

Data Warehouse para Plataforma SaaS

Neste projeto, não vamos apenas "limpar dados", vamos **estruturar um armazém de dados**. Imagine que uma empresa de Software (SaaS) tem milhões de registros de vendas, assinaturas e logs. Se tentarmos gerar relatórios diretamente no banco de produção, o sistema trava. O nosso objetivo é criar um **Data Warehouse (DW)** otimizado para análise.

1. 🎯 Objetivos de Aprendizagem

- **Diferença entre OLTP e OLAP:** Entender por que o banco da aplicação é diferente do banco da análise.
- **Modelagem Dimensional:** Aprender a criar tabelas **Fato** e tabelas **Dimensão**.
- **Esquema Snowflake (Floco de Neve):** Implementar uma estrutura de dados normalizada para BI.
- **SQL de Performance:** Criar Views e tabelas agregadas para dashboards rápidos.

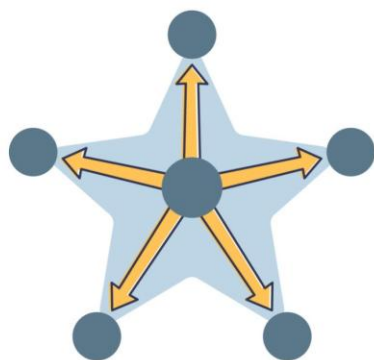
2. 🛠️ Stack Tecnológica

- **PostgreSQL:** O nosso servidor de Data Warehouse.
 - **Python (Pandas/SQLAlchemy):** Para simular a carga de dados (ETL Dimensional).
 - **SQL Avançado:** Para consultas de alta complexidade.
-

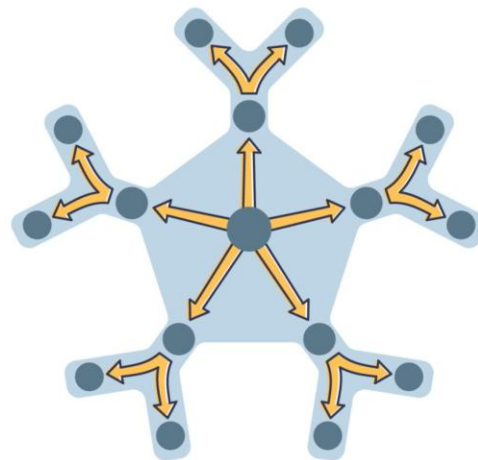
3. 🗺️ Roadmap do Projeto

Fase 1: O Modelo Dimensional (Conceitos)

Antes de codificar, precisamos desenhar. Vamos definir o que são os processos de negócio e como eles se tornam tabelas.



STAR SCHEMA



SNOWFLAKE SCHEMA

Getty Images

- **Tabelas Dimensão (O "Quem", "Onde", "Quando"):** DimCliente, DimProduto, DimTempo.
- **Tabelas Fato (O "Quanto"):** FatoVendas, FatoAssinaturas (contém métricas e chaves estrangeiras).

Fase 2: Implementação do Schema (SQL DDL)

Vamos criar o banco de dados no PostgreSQL usando scripts SQL puros para definir chaves primárias, estrangeiras e restrições de integridade.

Fase 3: ETL Dimensional (Python)

Este é um desafio técnico: Como inserir dados na tabela Fato garantindo que os IDs correspondam às tabelas Dimensão? Vamos usar Python para coordenar essa carga.

Fase 4: Consultas Analíticas de Elite

Aqui você aprenderá a usar **Window Functions (RANK, OVER)** e **CTEs** para responder perguntas como:

- "Qual é o crescimento mensal de receita (MoM) por categoria de produto?"
- "Quem são os Top 10% clientes que mais geram receita em cada estado?"