

Trabalho Final

Rafael Zimbrão, Kayan Martins e João Pedro Ferreira

<https://github.com/razimbrao/dw-trabalho-final>

Introdução

Este trabalho tem como objetivo analisar os dados de crimes cometidos em Chicago, fornecendo relatórios e dashboards que auxiliem na compreensão das ocorrências. Para isso, utilizamos a arquitetura Kimball, estruturando um Data Warehouse que facilita a análise dos dados.

A tecnologia adotada inclui **PostgreSQL** para armazenamento, **PHP** para tratamento dos dados e **Power BI** para a criação dos relatórios.

Processo ETL (Extract, Transform, Load)

1. Extração (Extract)

O processo começa lendo os dados do conjunto **Crimes - 2001 to Present** e inserindo-os na tabela **staging_area**

2. Transformação (Transform)

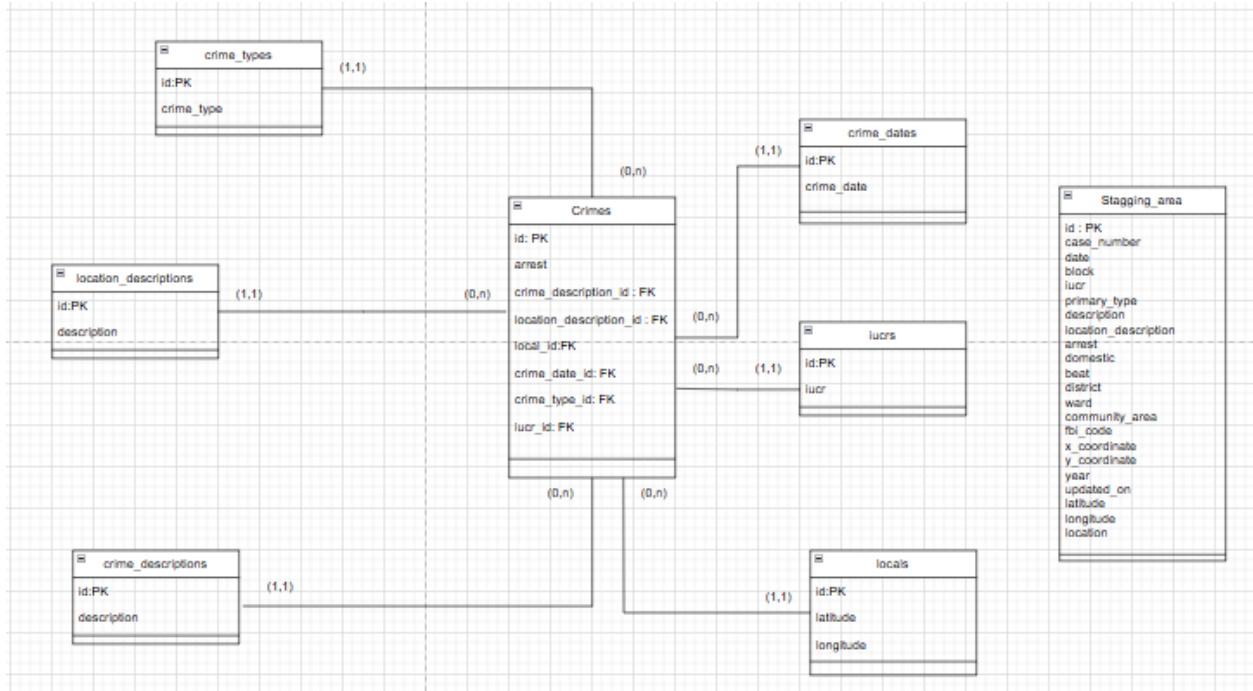
A transformação dos dados foi realizada em lotes de **1.000 registros** por vez. Durante essa etapa, os dados foram normalizados e organizados nas tabelas **locals**, **location_descriptions**, **crime_descriptions**, **crime_types**, **iucrs** e **crime_dates** sem duplicatas.

3. Carga (Load)

Após processar todas as dimensões, os registros da **fato crimes** são criados na tabela **crimes**, substituindo valores textuais da **staging_area** pelos IDs das dimensões criadas

Estrutura do Data Warehouse

Modelo Dimensional



Após o tratamento na Staging Area, os dados são organizados em um esquema dimensional:

Dimensão CrimeDescription

- id
- description

Dimensão LocationDescription

- id
- description

Dimensão Local

- id
- latitude
- longitude

Dimensão CrimeDate

- id
- crime_date

Dimensão CrimeType

- id
- crime_type

Dimensão IUCR

- id
- iucr

Fato Crimes

- id
- arrest
- crime_description_id
- location_description_id
- local_id
- crime_date_id
- crime_type_id
- iucr_id

Granularidade do Data Warehouse

A granularidade do nosso Data Warehouse define o nível de detalhe armazenado em cada fato e dimensão. No nosso modelo:

- **CrimeDate:** Armazenamos apenas a data completa do crime (**dia, mês e ano**), sem considerar a hora do ocorrido.
- **Locals:** Cada localização é registrada de forma única, baseada em sua **latitude e longitude**.

Métricas Geradas

A partir da tabela fato **Crimes**, geramos as seguintes métricas dentro do **Power BI**:

- **Crimes com Prisão:** Quantidade de crimes onde ocorreu uma prisão.
- **Taxa de Prisões:** Percentual de crimes que resultaram em prisão em relação ao total de crimes registrados.
- **Total Crimes:** Conta a quantidade de crimes que possuem as mesmas características