Data Mart Implementation (P01)

DECISION SUPPORT SYSTEMS, 2022-23

**João Apresentação (21152), Gonçalo Cunha (21145), Pedro Simões (21140)**

Índice

[1. Introdução 3](#_Toc132041819)

[1.1 Visão 3](#_Toc132041820)

[1.2 Objetivos 3](#_Toc132041821)

[1.3 Descrição de Processos de Negócio 4](#_Toc132041822)

[2. Recurso de dados 5](#_Toc132041823)

[3. Modelação Dimensional 6](#_Toc132041824)

[3.1 Design do modelo de dados dimensional 7](#_Toc132041825)

[4. Implementação da DataMart 11](#_Toc132041826)

[4.1 Jobs 11](#_Toc132041827)

[4.2 Transformations 12](#_Toc132041828)

[5. Conclusão 21](#_Toc132041829)

**Índice de imagens**

[Figura 1 - WWI Modelo BaseDados 6](file:///C:\Universidade\3º%20Ano%20-%20Semestre%202\SAD\ETL\Relatório\DSS_G07_P01.docx#_Toc132043312)

[Figura 2 – Job\_dims\_ft 12](file:///C:\Universidade\3º%20Ano%20-%20Semestre%202\SAD\ETL\Relatório\DSS_G07_P01.docx#_Toc132043313)

[Figura 3 - Job\_dims 12](file:///C:\Universidade\3º%20Ano%20-%20Semestre%202\SAD\ETL\Relatório\DSS_G07_P01.docx#_Toc132043314)

[Figura 4 - Transformation Dim\_Costumers 13](file:///C:\Universidade\3º%20Ano%20-%20Semestre%202\SAD\ETL\Relatório\DSS_G07_P01.docx#_Toc132043315)

[Figura 5 -Transformation Dim\_CostumerTransactions 14](file:///C:\Universidade\3º%20Ano%20-%20Semestre%202\SAD\ETL\Relatório\DSS_G07_P01.docx#_Toc132043316)

[Figura 6 - Transformation Dim\_StockItems 15](file:///C:\Universidade\3º%20Ano%20-%20Semestre%202\SAD\ETL\Relatório\DSS_G07_P01.docx#_Toc132043317)

[Figura 7 - Transformation Ft\_invoiceLines 16](file:///C:\Universidade\3º%20Ano%20-%20Semestre%202\SAD\ETL\Relatório\DSS_G07_P01.docx#_Toc132043318)

[Figura 8 - ft\_invoiceLines get\_max 16](file:///C:\Universidade\3º%20Ano%20-%20Semestre%202\SAD\ETL\Relatório\DSS_G07_P01.docx#_Toc132043319)

[Figura 9 - ft\_invoiceLines input 16](file:///C:\Universidade\3º%20Ano%20-%20Semestre%202\SAD\ETL\Relatório\DSS_G07_P01.docx#_Toc132043320)

[Figura 10 - ft\_invoicelines CostumerLookup 17](file:///C:\Universidade\3º%20Ano%20-%20Semestre%202\SAD\ETL\Relatório\DSS_G07_P01.docx#_Toc132043321)

[Figura 11 - ft\_invoicelines StockItemsLookup 17](file:///C:\Universidade\3º%20Ano%20-%20Semestre%202\SAD\ETL\Relatório\DSS_G07_P01.docx#_Toc132043322)

[Figura 12 - ft\_invoicelines SelectValues 18](file:///C:\Universidade\3º%20Ano%20-%20Semestre%202\SAD\ETL\Relatório\DSS_G07_P01.docx#_Toc132043323)

[Figura 13 - ft\_invoicelines output 18](file:///C:\Universidade\3º%20Ano%20-%20Semestre%202\SAD\ETL\Relatório\DSS_G07_P01.docx#_Toc132043324)

[Figura 14 - ft\_orderlines get\_max 19](file:///C:\Universidade\3º%20Ano%20-%20Semestre%202\SAD\ETL\Relatório\DSS_G07_P01.docx#_Toc132043325)

[Figura 15 - ft\_orderlines input 19](file:///C:\Universidade\3º%20Ano%20-%20Semestre%202\SAD\ETL\Relatório\DSS_G07_P01.docx#_Toc132043326)

[Figura 16 - ft\_orderlines CostumersLookup 20](file:///C:\Universidade\3º%20Ano%20-%20Semestre%202\SAD\ETL\Relatório\DSS_G07_P01.docx#_Toc132043327)

[Figura 17 - ft\_orderlines StockLookup 20](file:///C:\Universidade\3º%20Ano%20-%20Semestre%202\SAD\ETL\Relatório\DSS_G07_P01.docx#_Toc132043328)

[Figura 19 - ft\_orderlines SelectValues 21](file:///C:\Universidade\3º%20Ano%20-%20Semestre%202\SAD\ETL\Relatório\DSS_G07_P01.docx#_Toc132043329)

# Introdução

Este projeto de uma forma resumida terá a implementação de um Data Mart com recurso a uma base de dados que armazena atividades de uma empresa de importação e distribuição de produtos, no caso em estudo, Wide World Importers (WWI).

A empresa WWI é uma empresa de importação e distribuição de produtos, que atende principalmente a empresas que revendem para consumidores individuais (público geral). Os clientes da WWI incluem lojas de especialidades, supermercados, lojas de informática, lojas de atrações turísticas e alguns indivíduos.

O projeto passará pela análise de requisitos de negócios, montagem de dimensões e tabelas de factos e implementação de processos ETL.

Nesta introdução, será falada a visão geral do projeto, os objetivos e uma descrição dos processos de negócios envolvidos.

## Visão

O objetivo principal deste projeto é criar um Data Mart eficiente e preciso que possa ser usado para análises de negócios na empresa WWI.

O Data Mart será projetado para fornecer as respostas necessárias e relevantes para os utilizadores, permitindo que estes tomem decisões com base nessas informações.

## Objetivos

Os objetivos deste projeto passam por:

* Realizar a análise de requisitos de negócios para entender as necessidades da empresa WWI;
* Realizar a análise de perfil de dados para entender a qualidade e a integridade dos dados disponíveis;
* Projetar e implementar processos de extração, transformação e carregamento dos dados para a Data Mart;
* Projetar e implementar um modelo dimensional que atenda aos requisitos de negócios da WWI;
* Executar os processos ETL para garantir a integridade dos dados no Data Mart.

## Descrição de Processos de Negócio

Os processos de negócios da WWI inicia-se pelo pedido de produtos, da parte do cliente. Se a WWI não tiver stock suficiente, eles encomendam o stock adicional dos fornecedores. Se os clientes não quiserem esperar pelos produtos que não estão disponíveis, os produtos são enviados posteriormente em um envio separado.

A WWI rentabiliza através dos produtos em stock, convertendo o pedido em fatura. Quando os clientes fazem pedidos de produtos que não estão em stock, esses produtos são colocados em backorder. A WWI entrega os produtos em stock aos clientes, seja por meio de seus próprios veículos de entrega ou por meio de outros correios ou métodos de frete.

Os clientes pagam as faturas à WWI. As faturas e pagamentos são registados na tabela de transações de clientes.

# Recurso de dadosUma imagem com diagrama Descrição gerada automaticamente

Figura 1 - WWI Modelo BaseDados

Table 1: Summary of WWI database contents

|  |  |  |
| --- | --- | --- |
| **Event / object** | **Table** | **Nr. Records** |
| Clientes | *Customers* | 663 |
| Encomendas feitas | *Orders* | 73595 |
| Registo de Faturas | *Invoices* | 70510 |
| Transações de dinheiro entre WWI e Costumers | *Customertransactions* | 97147 |
| Inventário de produtos | *Stockitems* | 227 |
| Lista de descontos | *Specialdeals* | 2 |

# Modelação Dimensional

Table 2: Data Warehouse Matrix

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **DIMENSIONS**  **BUSINESS PROCESSES** | Dim\_costumers | Dim\_stockitems | Dim\_costumertransactions | Ft\_orderlines | Ft\_invoicelines |
| Top de países que compraram mais o item X | X | X |  | X |  |
| Top de items mais comprados no WWI |  | X |  | X |  |
| Top pessoas que mais fizeram encomendas | X |  |  | X |  |
| Mês que tem mais vendas em quantidade |  |  |  |  | X |
| Top faturas com mais lucrativas |  |  |  |  | X |
| Média da quantidade de vendas de sempre |  |  |  |  | X |
| Top de países que mais dinheiro gastaram com o WWI | X |  |  |  | X |
| Top marcas mais vendidas |  | X |  |  | X |
| Qual a categoria de clientes que mais gera receita à WWI | X |  |  |  | X |
| Qual é o período médio entre a receção do pedido e a entrega do produto? |  |  |  | X | X |
| Top de países que demora mais a receber produtos? | X |  |  | X | X |
| Média de lucro mensal |  |  |  |  | X |
| Percentagem de pedidos que cumpriram o prazo expectável de entrega |  |  |  | X | X |

## Design do modelo de dados dimensional

Abaixo está representado, em tabelas, informações sobre os atributos mais importantes de cada tabela

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Name** | **Type of table** | **Nr. Records** | | **Description** |
| dim\_costumers | Dimension | 664 | | Dimensão que apresenta os dados pessoais de um cliente |
| **Target (Data mart)** | | | | **Source (OLTP)** |
| **Column** | **SCD** | **Data type** | **Description** | |
| Costumerkey |  | Int |  | |
| Costumerid |  | Int |  | |
| Costumername |  | Varchar |  | |
| Cityname |  | Varchar |  | |
| Stateprovincename |  | Varchar |  | |
| Formalname |  | Varchar |  | |
| Costumercategoryname |  | Varchar |  | |
| Isoncredithold |  | Bit |  | |
| Phonenumber |  | Varchar |  | |
| Faxnumber |  | Varchar |  | |
| Deliveryadressline1 |  | Varchar |  | |
| Deliverypostalcode |  | Varchar |  | |
| Deliverylocation |  | Varchar |  | |
| Postaladressline1 |  | Varchar |  | |
| Postalpostalcode |  | Varchar |  | |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Name** | **Type of table** | **Nr. Records** | | **Description** |
| dim\_stockitems | Dimension | 228 | | Dimensão que apresenta os dados sobre os items em stock e suas características |
| **Target (Data mart)** | | | | **Source (OLTP)** |
| **Column** | **SCD** | **Data type** | **Description** | |
| Stockitemkey |  | Int |  | |
| Stockitemid |  | Int |  | |
| Stockitemname |  | Varchar |  | |
| Colorname |  | Varchar |  | |
| Unitprice |  | Double |  | |
| Brand |  | Varchar |  | |
| Size |  | Varchar |  | |
| Ischillerstock |  | Varchar |  | |
| Barcode |  | Varchar |  | |
| Taxrate |  | Numeric |  | |
| Unitprice |  | Numeric |  | |
| Recommendedretailprice |  | Numeric |  | |
| Typicalweightperunit |  | Numeric |  | |
| Tags |  | Text |  | |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Name** | **Type of table** | **Nr. Records** | | **Description** |
| dim\_customertransaction | Dimension | 26638 | | Dimensão que apresenta todas as transações que ocorreram no WWI |
| **Target (Data mart)** | | | | **Source (OLTP)** |
| **Column** | **SCD** | **Data type** | **Description** | |
| Customertransactionkey |  | Int |  | |
| Customertransactionid |  | Int |  | |
| Costumerid |  | Int |  | |
| Invoiceid |  | Varchar |  | |
| Transactiontypename |  | Int |  | |
| Paymentmethodname |  | Varchar |  | |
| Transactiondate |  | Datetime |  | |
| Taxammount |  | Numeric |  | |
| Transactionamount |  | Numeric |  | |
| Finalizationdate |  | Datetime |  | |
| Isfinalized |  | Bit |  | |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Name** | **Type of table** | **Nr. Records** | | **Description** |
| Ft\_orderlines | Fact Table | 1157060 | | Tabela de factos representante de cada linha de cada encomenda |
| **Target (Data mart)** | | | | **Source (OLTP)** |
| **Column** | **SCD** | **Data type** | **Description** | |
| Orderlineid |  | Int |  | |
| Orderid |  | Int |  | |
| Costumerkey |  | Int |  | |
| Stockitemkey |  | Int |  | |
| Orderdate |  | Datetime |  | |
| Orderprice |  | Numeric |  | |
| Expecteddeliverydate |  | Datetime |  | |
| Costumerpurchaseordernumber |  | Varchar |  | |
| Isundersupplybackordered |  | bit |  | |
| Pickingcompletedwhen |  | Datetime |  | |
| Backorderid |  | Int |  | |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Name** | **Type of table** | **Nr. Records** | | **Description** |
| Ft\_invoicelines | Fact Table | 456530 | | Tabela de factos representante de cada linha de cada fatura |
| **Target (Data mart)** | | | | **Source (OLTP)** |
| **Column** | **SCD** | **Data type** | **Description** | |
| Invoicelineid |  | Int |  | |
| Invoiceid |  | Int |  | |
| Ordered |  | Int |  | |
| Costumerkey |  | Int |  | |
| Stockitemkey |  | Int |  | |
| Ordered |  | Int |  | |
| Extendedprice |  | Numeric |  | |
| Lineprofit |  | Numeric |  | |
| Invoicedate |  | Datetime |  | |
| Quantity |  | Int |  | |
| Unitprice |  | Numeric |  | |
| Taxrate |  | Numeric |  | |
| Iscreditnote |  | Bit |  | |
| Totaldryitems |  | Int |  | |
| Totalchilleritems |  | Int |  | |
| Returneddeliverydata |  | Datetime |  | |
| Confirmeddeliverytime |  | Datetime |  | |

# Implementação da DataMart

## Jobs

Esta solução apresenta 2 “jobs” e 5 “transformations”, sendo estes 2 “jobs” igualmente importantes para esta solução.

* job\_dims\_ft

Uma imagem com diagrama

Descrição gerada automaticamente Este “job” está a fazer o tratamento de ficheiros para as tabelas de factos, (explicar quais são as tabelas de factos que escolhemos), estas tabelas são as mais importantes para o nosso projeto. Dentro do “job” há também um tratamento de erros que colocamos a não realizar nenhuma operação (importante para o auxílio e deteção de erros).

Figura 2 – Job\_dims\_ft

* job\_dims

Este “job” está a manipular os ficheiros de acordo com as dimensões que escolhemos no projeto.

(explicar quais são as dimensões e fundamentar escolha).

O processo começa por verificar se a tabela que pretendemos aceder existe, caso não exista vamos

aceder ao “Load\_dim\_date” (explicar por alto o que é), que por sua vez acede à base

de dados onde contém a informação que pretendemos alcançar.

Caso a informação seja corretamente carregada executamos um conjunto de transformações para manipular os dados que queremos. Por fim extraímos tudo para um ficheiro novo. Dentro do “job” há também um tratamento de erros, tal como na imagem apresentada anteriormente.

Uma imagem com diagrama

Descrição gerada automaticamente

Figura 3 - Job\_dims

## Transformations

* Dim\_costumers

Explicar

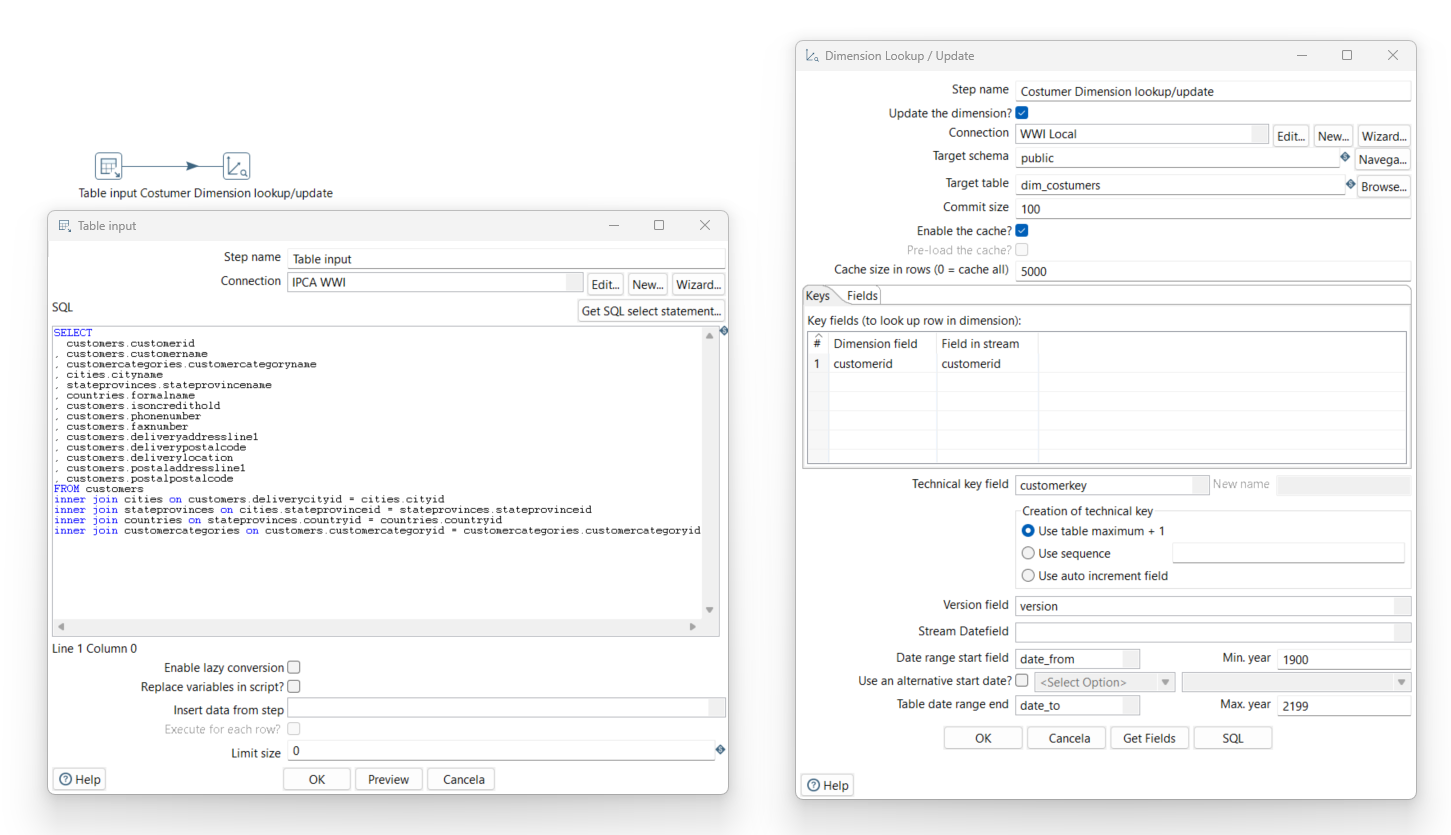


Figura 4 - Transformation Dim\_Costumers

Dim\_costumertransactions

Uma imagem com texto, captura de ecrã, interior, computador portátil

Descrição gerada automaticamenteExplicar

Figura 5 -Transformation Dim\_CostumerTransactions

Dim\_stockitems

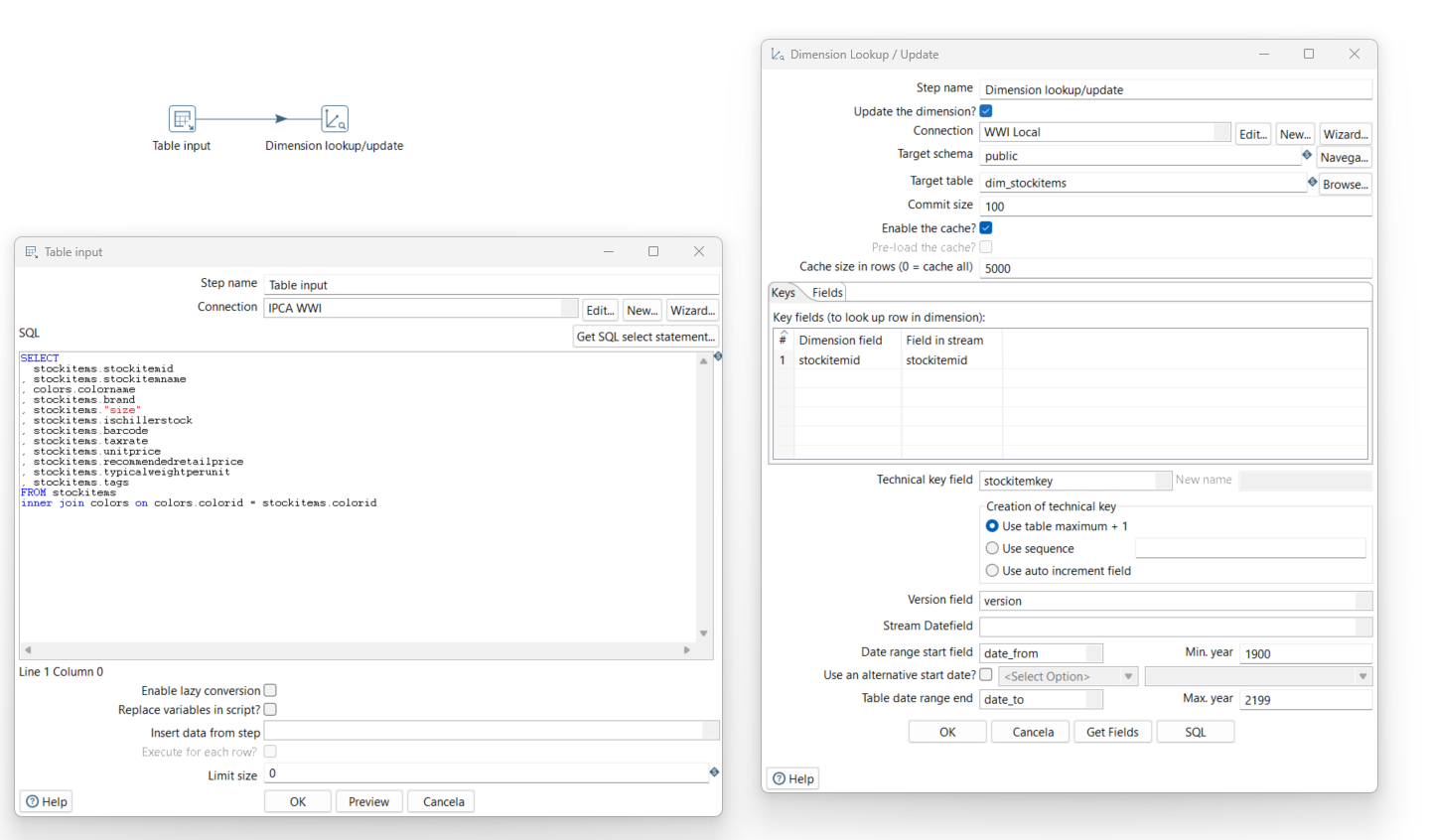
Explicar 

Figura 6 - Transformation Dim\_StockItems

* Ft\_invoicelines

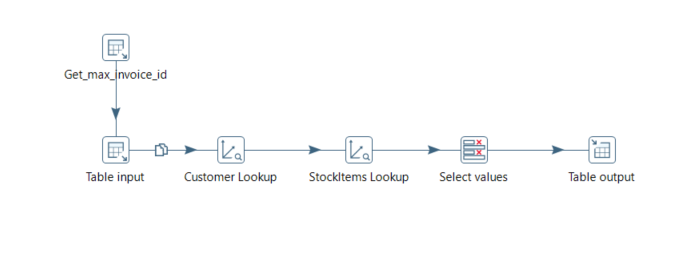
Explicar

Figura 7 - Transformation Ft\_invoiceLines

Uma imagem com texto

Descrição gerada automaticamente

Figura 8 - ft\_invoiceLines get\_max

Uma imagem com texto

Descrição gerada automaticamente

Figura 9 - ft\_invoiceLines input

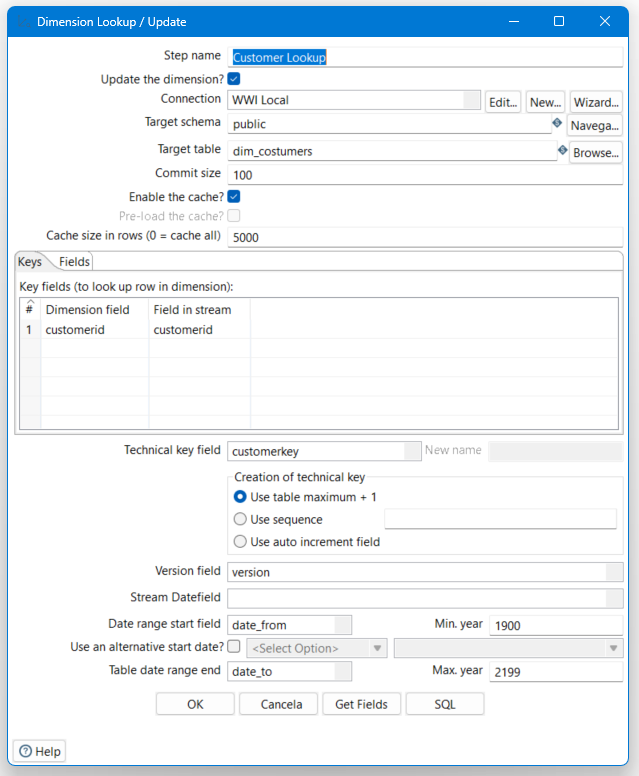


Figura 10 - ft\_invoicelines CostumerLookup

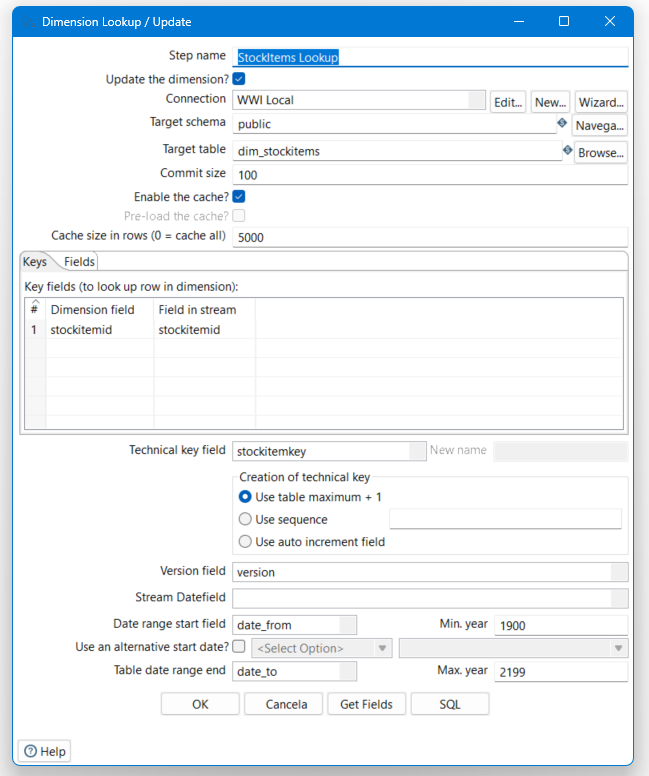


Figura 11 - ft\_invoicelines StockItemsLookup

Uma imagem com texto

Descrição gerada automaticamente

Figura 12 - ft\_invoicelines SelectValues

Uma imagem com texto

Descrição gerada automaticamente

Figura 13 - ft\_invoicelines output

* Ft\_orderlines

Uma imagem com texto

Descrição gerada automaticamente

Figura 14 - ft\_orderlines get\_max

Uma imagem com texto

Descrição gerada automaticamente

Figura 15 - ft\_orderlines input

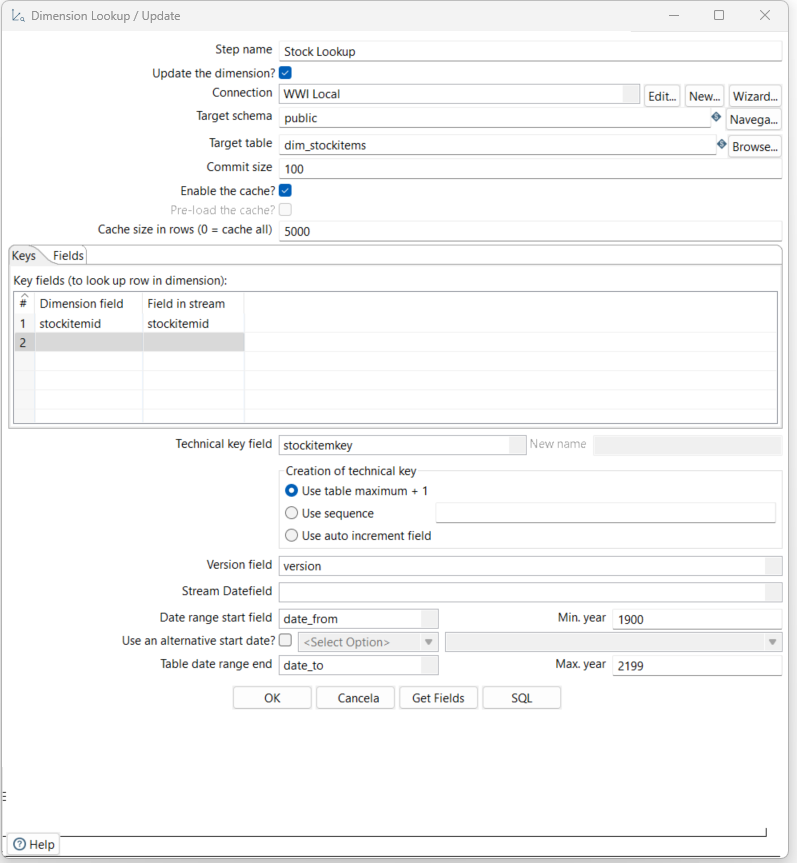
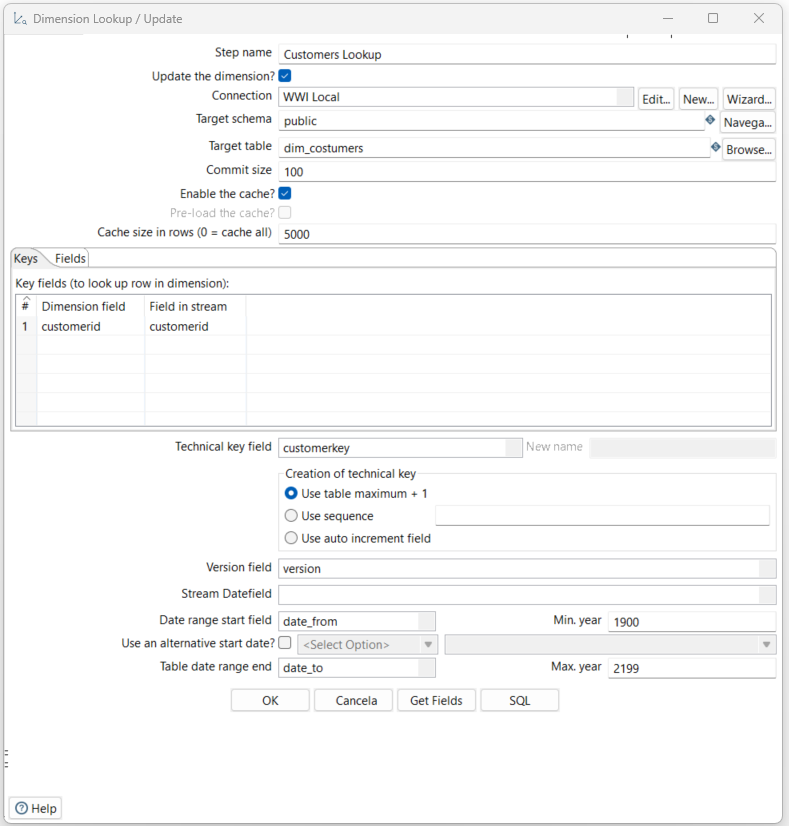


Figura 16 - ft\_orderlines CostumersLookup

Figura 17 - ft\_orderlines StockLookup

# Uma imagem com texto Descrição gerada automaticamenteUma imagem com texto Descrição gerada automaticamente

Figura 19 - ft\_invoices output

Figura 19 - ft\_orderlines SelectValues

# Conclusão

* Revisão crítica do trabalho feito
* Pontos fortes e fracos
* Se possível, o que se mudaria/adicionaria