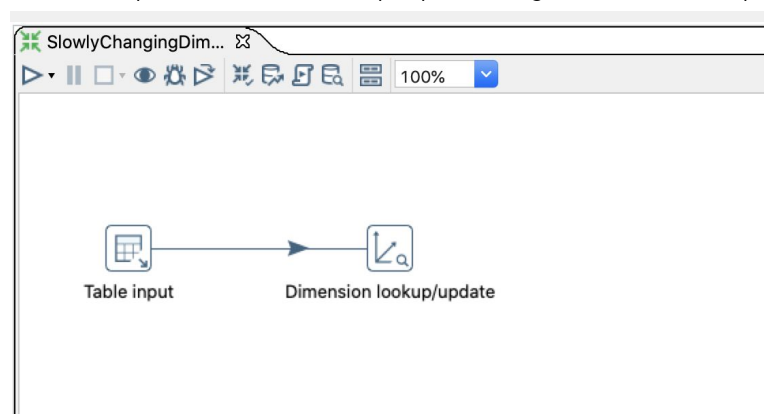


Step 2: Criar uma nova tabela na base de dados *target*, com os mesmos atributos que os da tabela *Suppliers* da *source*, adicionando os seguintes atributos:

- **technical_key:** É uma coluna extra que é preciso adicionar à nossa tabela. Esta tabela surge devido à replicação que ocorre na estratégia *Slowly Changing Dimensions* do Tipo 2, onde as chaves primárias da tabela *source* não podem ser chaves primárias da tabela *target* porque poderão ocorrer mudanças
- **version:** coluna que regista a versão do registo na base de dados
- **date_from e date_to:** registam o intervalo de tempo em que o registo foi válido (período de validade). O intervalo é composto por uma data inicial (*date_from*) e outra final (*date_to*), que são determinadas automaticamente durante a etapa.

```
CREATE TABLE SUPPLIER_NEW (  
  id_supplier INTEGER,  
  technical_key SERIAL PRIMARY KEY,  
  companyName VARCHAR(40),  
  contactName VARCHAR(30),  
  contactTitle VARCHAR(30),  
  supplierAddress VARCHAR(60),  
  cityName VARCHAR(15),  
  regionName VARCHAR(30),  
  postalCode VARCHAR(30),  
  countryName VARCHAR(30),  
  phone VARCHAR(60),  
  fax VARCHAR(30),  
  homePage VARCHAR(100),  
  version INTEGER,  
  date_from DATE,  
  date_to DATE  
);
```

Step 3: Inserir um novo step “*Dimension lookup/update*”, ligado ao Table Input anterior.



Para configurar este step, é necessário estabelecer-se uma ligação à base de dados *target* e indicar qual a tabela que será analisada e que conterá os *updates*, neste caso, a tabela criada anteriormente.

Dimension lookup/update

Step name: Dimension lookup/update

Update the dimension? ☒

Connection: Target - posteareSQL [Edit... New... Wizard... Navega...]

Target schema: [Browse...]

Target table: supplier_new [Browse...]

Commit size: 100

Enable the cache? ☐

Pre-load the cache? ☐

Cache size in rows (0 = cache all):

Step 4: Definir uma coluna de identificação dos dados de origem, que é usada para procurar linhas armazenadas anteriormente com a mesma key, para que o step possa comparar a linha recebida com as linhas já armazenadas na *target* e avaliar se a linha foi alterada ou não. Também é preciso definir o *Technical Key Field*, neste caso fica o atributo *technical_key* definido no processo de criação da nova tabela, indicar qual o campo de versão, e o *date range*, como mostra a figura.

Keys Fields

Key fields (to look up row in dimension):

#	Dimension field	Field in stream
1	id_supplier	SupplierID

Technical key field: technical key [New name]

Creation of technical key

☒ Use table maximum + 1

☐ Use sequence

☐ Use auto increment field

Version field: version

Stream Datefield:

Date range start field: date from [Min. year: 1900]

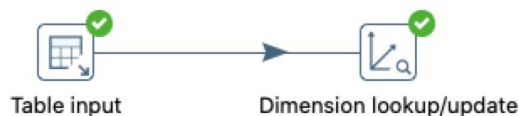
Use an alternative start date? ☐ <Select Option>

Table date range end: date to [Max. year: 2199]

Step 5: Para ultimar este step, é necessário ainda definir os campos que poderão ser alvo de mudanças.

Keys			
Fields			
Lookup/Update fields			
#	Dimension field	Stream field to compare with	Type of dimension update
2	contactName	ContactName	Insert
3	contactTitle	ContactTitle	Insert
4	supplierAddress	Address	Insert
5	cityName	City	Insert
6	regionName	Region	Insert
7	postalCode	PostalCode	Insert
8	countryName	Country	Insert
9	phone	Phone	Insert

Correndo a transformação criada pela primeira vez, a tabela na base de dados target é populada com todos os dados, sendo estas informações consideradas como primeira versão dos dados. Sempre que se realizar uma operação de update de um campo, é inserida uma nova linha associada ao *id_supplier* que está a sofrer alterações, e é registada como uma segunda versão da informação.



```

edados=# select id_supplier, technical_key, date_from, date_to, version, companyName, contactName, supplierAddress from supplier_new;
id_supplier | technical_key | date_from | date_to | version | companyname | contactname | supplieraddress
-----
1 | 1 | 1980-01-01 | 2199-12-31 | 1 | Exotic Liquids | Charlotte Cooper | 49 Gilbert St.
2 | 2 | 1980-01-01 | 2199-12-31 | 1 | New Orleans Cajun Delights | Shelley Burke | P.O. Box 78934
3 | 3 | 1980-01-01 | 2199-12-31 | 1 | Grandma Kelly's Homestead | Regina Murphy | 707 Oxford Rd.
4 | 4 | 1980-01-01 | 2199-12-31 | 1 | Tokyo Traders | Yoshi Nagase | 9-8 Sekimai Musashino-shi
5 | 5 | 1980-01-01 | 2199-12-31 | 1 | Cooperativa de Quesos 'Las Cabras' | Antonio del Valle Saavedra | Calle del Rosal 4
6 | 6 | 1980-01-01 | 2199-12-31 | 1 | Mayumi's | Mayumi Ohno | 92 Setsuko Chuo-ku
7 | 7 | 1980-01-01 | 2199-12-31 | 1 | Pavlova, Ltd. | Ian Devling | 74 Rose St. Moonie Ponds
8 | 8 | 1980-01-01 | 2199-12-31 | 1 | Specialty Biscuits, Ltd. | Peter Wilson | 29 King's Way
9 | 9 | 1980-01-01 | 2199-12-31 | 1 | PB Knäckebröd AB | Lars Peterson | Kaloadagatan 13
10 | 10 | 1980-01-01 | 2199-12-31 | 1 | Refrescos Americanas LTDA | Carlos Diaz | Av. das Americanas 12.898
11 | 11 | 1980-01-01 | 2199-12-31 | 1 | Heli Süßwaren GmbH & Co. KG | Petra Winkler | Tiergartenstraße 5
12 | 12 | 1980-01-01 | 2199-12-31 | 1 | Plutzer Lebensmittelgroßmärkte AG | Martin Bein | Bogenallee 51
13 | 13 | 1980-01-01 | 2199-12-31 | 1 | Nord-Ost-Fisch Handelsgesellschaft mbH | Sven Petersen | Frhmredder 112a
14 | 14 | 1980-01-01 | 2199-12-31 | 1 | Formaggi Fortini s.r.l. | Elio Rossi | Viale Dante, 75
15 | 15 | 1980-01-01 | 2199-12-31 | 1 | Norske Meierier | Beate Vileid | Hatlevegen 5
16 | 16 | 1980-01-01 | 2199-12-31 | 1 | Bigfoot Breweries | Cheryl Saylor | 3400 - 8th Avenue Suite 210
17 | 17 | 1980-01-01 | 2199-12-31 | 1 | Svensk Sjöföda AB | Michael Björn | Brovallavägen 231
18 | 18 | 1980-01-01 | 2199-12-31 | 1 | Aux joyeux ecclésiastiques | Guylène Nodier | 203, Rue des Francs-Bourgeois
19 | 19 | 1980-01-01 | 2199-12-31 | 1 | New England Seafood Cannery | Robb Merchant | Order Processing Dept. 2100 Paul Revere Blvd.
20 | 20 | 1980-01-01 | 2199-12-31 | 1 | Leka Trading | Chandre Leka | 471 Serangoon Loop, Suite #402
21 | 21 | 1980-01-01 | 2199-12-31 | 1 | Lyngbysild | Niels Petersen | Lyngbysild Fiskebakken 10
22 | 22 | 1980-01-01 | 2199-12-31 | 1 | Zaanse Snoepfabriek | Dirk Luchte | Verkoop Rijnweg 22
23 | 23 | 1980-01-01 | 2199-12-31 | 1 | Karkki Oy | Anne Heikkonen | Valtakatu 12
24 | 24 | 1980-01-01 | 2199-12-31 | 1 | O'Guy, Mate | Wendy Mackenzie | 170 Prince Edward Parade Hunter's Hill
25 | 25 | 1980-01-01 | 2199-12-31 | 1 | Ma Maison | Jean-Guy Lauzon | 2960 Rue St. Laurent
26 | 26 | 1980-01-01 | 2199-12-31 | 1 | Pasta Buttini s.r.l. | Giovanni Giudici | Via dei Gelsomini, 153
27 | 27 | 1980-01-01 | 2199-12-31 | 1 | Escargots Nouveaux | Marie Delamare | 22, rue H. Voiron
28 | 28 | 1980-01-01 | 2199-12-31 | 1 | Gai pâturage | Eliane Noz | Bat. B 3, rue des Alpes
29 | 29 | 1980-01-01 | 2199-12-31 | 1 | Forêts d'érables | Chantal Goulet | 148 rue Chasseur
(30 rows)
  
```