

Data Warehouse e Data Marts



Boletins de ocorrências de São Paulo

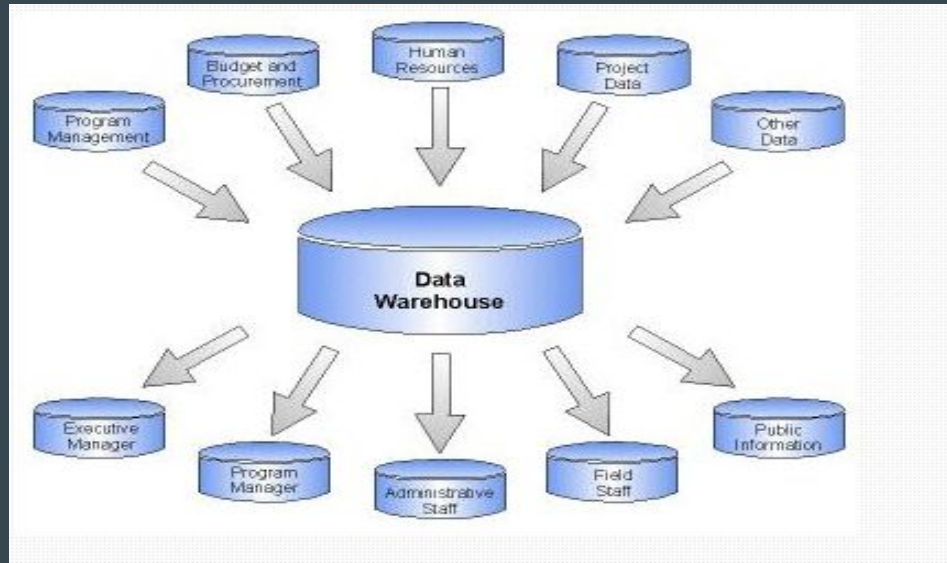
Ícaro Aragão e Sara Silva

Data Warehouse

- Armazena informações de uma organização/empresa;
- otimizado para consultas complexas e relatórios;
- integrado, não volátil, variante ao tempo;
- visão do todo;
- objetivo de fornecer suporte à tomada de decisão;
- analisar o passado para a tomada de decisão e fazer uma projeção para o futuro.

Data Warehouse

- Os dados podem ser coletados de várias fontes;
- os dados serão agrupados num único local para que o acesso às informações seja feita por uma interface única.



Data Mart

- Subconjunto de um data warehouse, orientado a uma determinada área da organização;
- oferece informações mais detalhadas sobre um departamento em questão;
- geralmente menor que 100GB;
- orienta decisões mais focadas ao invés de decisões mais gerais sobre a empresa

Online Transaction Processing - OLTP

- Monitorar e processar as funções rotineiras de uma organização;
- operações curtas de escrita;
- eliminar ao máximo a redundância;
- focados em busca rápida;
- só lidam com o dado atual

Online Analytical Processing - OLAP

- Operações de leitura complexa;
- realizar análise detalhada de dados transacionais históricos;
- ajudam a analisar um dado no Data Warehouse.

ETL X ELT

- ETL - Extract, Transform, Load:
 - extrai os dados de sistemas origens;
 - converte um dado para um formato adequado para busca e análise;
 - carrega os dados num Data Warehouse.
- ELT - Extract, Load, Transform:
 - coleta os dados de fontes diferentes;
 - carrega os dados num sistema, como um Data Warehouse;
 - transforma os dados sob demanda;
- O ELT fornece um carregamento mais rápido que o ETL, mas requer um sistema poderoso computacionalmente para ser capaz de realizar as transformações por demanda

OLTP

Online Transaction Processing

ERP

Sistema de Gestão Empresarial

Arquivos e documentos



CRM

Gestão de Relacionamento com o Cliente

Banco de dados operacional

ETL
Extract
Transform
Load

Data Warehouse

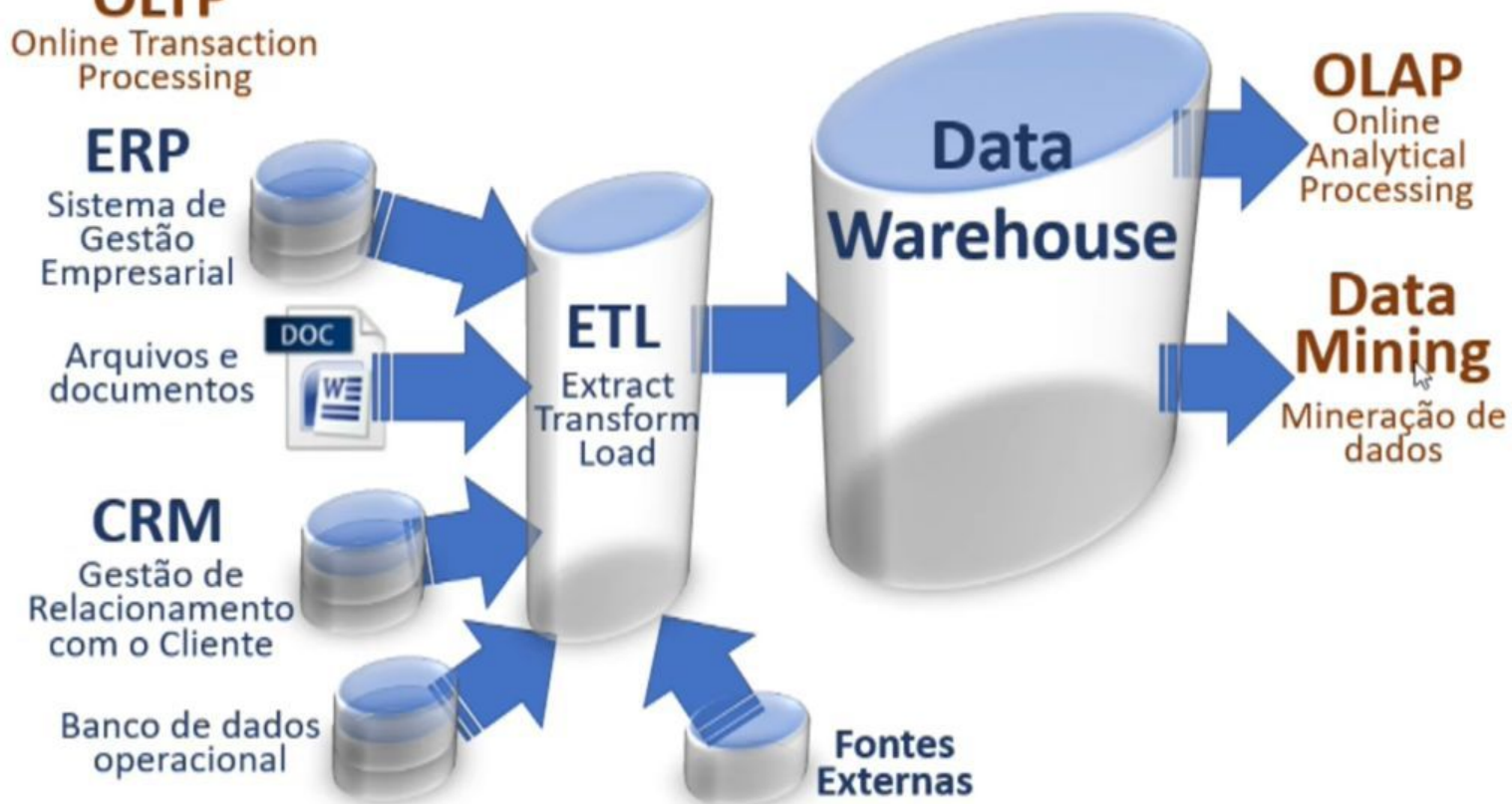
OLAP

Online Analytical Processing

Data Mining

Mineração de dados

Fontes Externas



Ferramentas de DW



Amazon **Redshift**



panoply



Google BigQuery

Base escolhida - critérios

- Grande volume de dados;
- tema instigante;
- possibilidade de geração de relatórios pertinentes;
- pelo menos três entidades relacionadas.

Base escolhida - base

- Ocorrências de crimes que foram executadas pela Polícia Militar de São Paulo em 2016;
- não contém todos os detalhes por questões legais e de privacidade;
- possui 775 mil tuplas;
- contém colunas a respeito da ocorrência, do boletim, do departamento e da delegacia;
- A base está disponível [aqui](#).

Diagrama Entidade-Relacionamento

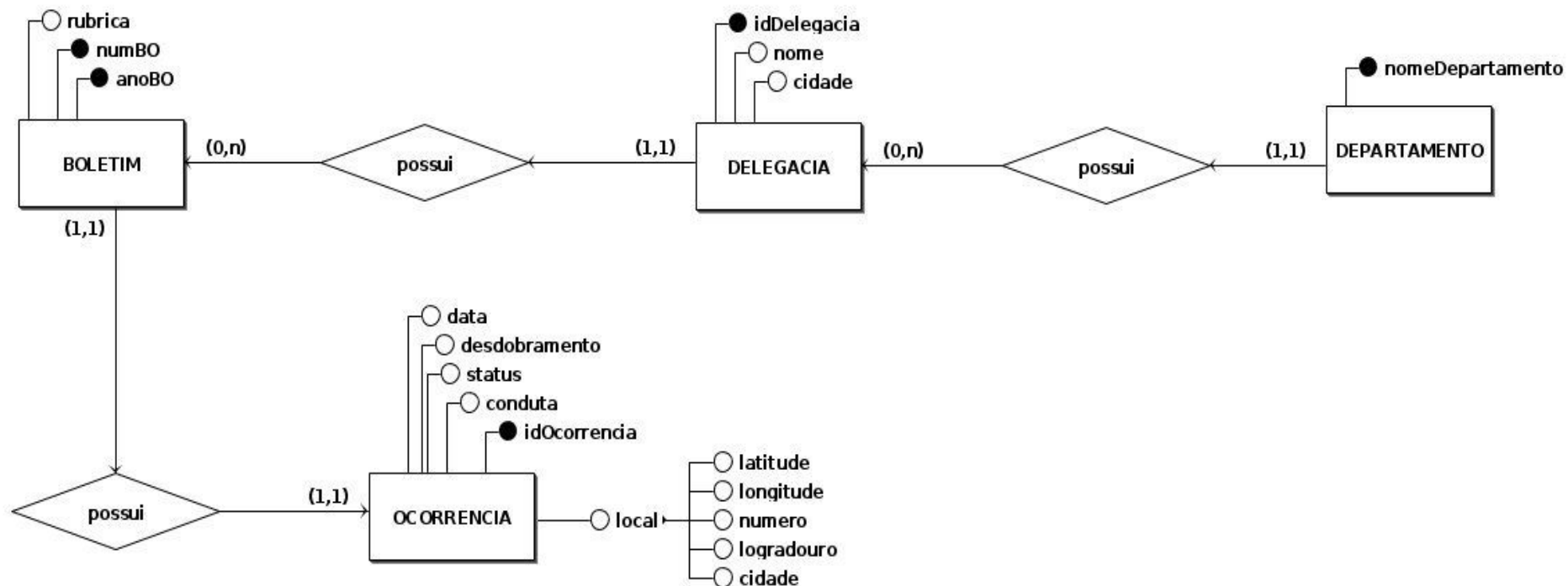
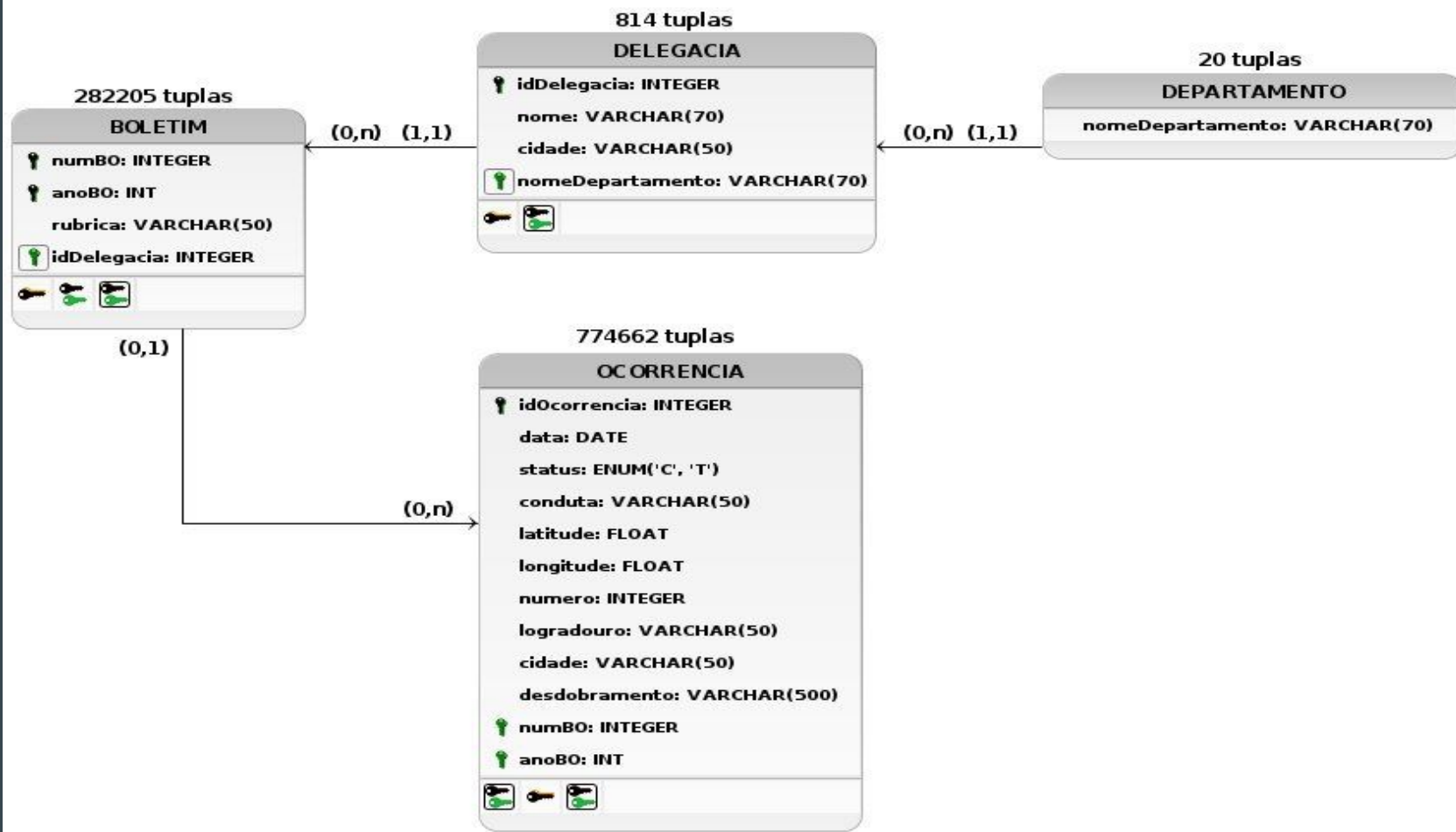


Diagrama Lógico



Referências

- **O que é Data Warehouse?**. Disponível em: <<https://dba-pro.com/o-que-e-data-warehouse/>>. Acesso em: 18/11/2019.
- **Crime data in Brazil**. Disponível em: <<https://www.kaggle.com/inquisitivecrow/crime-data-in-brazil>>. Acesso em: 18/11/2019.
- **Introdução à tecnologia Data Warehouse**. Disponível em: <<https://www.devmedia.com.br/introducao-a-tecnologia-data-warehouse/27629>>. Acesso em: 18/11/2019.
- **Data warehouse**. Disponível em: <<https://www.devmedia.com.br/data-warehouse/12609>>. Acesso em: 18/11/2019
- **Data Warehouse Concepts: Traditional vs. Cloud**. Disponível em: <<https://panoply.io/data-warehouse-guide/data-warehouse-concepts-traditional-vs-cloud/>>. Acesso em: 18/11/2019.