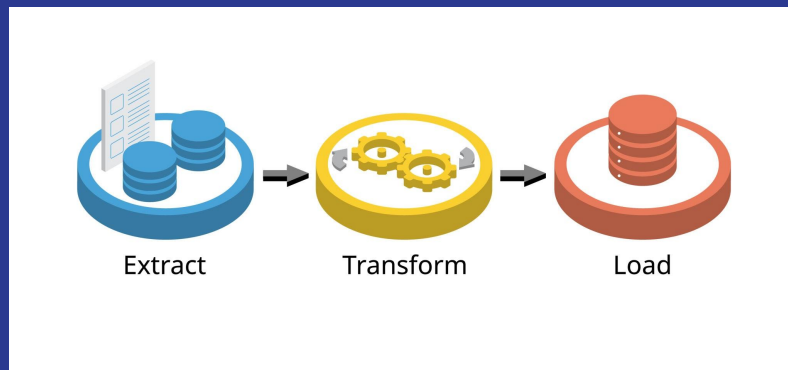


# Projeto 5 - Rota de ETL

- Criação de um sistema abrangente que permita extrair dados de diversas fontes;
- Transformação dos dados para necessidades específicas da Super Store;
- Carregamento em Data Warehouse;



## Processamento e Preparação da base de dados

- Importação base de dados bruta (CSV) - BigQuery;
- Verificação de nulos e duplicados;



BigQuery

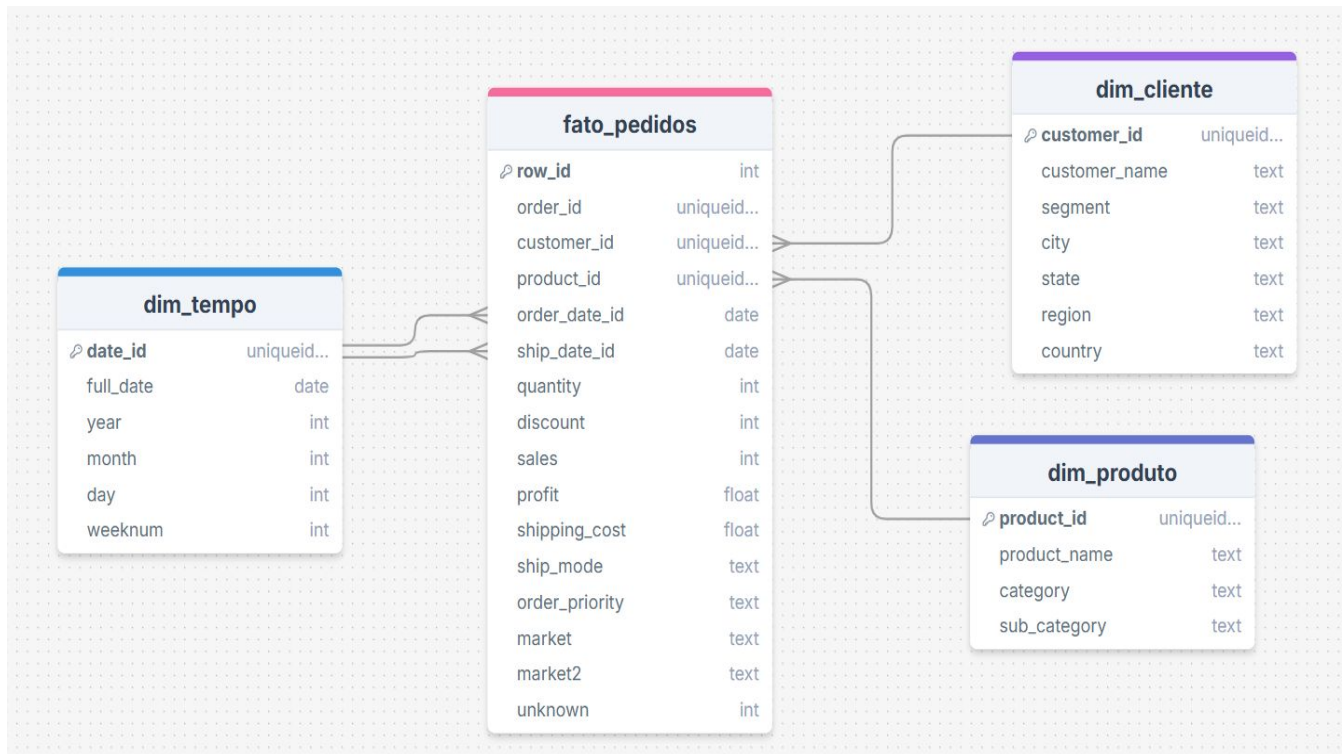


Google Sheets



drawSQL

## Estruturação da base de dados - esquema estrela



**Tabela fato pedidos** -  
tabela central com  
dados quantitativos /  
transacionais;

**Tabelas dimensões** -  
tabelas com descrições  
e categorias;

## chaves primárias (PK)

### chaves estrangeiras (FK)

dim_tempo	
🔑 date_id	uniqueid...
full_date	date
year	int
month	int
day	int
weeknum	int

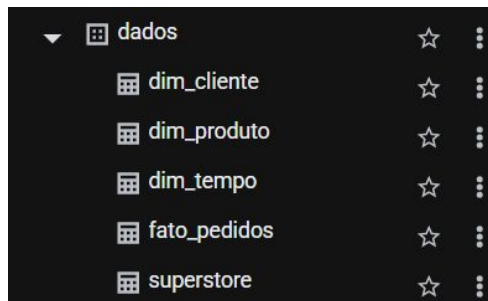
fato_pedidos	
🔑 row_id	int
order_id	uniqueid...
customer_id	uniqueid...
product_id	uniqueid...
order_date_id	date
ship_date_id	date
quantity	int
discount	int
sales	int
profit	float
shipping_cost	float
ship_mode	text
order_priority	text
market	text
market2	text
unknown	int

dim_cliente	
🔑 customer_id	uniqueid...
customer_name	text
segment	text
city	text
state	text
region	text
country	text

dim_produto	
🔑 product_id	uniqueid...
product_name	text
category	text
sub_category	text



## Criação das tabelas no BigQuery



```
CREATE OR REPLACE TABLE
`projetoetl-466019.dados.fato_pedidos` AS
SELECT
  row_id,
  order_id,
  customer_id,
  product_id,
  order_date AS order_date_id,
  ship_date AS ship_date_id,
  quantity,
  discount,
  sales,
  profit,
  shipping_cost,
  ship_mode,
  order_priority,
  market,
  market2,
  unknown
FROM `projetoetl-466019.dados.superstore`
```

```
--CRIACAO DE TABELA DIM_CLIENTE
CREATE OR REPLACE TABLE
`projetoetl-466019.dados.dim_cliente` AS
SELECT DISTINCT
  customer_id,
  customer_name,
  segment,
  city,
  state,
  region,
  country
FROM `projetoetl-466019.dados.superstore`
```

## Pipeline Estruturado

1. Recebimento do arquivo CSV - disponibilizado em frequência diária, ou de forma manual, conforme alguma demanda específica;
2. Carga inicial da tabela “bruta” superstore, mantendo a estrutura original da fonte;
3. Limpeza e transformação dos dados - correção de possíveis inconsistências, conversão de formatos (data, por exemplo) e criação de colunas auxiliares e chaves (como date\_id);
4. Atualização das tabelas de dimensões, na seguinte ordem: dim\_cliente, dim\_produto, dim\_tempo;
5. Atualização da tabela fato\_pedidos - essa tabela armazena registros transacionais (vendas, quantidade, lucro) e se referem as chaves das tabelas dimensão;

**A atualização é realizada de forma incremental, utilizando apenas dados novos ou modificados, o que evita recriar as tabelas novamente a cada execução.**



# Obrigada!

