

GlobalTech

[Explicação do domínio](#)

[Problemas que desencadearam a implementação do sistema](#)

[Volume de dados](#)

[Outros registos](#)

[Fonte de dados alternativa](#)

[Processos de negócio](#)

[Encomenda do\(s\) produto\(s\)](#)

[Pagamento da encomenda](#)

[Preparo do envio da encomenda](#)

[Cancelamento da encomenda](#)

[Motivos para a implementação do Data Warehouse](#)

[Schema da base de dados](#)

[Data Profiling \(SSIS\)](#)

[Modelação Dimensional](#)

[Questões Típicas](#)

[FACT_ORDER](#)

[FACT_ORDER_PAYMENT](#)

[FACT_SHIPMENT](#)

[FACT_CANCELTATION](#)

[Matriz de Conformidade](#)

[BPMN](#)

[Data Staging Area](#)

[Data Warehouse](#)

[Outras informações](#)

Explicação do domínio

A GlobalTech é uma empresa que se insere no domínio de retalho e tem como principais atividades a revenda de componentes informáticos. As encomendas podem ser realizadas pelos clientes que entram em contacto com os responsáveis de vendas ou então os responsáveis a contactar diretamente os clientes conforme as necessidades previstas. A encomenda deve de seguida ser registada e/ou atualizada no sistema da empresa ficando inicialmente com o estado PENDING.

Possíveis sequências de estado da encomenda:

- PENDING → SHIPPED
- PENDING → CANCELED
- PENDING → SHIPPED → CANCELED

As encomendas apenas são enviadas (SHIPPED) se o cliente não tiver ultrapassado o limite de crédito e se existir stock no armazém mais próximo, na falta de stock naquele armazém, é necessário efetuar transferência de stock entre armazéns. Caso o valor em falta de pagamento for superior ao limite de crédito, a encomenda continua suspensa (PENDING).

O limite de pagamento da encomenda será 3 meses depois da data em que a mesma for realizada.

É também importante mencionar que a empresa define o limite de crédito de cada cliente consoante o número de encomendas realizadas e o nível de confiança do mesmo.

Para além disso as encomendas canceladas podem ter sido recebidas ou não. Essa mesma informação é proveniente do excel fornecido pela transportadora. Se a encomenda foi enviada e não foi cancelada, então está predefinido que esta foi recebida.

Problemas que desencadearam a implementação do sistema

Com o passar do tempo e devido ao crescimento da empresa foram aparecendo certas adversidades/dúvidas cujas respostas são até hoje incompletas/desconhecidas. Dúvidas essas como:

- ▼ Clientes insatisfeitos com a demora de algumas encomendas
 - ▼ Apenas acontece com certa transportadora?
 - ▼ Altura do ano terá influência?

- ▼ Falta de comunicação e atenção para com o cliente?
- ▼ Encomendas não entregues
 - ▼ Qual a transportadora?
 - ▼ Volume das encomendas?
 - ▼ Destino e região?
- ▼ Dificuldade em dar resposta a encomendas
 - ▼ Muita afluência em momentos especiais (natal, ano novo, black friday)?
 - ▼ Stock insuficiente?

A empresa procura então dar resposta a estas questões investindo na implementação de um sistema analítico capaz de apresentar e auxiliar o negócio com diversos dashboards e diferentes perspetivas dos dados existentes.

Volume de dados

A empresa GlobalTech em média recebe 1000 encomendas por dia. No entanto, é também armazenada informação relativa aos produtos encomendados (quantidade, unit_price) e ao pagamentos realizados. No caso de termos uma encomenda de dois produtos e com um único pagamento, teremos sempre que adicionar os seguintes registos:

order_items

item_id	order_id	product_id	quantity	unit_price
1	5	3	25	400
2	5	4	100	20

orders

order_id	customer_id	salesman_id	status	order_date	shipped_date	canceled_date	limit_pay
5	4	3	PENDING	20-12-2013	NULL	NULL	20-03-2014

Quando a encomenda passa para o estado SHIPPED o registo passa a ficar:

5	4	3	SHIPPED	20-12-2013	27-12-2013	NULL	20-03-2014
---	---	---	---------	------------	------------	------	------------

payments

payment_id	order_id	method_id	payment_date	value
1	5	1	25-12-2013	12000

Outros registos

- Funcionários e Clientes
- Produtos e stock em armazém (localização)

Fonte de dados alternativa

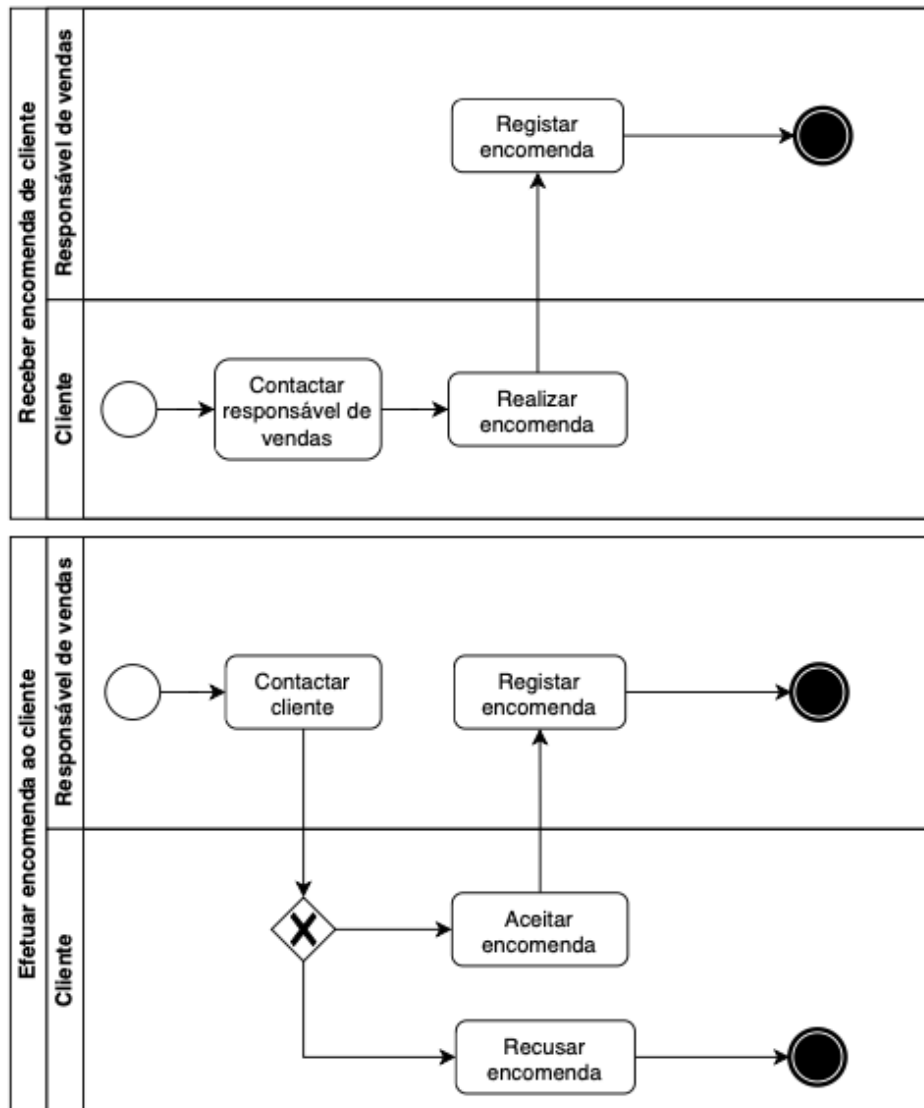
A empresa transportadora fornece mensalmente um ficheiro excel com as encomendas que foram entregues ou não entregues.

Exemplo do documento em formato .xlsx ou .csv

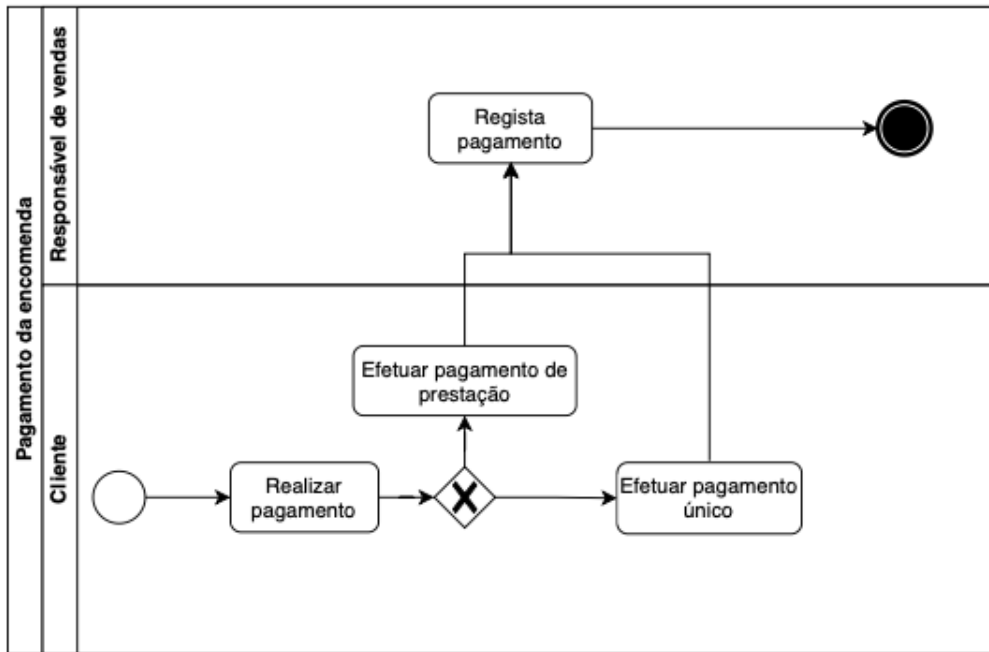
order_id	was_received	delivery_date
5	Yes	25-12-2013
2	No	28-12-2013
3	Yes	30-12-2013

Processos de negócio

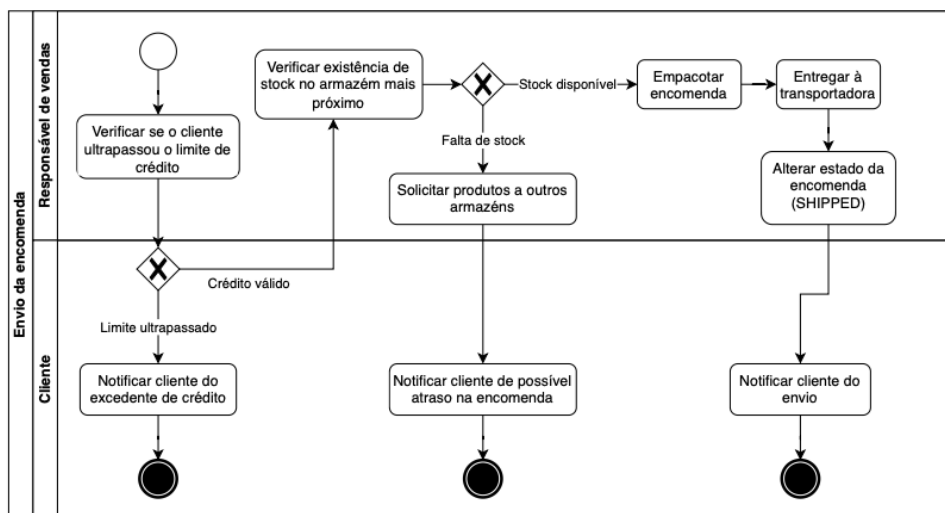
Encomenda do(s) produto(s)



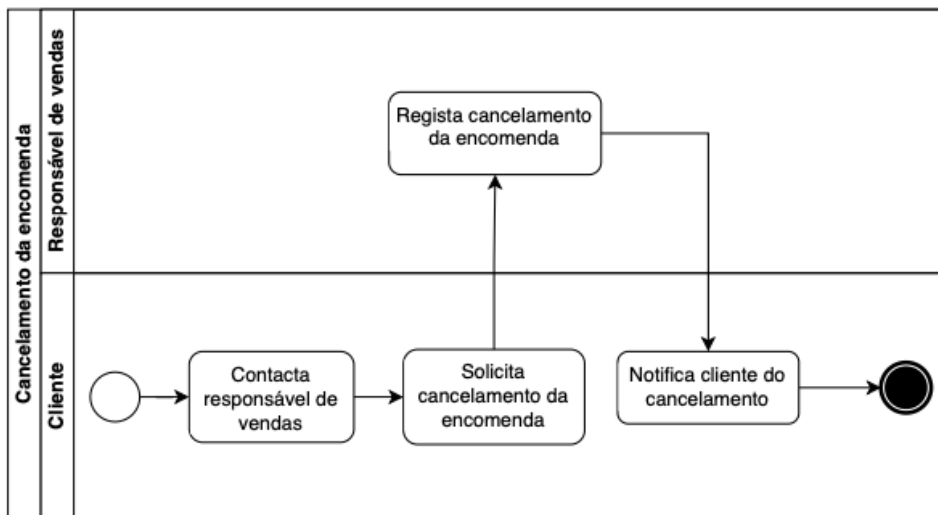
Pagamento da encomenda



Preparo do envio da encomenda



Cancelamento da encomenda

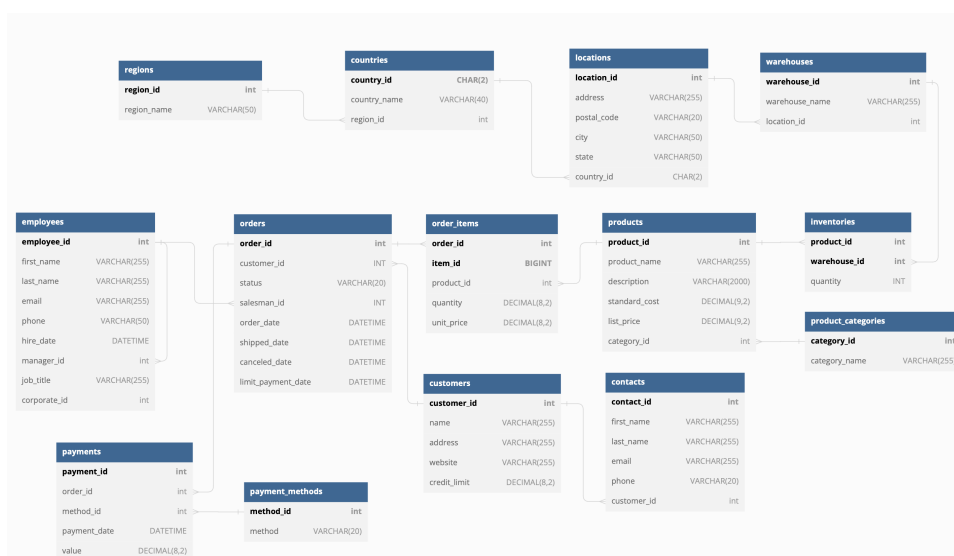


Motivos para a implementação do Data Warehouse

- A existência de um **repositório central** que representa uma única **realidade da organização** (união das diversas fontes de dados) com vantagens a nível de desempenho
- Flexibilidade para os gestores na forma como acedem à informação
- Definir **estratégias/rankings** para avaliação de responsáveis de venda
- Desenvolver **campanhas** promocionais orientadas a produtos/clientes/regões/países
- Identificar por região/país/localização/armazém o número volume de vendas por produto/categoria de forma a permitir uma melhor **gestão de stocks**
- Melhorar a **qualidade** do serviço de vendas
- Identificar **oportunidades** de expansão de negócio

Schema da base de dados

<https://dbdiagram.io/d/637c0bf6c9abfc61117447ac>



Data Profiling (SSIS)

O Data Profiling é uma ferramenta bastante importante e essencial que deve ser realizada previamente sobre um conjunto de dados. É uma atividade de Qualidade de Dados que é caracterizada pelo uso de técnicas analíticas sobre os dados para assim conhecer-mos o conteúdo, estrutura e qualidade dos mesmos.

Depois de realizada a análise com recurso à ferramenta **SQL Server Integration Services**, foi possível concluir algumas métricas importantes para o negócio:

- Apenas existem valores **NULL** nas tabelas: orders, locations e employees.
 - Podemos concluir que **74%** das encomendas foram enviadas, **16%** canceladas e **10%** estão pendentes

Column	Null Count	Null Percentage
corporate_id	0	0.0000 %
email	0	0.0000 %
employee_id	0	0.0000 %
first_name	0	0.0000 %
hire_date	0	0.0000 %
job_title	0	0.0000 %
last_name	0	0.0000 %
manager_id	1	0.9346 %
phone	0	0.0000 %

Column	Null Count	Null Percentage
address	0	0.0000 %
city	0	0.0000 %
country_id	0	0.0000 %
location_id	0	0.0000 %
postal_code	1	4.3478 %
state	6	26.0870 %

Column	Null Count	Null Percentage
canceled_date	89	84.7619 %
customer_id	0	0.0000 %
limit_paymen...	0	0.0000 %
order_date	0	0.0000 %
order_id	0	0.0000 %
salesman_id	35	33.3333 %
shipped_date	28	26.6667 %
status	0	0.0000 %

- Última encomenda realizada a **01/11/2017** e enviada a **06/11/2017**

Column	Minimum	Maximum
canceled_date	5/5/2015 12:00:00 AM	8/22/2017 12:00:00 AM
customer_id	1	70
limit_paymen...	9/21/2013 12:00:00 AM	2/1/2018 12:00:00 AM
order_date	6/21/2013 12:00:00 AM	11/1/2017 12:00:00 AM
order_id	1	105
salesman_id	54	64
shipped_date	6/23/2013 12:00:00 AM	11/6/2017 12:00:00 AM

- Mínimo, Máximo** e a **média** do limite de crédito atribuído aos clientes

Column	Minimum	Maximum	Mean	Standard Dev...
credit_limit	1250.00	62500.00	23683.385...	18755.36679...
customer_id	1	319	160	92.08691546...

- O **preço** e **custo** do produto mais caro e mais barato

Column	Minimum	Maximum	Mean	Standard Dev...
category_id	1	5	3.2986111...	1.662828334...
list_price	15.55	8867.99	903.24138...	887.3865952...
product_id	1	288	144.5	83.13793758...
standard_cost	12.63	7123.66	727.61538...	720.0403718...

- A coluna com **maior** número de caracteres na definição da descrição do produto, e a frequência do tamanho

Column	Minimum Length	Maximum Length	Ignore Leading Spaces	Ignore Trailing Spaces
description	28	69	<input type="checkbox"/>	<input checked="" type="checkbox"/>
product_name	7	43	<input type="checkbox"/>	<input checked="" type="checkbox"/>

Length	Count	Percentage
28	9	3.1250 %
29	28	9.7222 %
30	33	11.4583 %
45	6	2.0833 %
46	7	2.4306 %
47	2	0.6944 %
48	5	1.7361 %
49	13	4.5139 %
50	12	4.1667 %
51	3	1.0417 %
52	12	4.1667 %
53	23	7.9861 %
54	4	1.3889 %
55	17	5.9028 %
56	13	4.5139 %
57	21	7.2917 %
58	5	1.7361 %
59	6	2.0833 %
60	4	1.3889 %
61	16	5.5556 %
62	31	10.7639 %
63	15	5.2083 %
65	1	0.3472 %
66	1	0.3472 %
69	1	0.3472 %

- **Primeira** e **última** data de contratação de funcionários

Column	Minimum	Maximum	Mean	Standard Dev...
corporate_id	1	107	54	30.88689042...
employee_id	1	107	54	30.88689042...
hire_date	1/3/2016 12:00:00 AM	12/24/2016 12:00:00 AM		
manager_id	1	106	25.764150...	20.21934058...

Modelação Dimensional

Seguindo o método dos 4 passos abordado nas aulas teóricas para desenvolver os esquemas dimensionais obtivemos os seguintes resultados.

Questões Típicas

Questões que os gestores da empresa querem ver respondidas. Estas questões são fundamentais para a identificação das medidas associadas a cada evento e consequentemente a cada tabela de factos.

- Número de encomendas canceladas por dia de semana, mês
- Número de encomendas canceladas por país, região, categoria, armazém
- Percentagem de encomendas entregues por mês
- Top 10 clientes que pagam em menos prestações e no menor tempo
- Top 10 funcionários com maior faturação por mês, ano

FACT_ORDER

WHAT → Product

WHERE → Warehouse

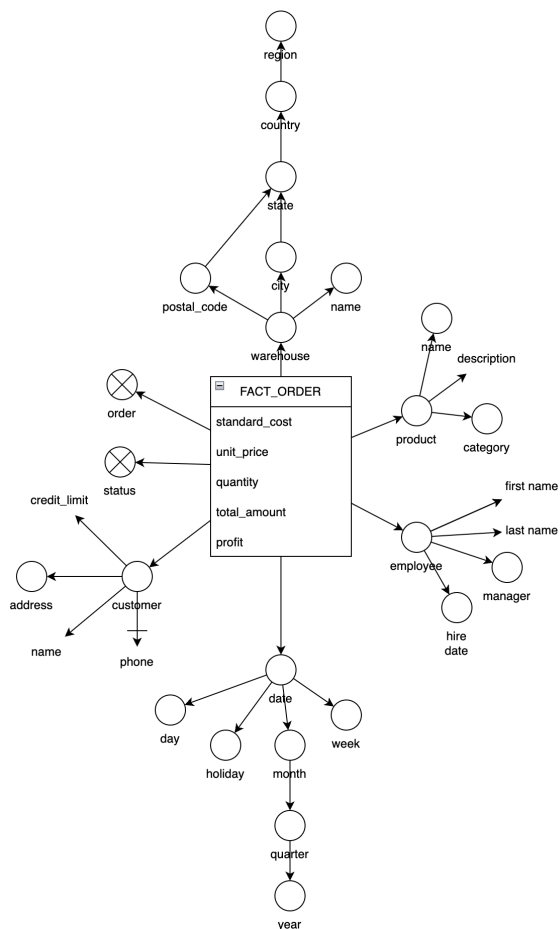
WHEN → Date

WHO → Customer, Employee

1. **Processo de negócio:** Encomenda de Produto
2. **Grão:** Quantidade encomendada de um determinado produto por um funcionário para um cliente num dado dia.
3. **Dimensões:** Product, Warehouse, Date, Customer, Employee
4. **Medidas:**

Atributo	Descrição	Tipo	Funções de agregação	Exemplo
standard_cost	Preço de custo do produto	Aditivo	SUM	450.36
unit_price	Preço unitário de venda	Aditivo	SUM	65.40
quantity	Quantidade do produto enviada	Aditivo	SUM	20
total_amount	Valor da encomenda	Aditivo	SUM	3000
profit	Lucro gerado pela encomenda	Aditivo	SUM	500





FACT_ORDER_PAYMENT

WHAT → Order

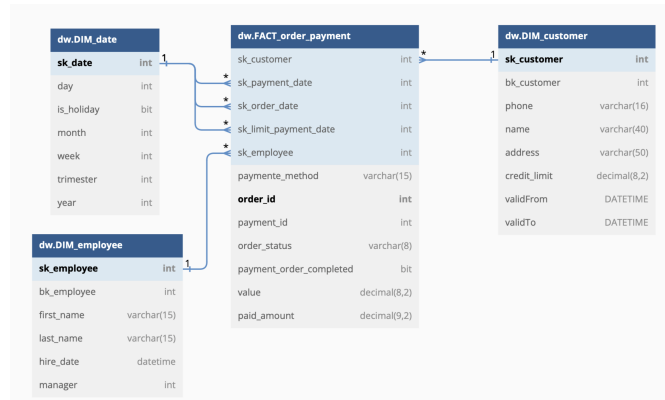
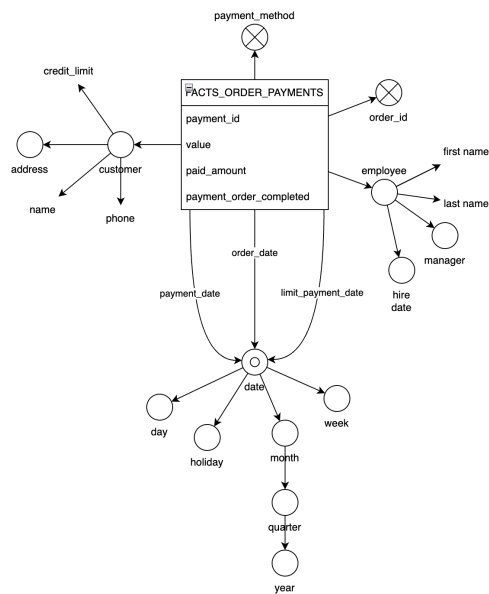
WHEN → Date

WHO → Customer, Employee

HOW → Payment Method

1. **Processo de negócio:** Pagamento de encomenda
2. **Grão:** Pagamento de uma prestação relativa a uma encomenda.
3. **Dimensões:** Order, Date, Customer, Employee, Payment Method
4. **Medidas:**

Atributo	Descrição	Tipo	Funções de agregação	Exemplo
value	Valor do pagamento	Aditivo	SUM	500
paid_amount	Valor pago até ao momento	Não aditivo		250



FACT_SHIPPMENT

Tabela Acumulativa devido ao campo **was_received** que é atualizado ao longo do tempo

WHAT → Order, Product

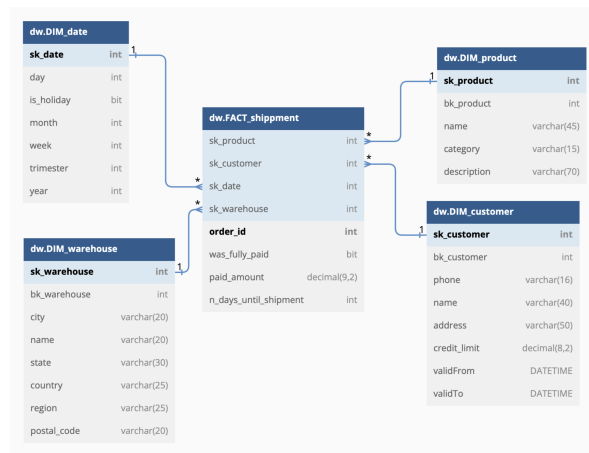
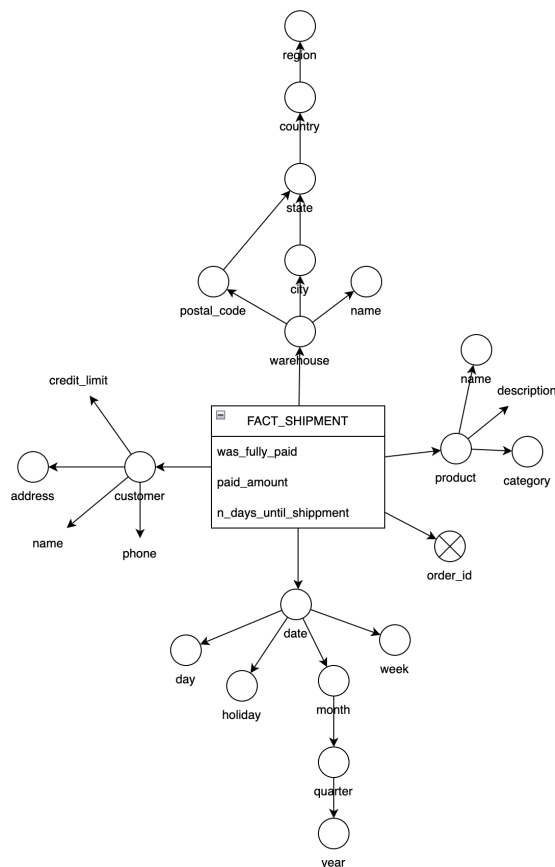
WHERE → Warehouse

WHEN → Date

WHO → Customer

1. **Processo de negócio:** Preparo do envio da encomenda
2. **Grão:** Envio de uma encomenda de um determinado produto para um cliente.
3. **Dimensões:** Order, Product, Warehouse, Date, Customer
4. **Medidas:**

Atributo	Descrição	Tipo	Funções de agregação	Exemplo
paid_amount	Quantidade paga até ao momento	Aditivo	SUM	1500
n_days_until_shippment	Número de dias até envio/Tempo de demora de envio	Não aditivo		23



FACT_CANCELATION

WHAT → Order, Product

WHERE → Warehouse

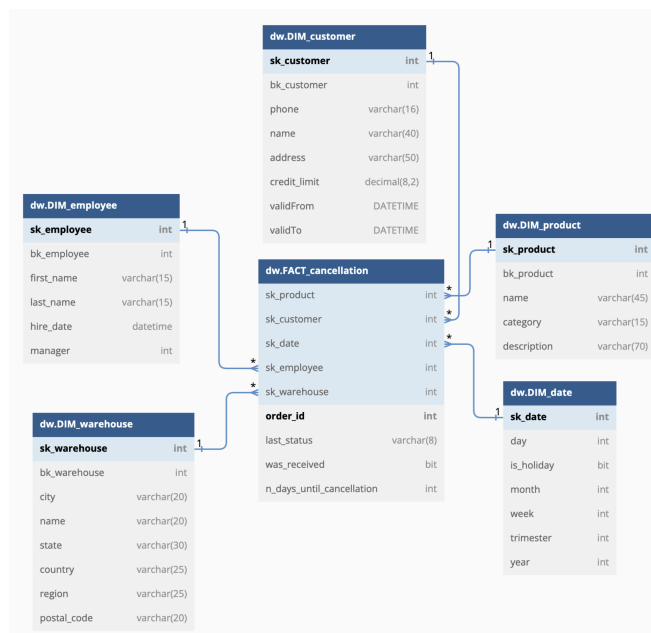
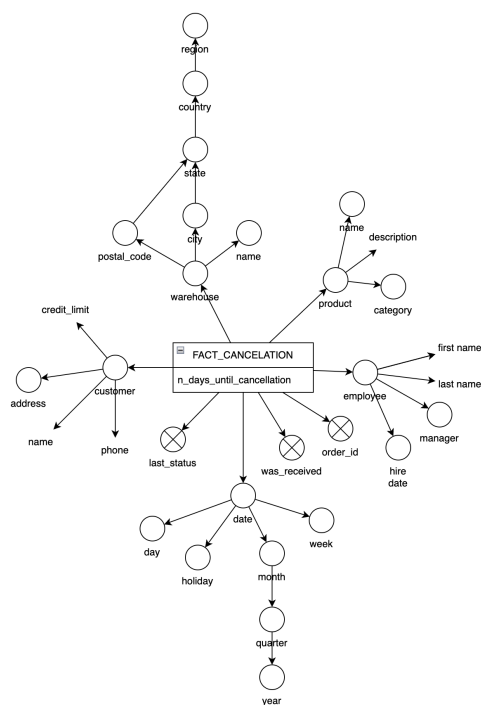
WHEN → Date

WHO → Customer, Employee

É importante mencionar que as encomendas canceladas podem ter sido recebidas ou não. Essa mesma informação é proveniente do excel fornecido pela transportadora. Se a encomenda foi enviada e não foi cancelada, então está predefinido que esta foi recebida.

1. **Processo de negócio:** Cancelamento da encomenda
2. **Grão:** Encomenda cancelada por um cliente
3. **Dimensões:** Order, Product, Warehouse, Customer, Employee, Date
4. **Medidas:**

Atributo	Descrição	Tipo	Funções de agregação	Exemplo
n_days_until_cancellation	Número de dias até a encomenda ser cancelada	Não aditivo		230



Matriz de Conformidade

A matriz de conformidade mostra como as tabelas de factos se intersectam. Isto permite que cada tabela de factos seja implementada individualmente, com a confiança de que, à medida que cada uma for implementada, funcionará em conjunto com as que já foram implementadas.

Process	Customer	Employee	Warehouse	Date	Product
Encomenda do(s) produto(s)	X	X	X	X	X
Pagamento da encomenda	X	X		X	
Envio da encomenda	X		X	X	X
Cancelamento da encomenda	X	X	X	X	X

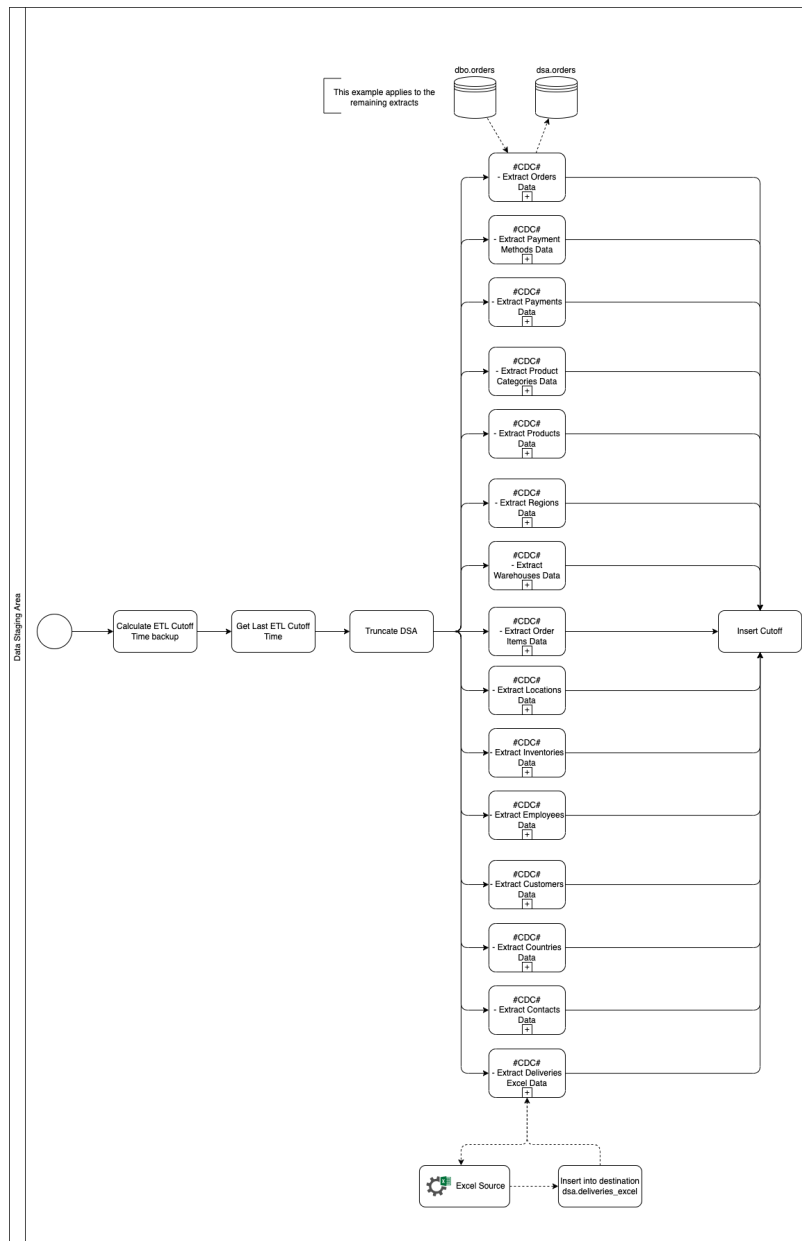
BPMN

Data Staging Area

Data Staging Area é uma cópia da base de dados original que será utilizada para o povoamento futuro da Data Warehouse.

Todas as tabelas do DSA vão buscar a informação à própria tabela do DBO, existindo apenas uma ligação de uma tabela para outra tabela.

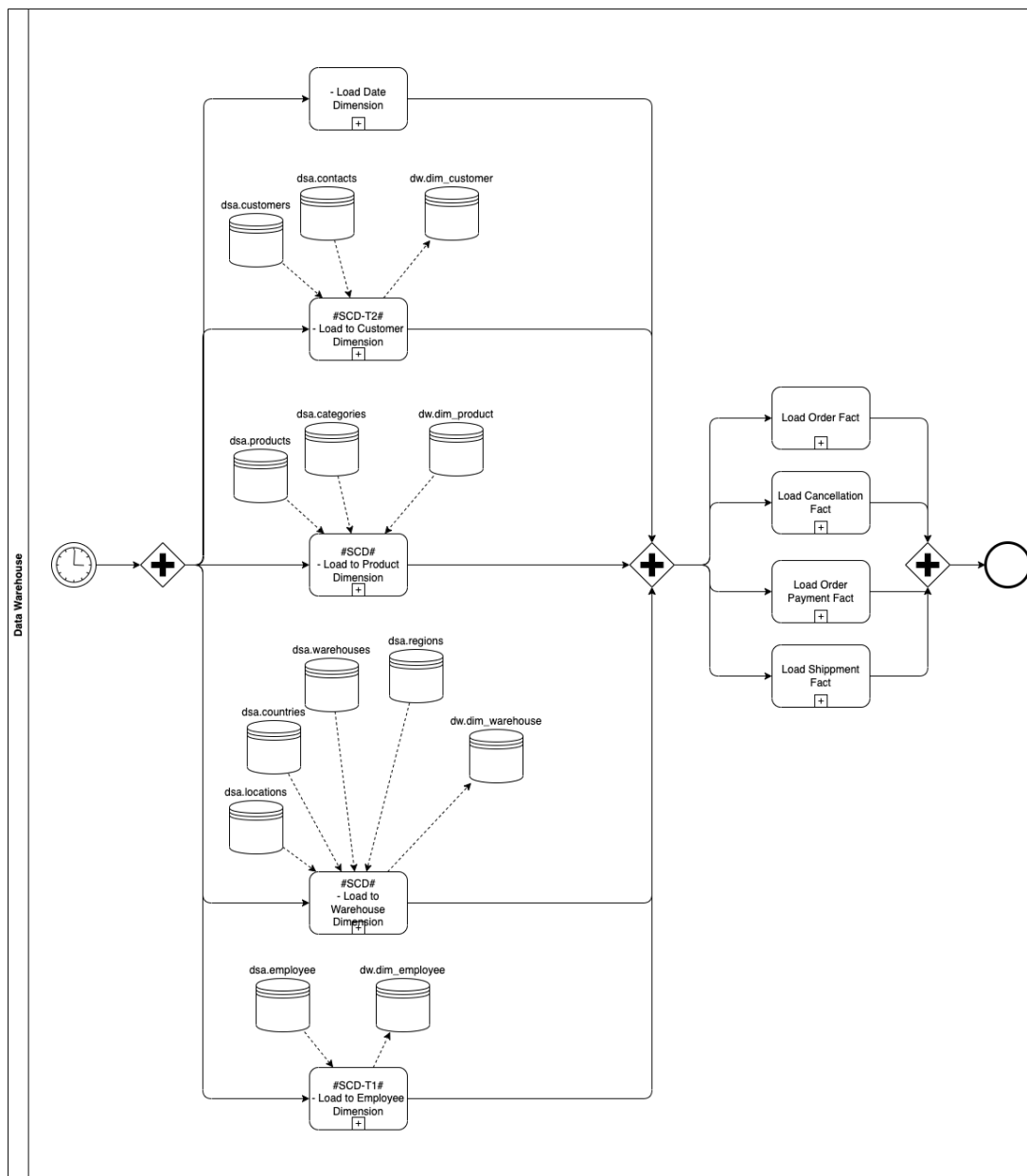
No caso do Deliveries Excel é utilizado uma tarefa de extração de dados (excel) disponibilizada pelo SSIS que depois é carregado numa tabela com exatamente as mesmas colunas.



Data Warehouse

Primeiramente é efetuado o povoamento das dimensões com recurso à tarefa **SCD (Slowly Changing Dimension)** dependendo se a dimensão é sem variação, com variação ou de variação com história. Neste processo, podem ser utilizadas várias tabelas da DSA. É importante mencionar que esta ferramenta apenas adiciona às dimensões os dados que não existem na mesma.

De seguida é realizado o povoamento das tabelas de factos e o processo dá-se como terminado.



Outras informações

Documentação:

<https://s3-us-west-2.amazonaws.com/secure.notion-static.com/450989ff-a42c-4581-b6d7-cf861b18ff42/documentacao.xlsx>

GitHub do projeto:

https://github.com/joaobraganca555/GlobalTech_ETL

Links para visualização dos schemas:

- Schema DB

<https://dbdiagram.io/d/637c0bf6c9abfc61117447ac>

- Schema DW

<https://dbdiagram.io/d/63a1133699cb1f3b55a2610c>