

# Data Mart Implementation (P01)

DECISION SUPPORT SYSTEMS, 2022-23

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

# Índice

1. Introdução	4
1.1 Visão	4
1.2 Objetivos	4
1.3 Descrição de Processos de Negócio	5
2. Recurso de dados	6
2.1 Data Profiling	<b>Erro! Marcador não definido.</b>
3. Modelação Dimensional	10
3.1 Design do modelo de dados dimensional	11
4. Implementação da DataMart (ETL)	15
4.1 Jobs	15
4.2 Transformations	<b>Erro! Marcador não definido.</b>
5. Conclusão	20

## Índice de imagens

Figura 1 - WWI Modelo BaseDados.....	6
Figura 2 – Job_dims_ft.....	15
Figura 3 - Job_dims .....	16
Figura 4 - Transformation Dim_Costumers.....	17
Figura 5 -Transformation Dim_CostumerTransactions .....	18
Figura 6 - Transformation Dim_StockItems .....	18
Figura 7 - Transformation ft_invoicelines .....	19
Figura 8 - ft_orderlines .....	19

# 1. 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.

## 1.1 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.

## 1.2 Objetivos

Os objetivos deste projeto passam por:

- Realizar a análise de requisitos de negócios e dos recursos 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 um modelo dimensional que atenda aos requisitos de negócios da WWI;
- Projetar e implementar processos de extração, transformação e carregamento dos dados para a Data Mart;
- Executar os processos ETL para garantir a integridade dos dados no Data Mart.

### **1.3 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.

## 2. Recurso de dados

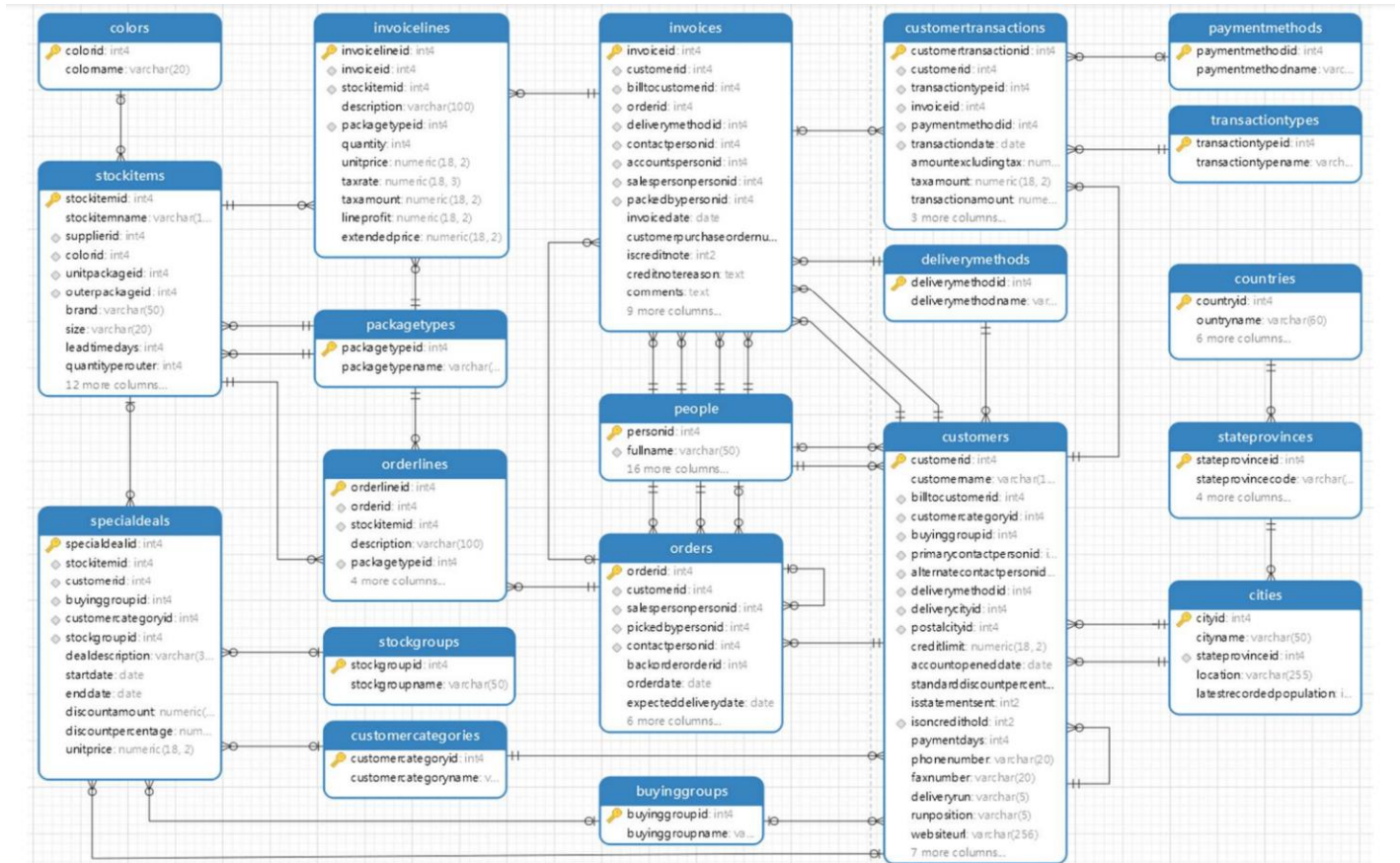


Figura 1 - WWI Modelo BaseDados

Table 1: Summary of important WWI database contents

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

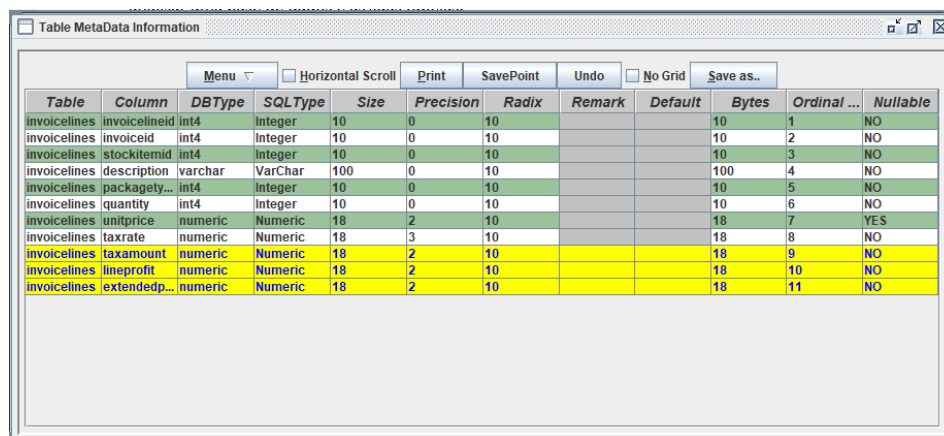
## 2.1 Perfil dos dados

De seguida será feita a análise às tabelas principais da base de dados WWI, fornecida pelo docente.

Para isto foi utilizado o software Open Source Data Quality and Profiling, que tem funcionalidades para analisar uma base de dados e as suas tabelas.

Para este caso irá ser feita a análise das tabelas que formarão as tabelas de facto:

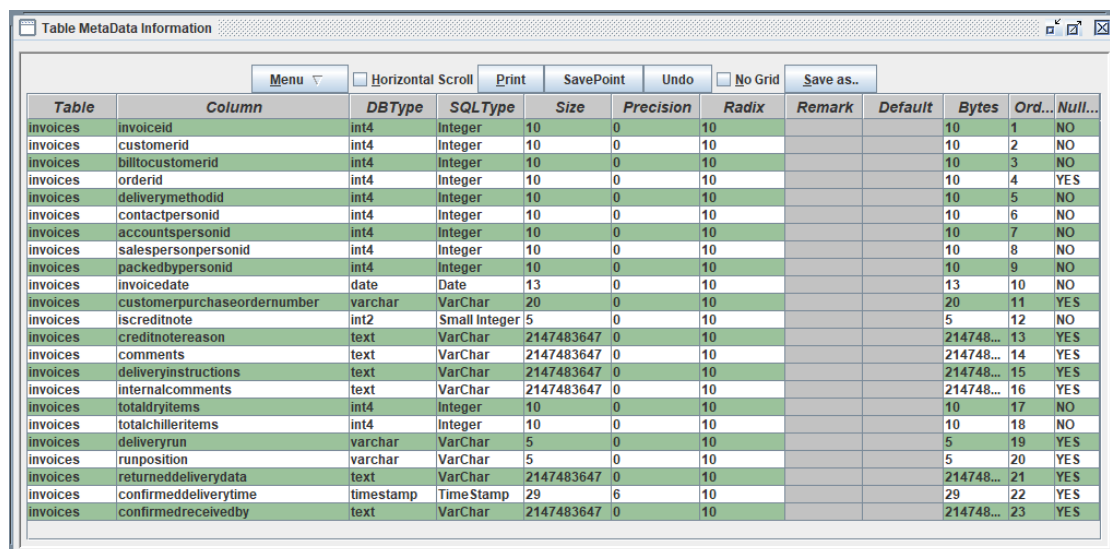
- Invoicelines e invoice
- Orderlines e orders



The screenshot shows the 'Table MetaData Information' window for the 'invoicelines' table. The table has 11 columns: Table, Column, DBType, SQLType, Size, Precision, Radix, Remark, Default, Bytes, Ordinal, and Nullable. The data is as follows:

Table	Column	DBType	SQLType	Size	Precision	Radix	Remark	Default	Bytes	Ordinal	Nullable
invoicelines	invoicelineid	int4	Integer	10	0	10			10	1	NO
invoicelines	invoiceid	int4	Integer	10	0	10			10	2	NO
invoicelines	stockitemid	int4	Integer	10	0	10			10	3	NO
invoicelines	description	varchar	VarChar	100	0	10			100	4	NO
invoicelines	packagety...	int4	Integer	10	0	10			10	5	NO
invoicelines	quantity	int4	Integer	10	0	10			10	6	NO
invoicelines	unitprice	numeric	Numeric	18	2	10			18	7	YES
invoicelines	taxrate	numeric	Numeric	18	3	10			18	8	NO
invoicelines	taxamount	numeric	Numeric	18	2	10			18	9	NO
invoicelines	lineprofit	numeric	Numeric	18	2	10			18	10	NO
invoicelines	extendedp...	numeric	Numeric	18	2	10			18	11	NO

Figura 3 – Tabela de Metadados da invoicelines



The screenshot shows the 'Table MetaData Information' window for the 'invoices' table. The table has 23 columns: Table, Column, DBType, SQLType, Size, Precision, Radix, Remark, Default, Bytes, Ord, and Null... The data is as follows:

Table	Column	DBType	SQLType	Size	Precision	Radix	Remark	Default	Bytes	Ord	Null...
invoices	invoiceid	int4	Integer	10	0	10			10	1	NO
invoices	customerid	int4	Integer	10	0	10			10	2	NO
invoices	billtocustomerid	int4	Integer	10	0	10			10	3	NO
invoices	orderid	int4	Integer	10	0	10			10	4	YES
invoices	deliverymethodid	int4	Integer	10	0	10			10	5	NO
invoices	contactpersonid	int4	Integer	10	0	10			10	6	NO
invoices	accountspersonid	int4	Integer	10	0	10			10	7	NO
invoices	salespersonpersonid	int4	Integer	10	0	10			10	8	NO
invoices	packedbypersonid	int4	Integer	10	0	10			10	9	NO
invoices	invoicedate	date	Date	13	0	10			13	10	NO
invoices	customerpurchaseordernumber	varchar	VarChar	20	0	10			20	11	YES
invoices	iscreditnote	int2	Small Integer	5	0	10			5	12	NO
invoices	creditnotereason	text	VarChar	2147483647	0	10			214748...	13	YES
invoices	comments	text	VarChar	2147483647	0	10			214748...	14	YES
invoices	deliveryinstructions	text	VarChar	2147483647	0	10			214748...	15	YES
invoices	internalcomments	text	VarChar	2147483647	0	10			214748...	16	YES
invoices	totaldryitems	int4	Integer	10	0	10			10	17	NO
invoices	totalchilleritems	int4	Integer	10	0	10			10	18	NO
invoices	deliveryrun	varchar	VarChar	5	0	10			5	19	YES
invoices	runposition	varchar	VarChar	5	0	10			5	20	YES
invoices	returndeliverydata	text	VarChar	2147483647	0	10			214748...	21	YES
invoices	confirmeddeliverytime	timestamp	Time Stamp	29	6	10			29	22	YES
invoices	confirmedreceivedby	text	VarChar	2147483647	0	10			214748...	23	YES

Figura 4 - Tabela de Metadados da invoices

Analisando primeiramente, verifica-se em “invoices” que todos os campos do tipo varchar, são nullable dado a entender que estes são apenas atributos com o objetivo de suporte a quem interage com a base de dados, sendo assim, estes atributos não serão relevantes para a Data Mart. De seguida, é notável que ambas as tabelas trabalham com bastantes ids, tornando a dimensão formada por ambas, numa forte concorrente a ser uma tabela de factos da Data Mart.

Table MetaData Information											
<div> Menu ▾ <input type="checkbox"/> Horizontal Scroll Print SavePoint Undo <input type="checkbox"/> No Grid Save as.. </div>											
Table	Column	DBType	SQLType	Size	Precision	Radix	Remark	Default	Bytes	Ordinal Pos	Nullable
orderlines	orderid	int4	Integer	10	0	10			10	1	NO
orderlines	orderid	int4	Integer	10	0	10			10	2	NO
orderlines	stockitemid	int4	Integer	10	0	10			10	3	NO
orderlines	description	varchar	VarChar	100	0	10			100	4	NO
orderlines	packagety...	int4	Integer	10	0	10			10	5	NO
orderlines	quantity	int4	Integer	10	0	10			10	6	NO
orderlines	unitprice	numeric	Numeric	18	2	10			18	7	YES
orderlines	taxrate	numeric	Numeric	18	3	10			18	8	NO
orderlines	pickedquant...	int4	Integer	10	0	10			10	9	NO

Figura 5 - Tabela de Metadados da orderlines

Table MetaData Information

Menu ▾☐ Horizontal ScrollPrintSavePointUndo☐ No GridSave as..

Table	Column	DBType	SQLType	Size	Precision	Radix	Remark	Default	Bytes	Ordinal...	Nullable
orders	orderid	int4	Integer	10	0	10			10	1	NO
orders	customerid	int4	Integer	10	0	10			10	2	NO
orders	salespersonpersonid	int4	Integer	10	0	10			10	3	NO
orders	pickedbypersonid	int4	Integer	10	0	10			10	4	YES
orders	contactpersonid	int4	Integer	10	0	10			10	5	NO
orders	backorderorderid	int4	Integer	10	0	10			10	6	YES
orders	orderdate	date	Date	13	0	10			13	7	NO
orders	expecteddeliverydate	date	Date	13	0	10			13	8	NO
orders	customerpurchaseordernumber	varchar	VarChar	20	0	10			20	9	YES
orders	isundersupplybackordered	int2	Small Integer	5	0	10			5	10	NO
orders	comments	text	VarChar	2147483647	0	10			2147483647	11	YES
orders	deliveryinstructions	text	VarChar	2147483647	0	10			2147483647	12	YES
orders	internalcomments	text	VarChar	2147483647	0	10			2147483647	13	YES
orders	pickingcompletedwhen	timestamp	Time Stamp	29	6	10			29	14	YES

Figura 6 - Tabela de Metadados da orders

Primeiramente, é possível observar que estas tabelas são semelhantes às tabelas “invoices” e “invoicelines”, trabalhando também com vários Ids e mostrando-se uma concorrente a ser uma tabela de factos.

Pelas informações dadas pela database, sabe-se que é possível fazer encomendas nos quais é preciso fazer um pedido de itens em falta de stock, tendo em conta que nem todas as encomendas necessitarão de fazê-lo, o campo “backorderorderid” é nullable. Os outros campos nullable são semelhantes aos da tabela “invoices”, tendo assim, as mesmas razões de esses atributos serem possivelmente nulos.



Information

/postgres2.ipca.pt/wwi

invoicelines

228265

Show Condition MetaData Summary Data Table Privilege Indexes

Menu

☐ Horizontal Scroll

Print

SavePoint

Undo

☐ No Grid

Save as...

invoicelineid	invoiceid	stockitemid	description	packagetypeid	quantity	unitprice	taxrate	taxamount	lineprofit	extendedprice
1	1	67	Ride on toy sedan car (Brown)	7	10	230	15	345	850	2.645
2	2	50	Developer joke mug - old	7	9	13	15	17,55	76,5	134,55
3	2	10	USB food flash drive - C...	7	9	32	15	43,2	180	331,2
4	3	114	Superhero action jacket...	7	3	30	15	13,5	24	103,5
5	4	206	Permanent marker black	7	96	2,7	15	38,88	96	298,88
6	4	130	Furry gorilla with big eyes	7	5	32	15	24	120	184
7	4	50	Developer joke mug - old	7	2	13	15	3,9	17	28,9
8	5	128	Plush shark slippers (Grey)	7	4	32	15	19,2	96	147,2
9	5	121	Dinosaur battery-powered	7	3	32	15	14,4	70,5	110,4
10	5	155	Large sized bubblewrap...	7	20	24	15	72	220	552
11	6	150	Pack of 12 action figures	9	4	16	15	9,6	42	73,6
12	6	40	Developer joke mug - (blue)	7	6	13	15	11,7	51	89,7
13	6	126	Plush shark slippers (Grey)	7	9	32	15	43,2	216	331,2
14	7	39	Developer joke mug - in...	7	9	13	15	17,55	76,5	134,55
15	7	184	Shipping carton (Brown)	7	75	3,5	15	39,38	120	301,88
16	7	9	USB food flash drive - b...	7	6	32	15	28,8	120	220,8
17	7	132	Furry gorilla with big eyes	7	1	32	15	4,8	24	36,8
18	8	9	USB food flash drive - b...	7	4	32	15	19,2	80	147,2
19	8	32	Developer joke mug - th...	7	9	13	15	17,55	76,5	134,55
20	8	50	Developer joke mug - old	7	1	13	15	1,95	8,5	14,95
21	9	58	RC toy sedan car with r...	7	2	25	15	7,5	25	57,5
22	9	134	Animal with big feet slip...	7	4	32	15	19,2	96	147,2
23	10	38	Developer joke mug - in...	7	5	13	15	9,75	42,5	74,75
24	10	34	Developer joke mug - fu...	7	3	13	15	5,85	25,5	44,85
25	11	45	Developer joke mug - th...	7	1	13	15	1,95	8,5	14,95
26	12	62	RC toy sedan car with r...	7	5	25	15	18,75	62,5	143,75
27	12	44	Developer joke mug - th...	7	8	13	15	15,6	68	119,6
28	13	156	10 mm Double sided bu...	7	10	15	15	22,5	75	172,5

Figura 7 - Tabela invoicelines

//postgres2.ipca.pt/wwi

orderlines

231412

Show Condition

MetaData

Summary Data

Table Privilege

Indexes

Menu

Horizontal Scroll

Print

SavePoint

Undo

No Grid

Save as...

orderlineid	orderid	stockitemid	description	packagetypeid	quantity	unitprice	taxrate	pickedquantity
1	45	164	32 mm Double sided bubble w...	7	50	112	15	50
2	1	67	Ride on toy sedan car (Black) ...	7	10	230	15	10
3	2	50	Developer joke mug - old C de...	7	9	13	15	9
4	46	89	"The Gu" red shirt XML tag t-s...	7	72	18	15	72
5	46	171	32 mm Anti static bubble wra...	7	90	32	15	90
6	2	10	USB food flash drive - chocola...	7	9	32	15	9
7	47	167	10 mm Anti static bubble wra...	7	50	99	15	50
8	47	219	Void fill 400 L bag (White) 400L	7	80	50	15	80
9	3	114	Superhero action jacket (Blue...	7	3	30	15	3
10	47	72	Ride on toy sedan car (Pink) 1...	7	9	230	15	9
11	4	206	Permanent marker black 5m...	7	96	2,7	15	96
12	4	130	Furry gorilla with big eyes slip...	7	5	32	15	5
13	4	50	Developer joke mug - old C de...	7	2	13	15	2
14	5	128	Plush shark slippers (Gray) L	7	4	32	15	4
15	48	100	"The Gu" red shirt XML tag t-s...	7	72	18	15	72
16	48	151	Pack of 12 action figures (mal...	9	8	16	15	8
17	5	121	Dinosaur battery-powered sli...	7	3	32	15	3
18	5	155	Large sized bubblewrap roll 5...	7	20	24	15	20
19	6	150	Pack of 12 action figures (vari...	9	4	16	15	4
20	6	40	Developer joke mug - (hip, hip...	7	6	13	15	6
21	6	126	Plush shark slippers (Gray) S	7	9	32	15	9
22	7	39	Developer joke mug - inherita...	7	9	13	15	9
23	7	184	Shipping carton (Brown) 305x...	7	75	3,5	15	75
24	7	9	USB food flash drive - banana	7	6	32	15	6

Figura 8 - Tabela orderlines

Analisando agora as possíveis tabelas de facto, tendo em conta que “orders” está em várias “orderlines” e “invoices” está em várias “invoicelines”, foi feita uma observação das tabelas “orderlines” e “invoicelines”. Como é possível verificar, temos bastantes dados inseridos nestas tabelas o que faz com que estas tenham um grau de importância maior. Com o grande número de chaves estrangeiras e registos nestas tabelas, podem ser determinadas como tabelas de facto da Data Mart a ser construída

### 3. Modelação Dimensional

Table 2: Data Warehouse Matrix

Processos de negócio \ Dimensões	Dim_customers	Dim_stockitems	Dim_customertransactions	Ft_orderlines	Ft_invoicelines
Top de países que compraram mais o item X	X	X		X	
Top de itens 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

### 3.1 Design do modelo de dados dimensional

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

Nome	Tipo	Nr. Registos	Descrição
dim_costumers	Dimension	664	Dimensão que apresenta os dados pessoais de um cliente
Target (Data Mart)			
Coluna	Tipo de dados	Descrição	
Costumerkey	Int	Chave identificadora do cliente	
Costumerid	Int	Identificador do cliente	
Costumername	Varchar	Nome do cliente	
Cityname	Varchar	Nome da cidade em que o cliente vive	
Stateprovincename	Varchar	Nome do estado em que o cliente vive	
Formalname	Varchar	Nome do país em que o cliente vive	
Costumercategoryname	Varchar	Nome do tipo de categoria do cliente	
Isoncredithold	Bit	Atributo que confirma se o cliente tem um crédito em espera	
Phonenumber	Varchar	Número de telemóvel do cliente	
Faxnumber	Varchar	Número de telefone fixo do cliente	
Deliveryadressline1	Varchar	Morada de entrega de encomendas do cliente	
Deliverypostalcode	Varchar	Código postal de entrega de encomendas do cliente	
Deliverylocation	Varchar	Localização em coordenadas de entrega de encomendas do cliente	
Postaladressline1	Varchar	Morada do cliente	
Postalpostalcode	Varchar	Código postal do cliente	

Nome	Tipo	Nr. Registos	Descrição
dim_stockitems	Dimension	228	Dimensão que apresenta os dados sobre os itens em stock e suas características
Target (Data Mart)			
Coluna	Tipo de dados	Descrição	
Stockitemkey	Int	Chave identificadora do item	
Stockitemid	Int	Identificador do item	
Stockitemname	Varchar	Nome do item	
Colorname	Varchar	Cor do item	
Unitprice	Numeric	Preço unitário do item	
Brand	Varchar	Marca à qual o item pertence	
Size	Varchar	Tamanho do item	
Ischillerstock	Varchar	Atributo que verifica se o item é fresco ou não	
Barcode	Varchar	Número do código de barras do item	
Taxrate	Numeric	Taxa aplicada no item para a sua venda	
Recommendedretailprice	Numeric	Preço recomendado para revenda do item	
Typicalweightperunit	Numeric	Peso por unidade	
Tags	Text	Etiquetas que se relacionam ao item	

Nome	Tipo	Nr. Registos	Descrição
dim_customertransaction	Dimension	26638	Dimensão que apresenta todas as transações que ocorreram no WWI
Target (Data Mart)			
Coluna	Tipo de dados	Descrição	
Customertransactionkey	Int	Chave identificador da transação	
Customertransactionid	Int	Identificador da transação	
Costumerid	Int	Identificador do cliente	
Invoiceid	Varchar	Identificador da fatura	
Transactiontypename	Int	Nome do tipo de transição efetuada	
Paymentmethodname	Varchar	Nome do tipo de pagamento efetuado	
Transactiondate	Datetime	Data em que a transação foi efetuada	
Taxammount	Numeric	Taxa aplicada na transação	
Transactionamount	Numeric	Valor monetário enviado a partir da transação	
Finalizationdate	Datetime	Data de finalização da transação	
Isfinalized	Bit	Atributo que confirma se a transação está terminada ou não	

Nome	Tipo	Nr. Registos	Descrição
Ft_orderlines	Fact Table	1157060	Tabela de factos representante de cada linha de cada encomenda
Target (Data Mart)			
Coluna	Tipo de dados	Descrição	
Orderlineid	Int	Identificador da linha de encomenda	
Orderid	Int	Identificador da encomenda	
Costumerkey	Int	Chave identificadora do cliente	
Stockitemkey	Int	Chave identificadora do item a comprar	
Orderdate	Datetime	Data em que a encomenda foi feita	
Orderprice	Numeric	Preço total da encomenda	
Expecteddeliverydate	Datetime	Data expectável de receção da encomenda por parte do cliente	
Costumerpurchaseordernumber	Varchar	Número de pagamento da encomenda do cliente	
Isundersupplybackordered	bit	Atributos que regista se a encomenda atual tem stock em falta para fazer uma encomenda desse mesmo material	
Pickingcompletedwhen	Datetime	Data em que a obtenção de todos os itens a serem enviados por parte da WWI	
Backorderid	Int	Identificador da encomenda de stock em falta	

Nome	Tipo	Nr. Registos	Descrição
Ft_invoicelines	Fact Table	456530	Tabela de factos representante de cada linha de cada fatura
Target (Data Mart)			
Coluna	Tipo de dados	Descrição	
Invoicelineid	Int	Identificador da linha de fatura	
Invoiceid	Int	Identificador da fatura	
Orderid	Int	Identificador da encomenda	
Costumerkey	Int	Chave identificadora do cliente	
Stockitemkey	Int	Chave identificadora do item a comprar	
Extendedprice	Numeric	Resultado do preço final do produto, calculado a partir de unitprice, taxrate e quantity	
Lineprofit	Numeric	Lucro retirado da linha da fatura	
Invoicedate	Datetime	Data na qual a fatura foi tirada	
Quantity	Int	Quantidade de unidades registadas na fatura	
Unitprice	Numeric	Preço unitário registado na fatura	
Taxrate	Numeric	Taxa aplicada ao comprador por linha de fatura	
Iscreditnote	Bit	Confirma se a fatura é uma nota de crédito (documento emitido para corrigir faturas com erro) e usado para cancelar ou reduzir o valor da fatura anterior	
Totaldryitems	Int	Quantidade total de itens secos	
Totalchilleritems	Int	Quantidade total de itens frescos	
Returneddeliverydata	Datetime	Data de devolução da entrega	
Confirmeddeliverytime	Datetime	Data de confirmação da receção da encomenda	

## 4. Implementação da Data Mart (ETL)

Para implementar a nossa Data Mart foi utilizado o Kettle, ferramenta que permite fazer operações ETL em dados inseridos no mesmo.

Foram criadas transformações neste software de forma a montar as dimensões e tabelas de factos pretendidas para os nossos objetivos, adicionalmente foram implementados alguns jobs de forma a dar apoio ao funcionamento da montagem e atualização da Data Mart.

### 4.1 Jobs

Primeiramente é executado o “**job\_dims\_ft**” que está a fazer inicialmente a execução do job “**job\_dims**”, que será analisado mais abaixo no relatório, depois é feito o carregamento das transformações das tabelas de factos “**ft\_invoicelines**” e “**ft\_orderlines**”, no caso de erro de qualquer um destes processo, o processo ETL é cancelado identificado o erro na consola.

Este é o “**main job**” que faz o trabalho todo de registar as informações na Data Mart.

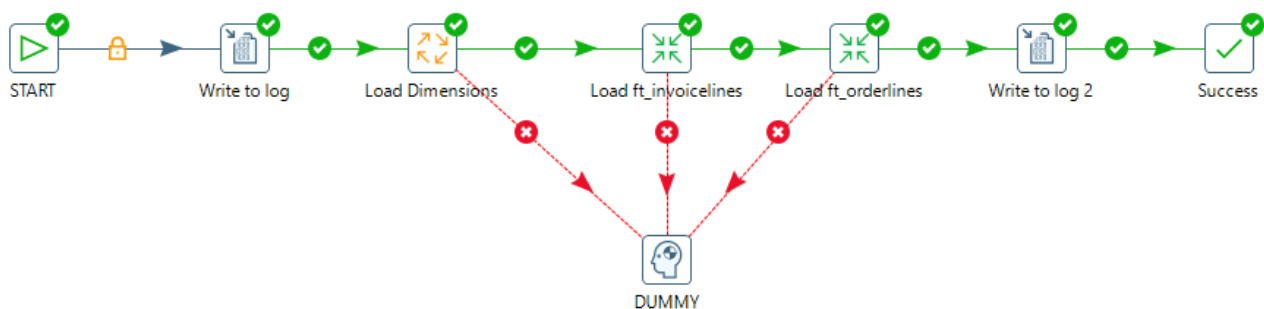


Figura 2 – Job\_dims\_ft

O outro job é “**jobs\_dims**” que tem como função fazer a execução das transformações referentes às dimensões geradas para a nossa Data Mart, no caso de erro o processo ETL é parado e o erro é registado na consola.

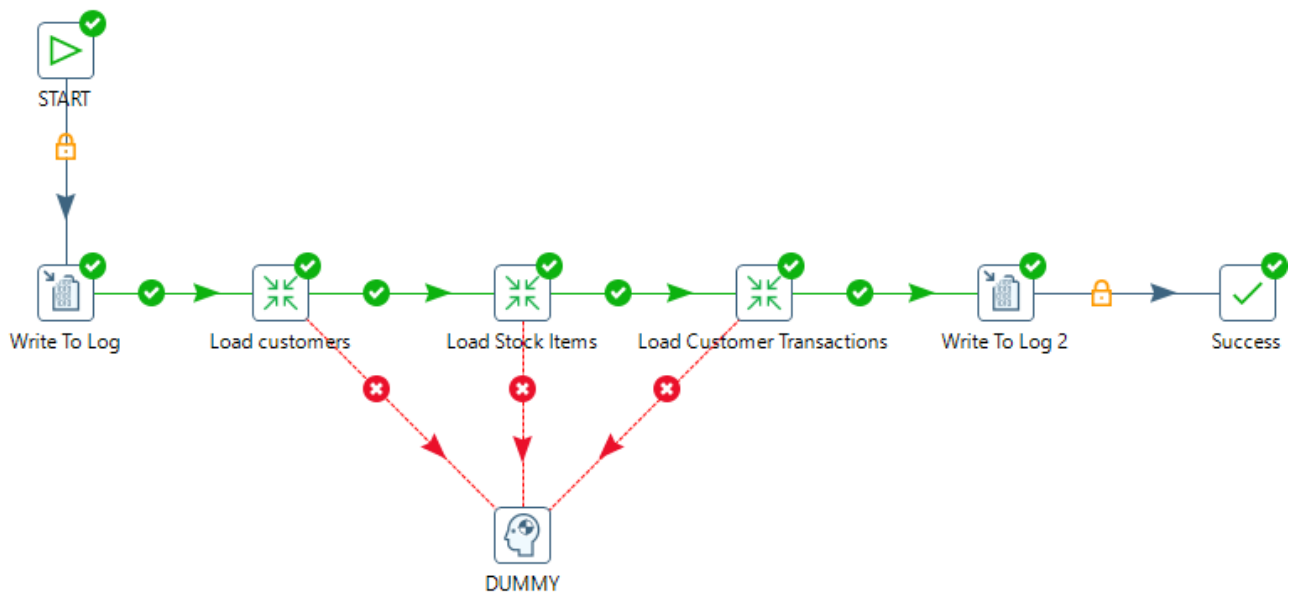


Figura 3 - Job\_dims



## 4.2 Transformations

Nas transformações houve 2 tipos de abordagens: uma para as **dimensões** e outra para as **tabelas de factos**.

Segue-se abaixo um exemplo de uma dimensão, neste caso, “dim\_costumers”:

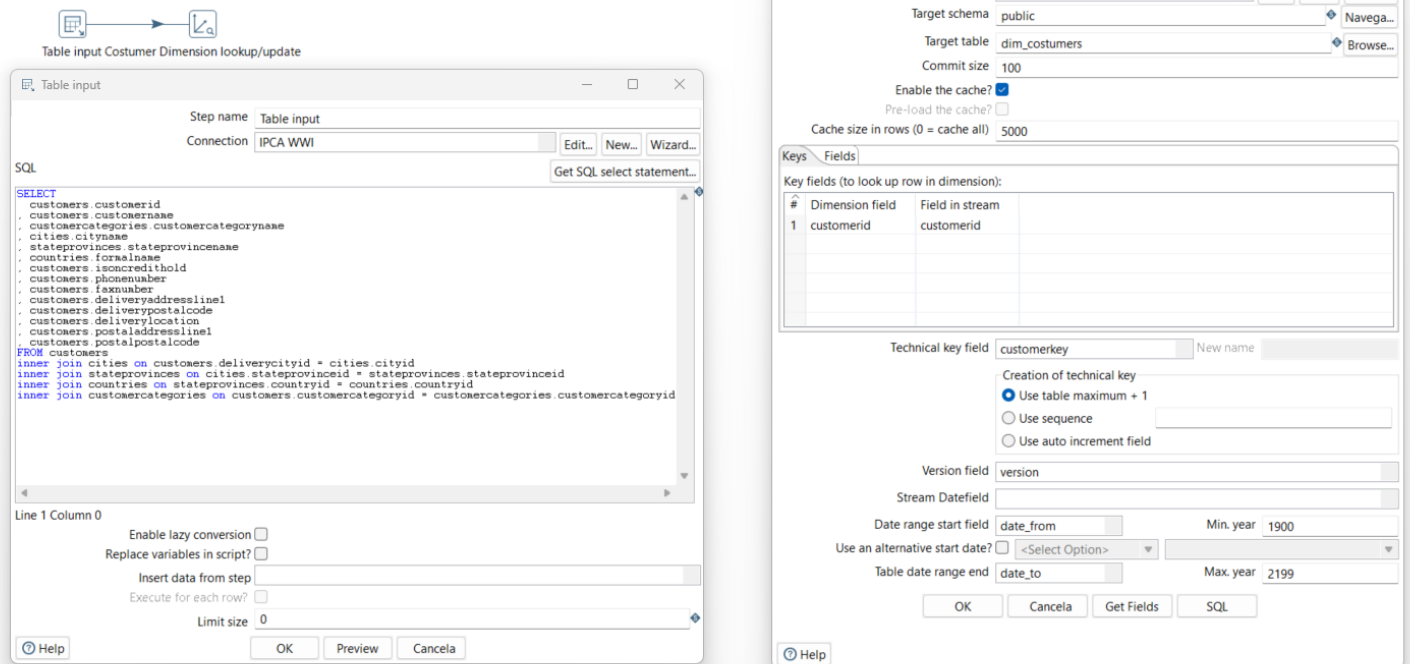


Figura 4 - Transformation Dim\_Costumers

Na construção de dimensões, é preciso apenas estar preocupado em pegar na tabela no formato que já queremos, incluído assim já todas as colunas das tabelas necessárias para construir a dimension e, por fim, aplicar o dimension lookup para fazer a criação da key da nossa dimensão.

Este passo permite também a criação da tabela na nossa Data Mart e o povoamento a partir dos dados obtidos do table input.

Estes passos foram feitos para as restantes dimensões:

- “dim\_stockitems”
- “dim\_costumertransactions”

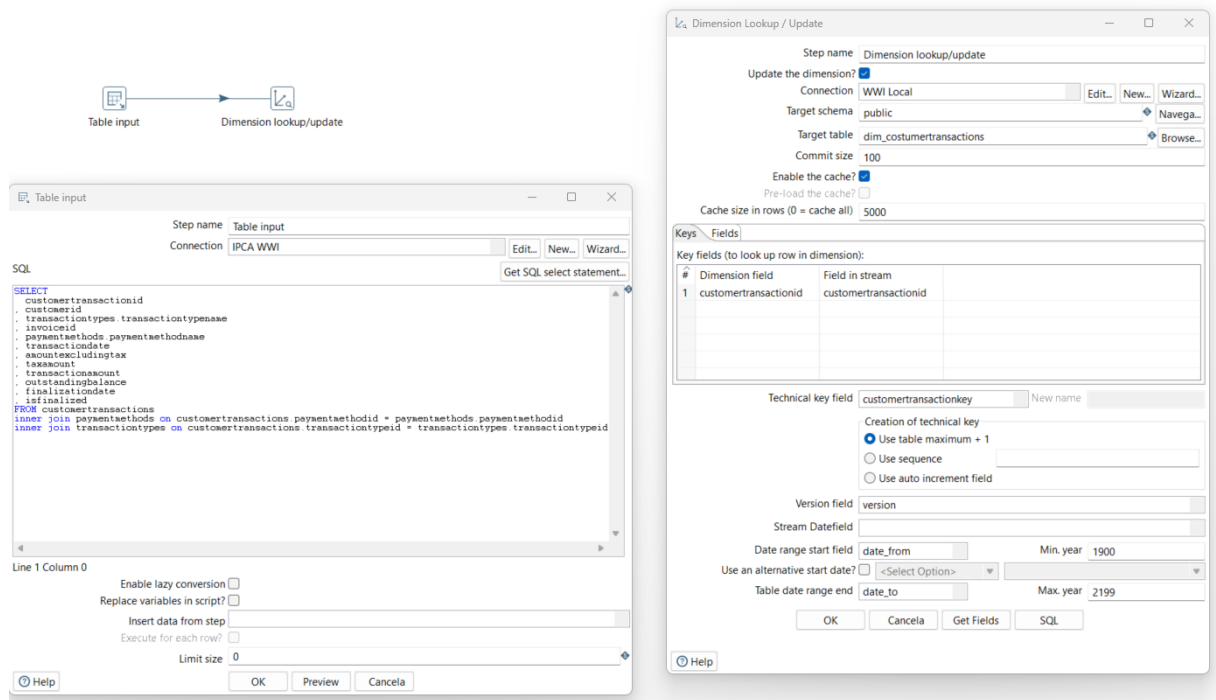


Figura 5 -Transformation Dim\_CostumerTransactions

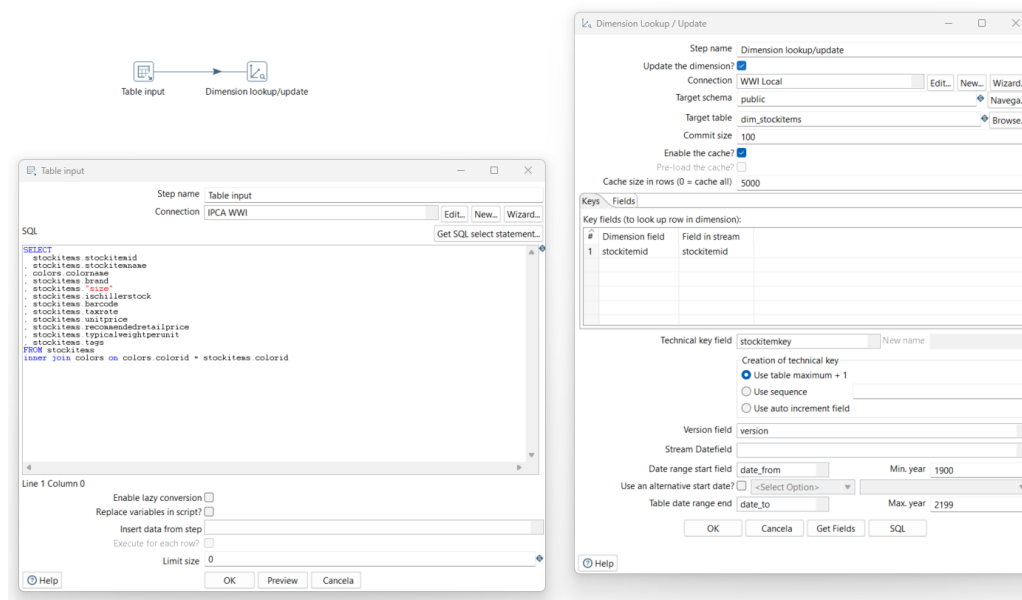


Figura 6 - Transformation Dim\_StockItems

Para o caso das tabelas de factos, a abordagem foi mais exigente, tendo em conta que estamos a trabalhar com uma fact table, foi necessário fazer o lookup para cada atributo que possuía um id de outra dimensão. Segue-se abaixo as transformações de “ft\_invoicelines” e “ft\_orderlines”:

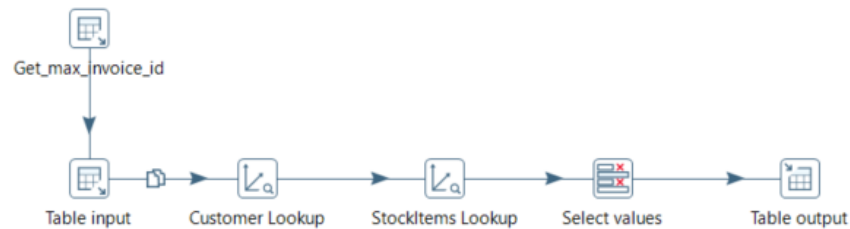


Figura 7 - Transformation ft\_invoicelines

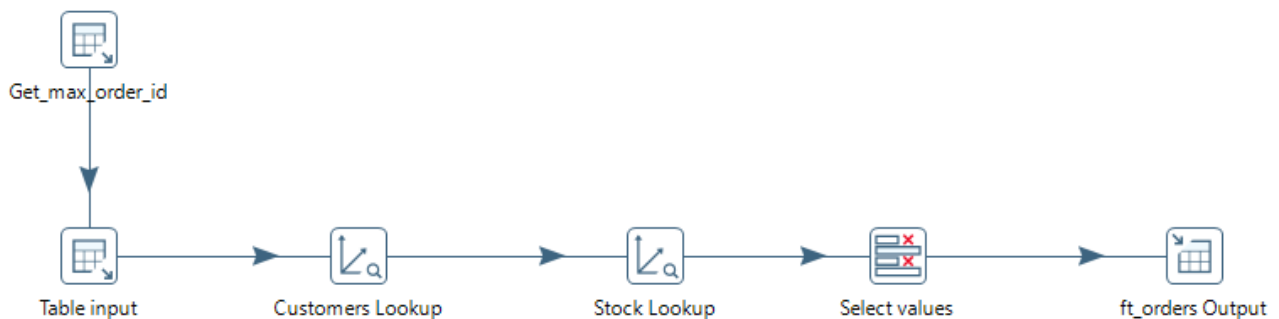


Figura 8 - ft\_orderlines

Inicialmente são obtidos os dados das tabelas pretendidas para a montagem das tabelas de factos. De seguida, dependendo das colunas adicionas, terá de ser feito então um lookup a todas as chaves estrangeiras exceto aos ids principais da própria fact table, por exemplo para “ft\_invoicelines” não foi necessário este passo para “invoiceid” e “invoicelineid”. Por fim são removidas colunas em excesso e é feito o output para a Data Mart da tabela de factos pronta.

## 5. Conclusão

De um ponto de vista geral deu para aplicar com recurso à maioria das ferramentas fornecidas, os objetivos propostos para o desenvolvimento do projeto, permitindo a abordagem de vários conceitos aprendidos e melhorando as nossas aptidões com processos ETL e construção de uma Data Mart.

Com isto foi também possível responder a algumas perguntas pensadas por nós, desenvolvendo uma ideia crítica de informações relevantes para negócio e entender a finalidade de um Data Mart.

No que diz respeito a mudanças, talvez pudesse ser feito um estudo mais detalhado de forma a poder ser respondido mais questões e de uma forma externa a este projeto seria interessante ser explorada a possibilidade de utilizar técnicas de Machine Learning para prever tendências ou identificar oportunidades de negócios para a WWI.