

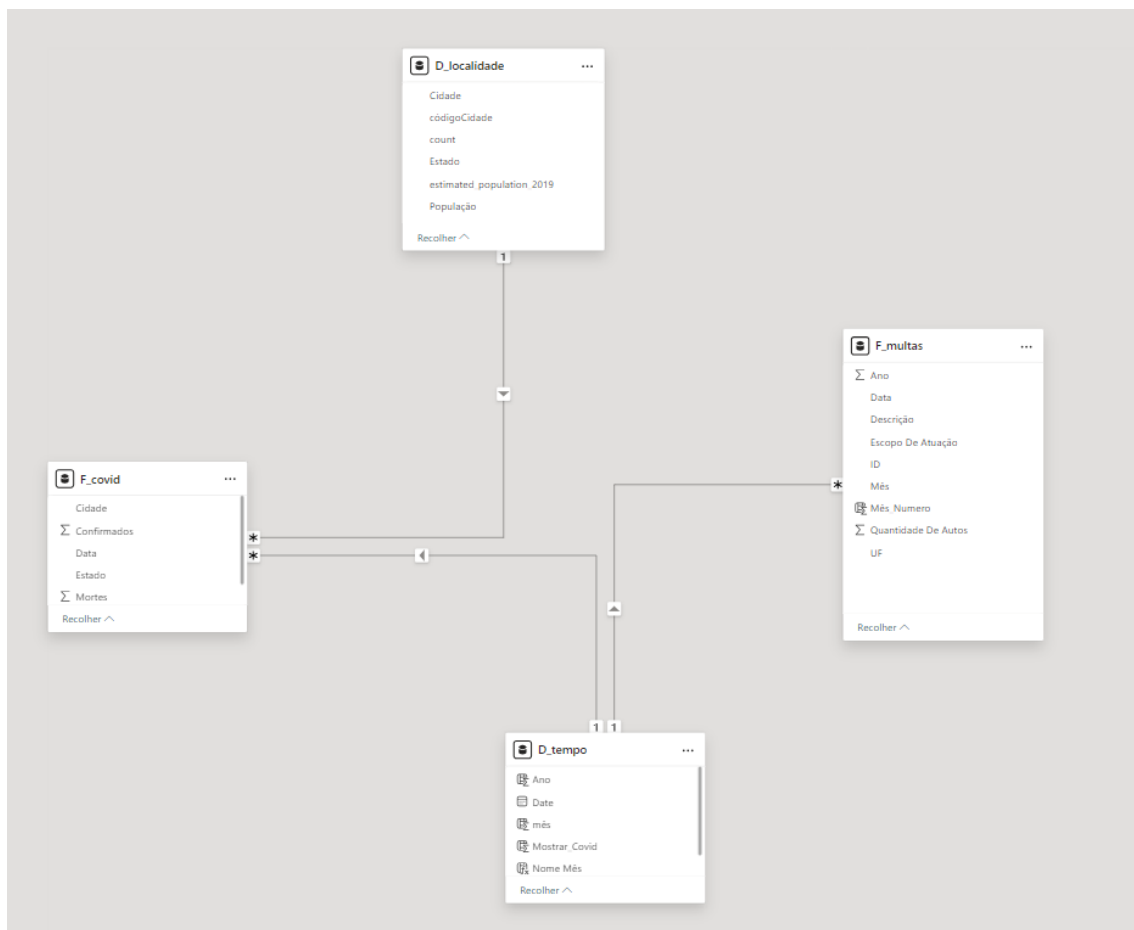
Processo de ETL e Consultas SQL no Contexto do Modelo Star Schema

Contexto

O objetivo deste processo de ETL foi preparar e organizar dados para análises detalhadas no Power BI. Utilizei o modelo de dados do tipo star schema, onde temos tabelas de fatos e dimensões. As consultas SQL foram projetadas para transformar e agregar os dados, facilitando a análise no contexto do esquema estrela.

Modelo de Dados

- D_localidade: Contém informações sobre cidades e estados, incluindo populações.
- D_tempo: Armazena informações temporais como datas, meses e anos.
- F_covid: Registra dados de COVID-19, incluindo casos confirmados e mortes.
- F_multas: Registra dados de multas, incluindo descrições de infrações e quantidades de autos.



Descrição das Consultas SQL

Consulta 1

```
SELECT
    date as Data,
    CASE
        WHEN city_ibge_code = 12 THEN 'Acre'
        WHEN city_ibge_code = 27 THEN 'Alagoas'
        WHEN city IS NULL THEN 'Sem Nome'
        WHEN city = 'Acrelândia' THEN 'Acrelândia'
        WHEN city = 'Brasília' THEN 'Brasília'
        WHEN city = 'Epitaciolândia' THEN 'Epitaciolândia'
        WHEN city = 'Feijó' THEN 'Feijó'
        WHEN city = 'Jordão' THEN 'Jordão'
        WHEN city = 'Maceió' THEN 'Maceió'
        WHEN city = 'Mâncio Lima' THEN 'Mâncio Lima'
        WHEN city = 'Plácido de Castro' THEN 'Plácido de Castro'
        WHEN city = 'Tarauacá' THEN 'Tarauacá'
        WHEN city = 'Porto Real do Colégio' THEN 'Porto Real do Colégio'
        ELSE city
    END AS Cidade,
    state as Estado,
    new_confirmed as Confirmados,
    new_deaths as Mortes
FROM
    DADOS_COVID;
```

O que foi feito:

- Esta consulta formata os nomes das cidades para garantir consistência e corrige possíveis erros de digitação ou substitui valores nulos por descrições padrão.
- Seleciona os dados de casos confirmados e mortes de COVID-19, agrupando por data e estado.

Consulta 2

SELECT

CASE

WHEN state = 'AC' AND city_ibge_code = 12 THEN 'Acre'

WHEN state = 'AL' AND city_ibge_code = 27 THEN 'Alagoas'

WHEN city IS NULL THEN 'Sem Nome'

WHEN city = 'Acrelândia' THEN 'Acrelândia'

WHEN city = 'Brasília' THEN 'Brasília'

WHEN city = 'Epitaciolândia' THEN 'Epitaciolândia'

WHEN city = 'Feijó' THEN 'Feijó'

WHEN city = 'Jordão' THEN 'Jordão'

WHEN city = 'Porto Real do Colégio' THEN 'Porto Real do Colégio'

WHEN city = 'Maceió' THEN 'Maceió'

WHEN city = 'Mâncio Lima' THEN 'Mâncio Lima'

WHEN city = 'Plácido de Castro' THEN 'Plácido de Castro'

WHEN city = 'Tarauacá' THEN 'Tarauacá'

ELSE city

END AS Cidade,

state as Estado,

city_ibge_code as códigoCidade,

estimated_population as População,

estimated_population_2019,

COUNT(*) AS count

FROM

DADOS_COVID

GROUP BY

CASE

WHEN state = 'AC' AND city_ibge_code = 12 THEN 'Acre'

WHEN state = 'AL' AND city_ibge_code = 27 THEN 'Alagoas'

WHEN city IS NULL THEN 'Sem Nome'

WHEN city = 'Acrelândia' THEN 'Acrelândia'

```

WHEN city = 'Brasília' THEN 'Brasília'

WHEN city = 'Epitaciolândia' THEN 'Epitaciolândia'

WHEN city = 'Feijó' THEN 'Feijó'

WHEN city = 'Jordão' THEN 'Jordão'

WHEN city = 'Porto Real do Colégio' THEN 'Porto Real do Colégio'

WHEN city = 'Maceió' THEN 'Maceió'

WHEN city = 'Mâncio Lima' THEN 'Mâncio Lima'

WHEN city = 'Plácido de Castro' THEN 'Plácido de Castro'

WHEN city = 'Tarauacá' THEN 'Tarauacá'

ELSE city

END,

state,

city_ibge_code,

estimated_population,

estimated_population_2019

```

O que foi feito:

- Esta consulta agrega os dados por cidade e estado, calculando a população estimada e a contagem de registros.
- Formata os nomes das cidades e estados para garantir consistência.
- Agrupa os dados e fornece um resumo populacional e a incidência de casos de COVID-19.

Contextualização no Modelo Star Schema

Estas consultas foram elaboradas para corresponder ao modelo star schema, onde as tabelas de dimensões (D_localidade e D_tempo) fornecem atributos descritivos que contextualizam os dados nas tabelas de fatos (F_covid e F_multas).

- Tabela D_localidade: Armazena informações detalhadas sobre cidades e estados, como população e código IBGE, fornecendo contexto geográfico aos dados de fatos.
- Tabela D_tempo: Contém informações temporais, como datas, meses e anos, permitindo análises temporais detalhadas.
- Tabela F_covid: Registra dados de COVID-19, incluindo casos confirmados e mortes, associados a datas e localizações.
- Tabela F_multas: Registra dados de multas, incluindo descrições de infrações e quantidades de autos, associados a datas e localizações.

Resumo

Este processo de ETL foi implementado para transformar e agregar dados de COVID-19 e multas, garantindo a consistência e a integridade dos dados. As consultas SQL foram projetadas para formatar e agrupar os dados, facilitando a análise no contexto do modelo star schema. O modelo star schema permite uma análise eficiente e organizada dos dados, combinando informações temporais e geográficas com os eventos registrados nas tabelas de fatos.

Se precisar de mais alguma explicação, estou à disposição!