



Data Warehouses

Introdução



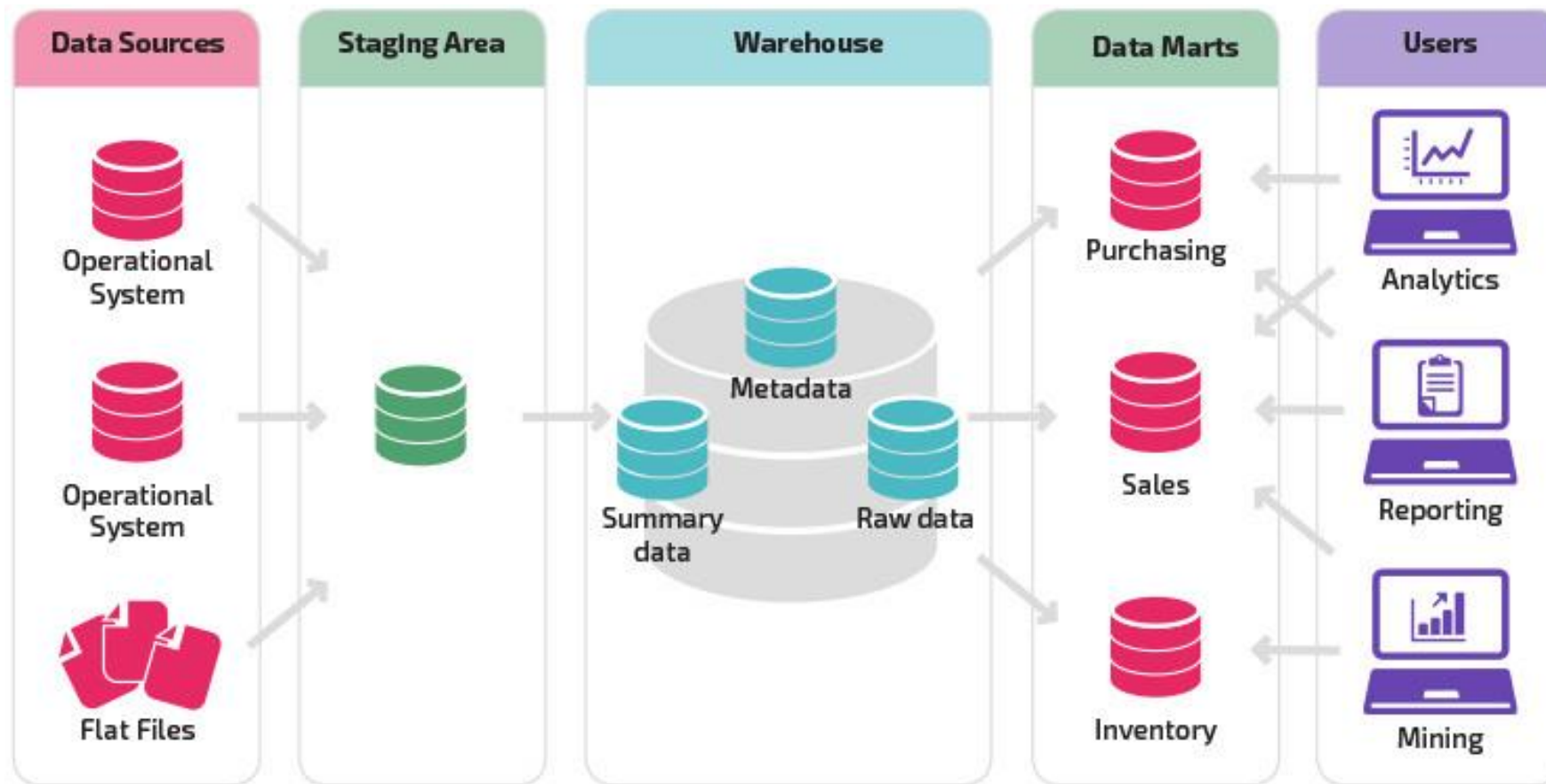
- Grande crescimento do fluxo de dados
- Surgem vários temas em relação a esse crescimento como:
 - Coletamos uma quantidade significativa de dados, mas não conseguimos acessá-los;
 - Precisamos fatiar e dividir os dados de todas as maneiras;
- Esses temas se transformam em requerimentos como:
 - O sistema *data warehouse* e *business intelligence* deve tornar a informação acessível;
 - O sistema DW/BI deve apresentar a informação de maneira consistente.

Definição da tecnologia

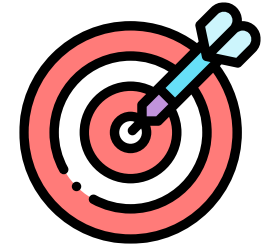


- Armazém de dados
- Coleção de dados orientada a assunto
- Características:
 - Base SQL
 - Pré-processamento
 - Visualização
 - ML/AI
- Otimizada
- Diversas tecnologias
- Data Mart

Definição da tecnologia



Objetivos Principais

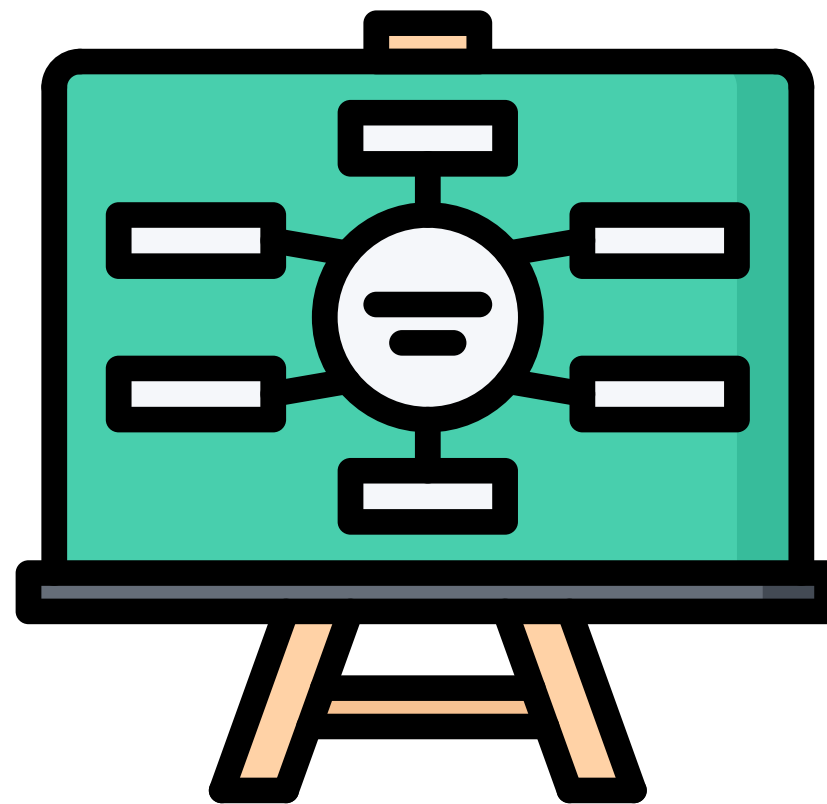


- ETL
- Principais características:
 - Atualização automática regular da base;
 - Armazena dados históricos de meses, ano e até décadas;
 - Proporciona análises avançadas e consultas ad-hoc;
 - Usa modelagem parcialmente normalizada para otimizar o desempenho;
 - Acessa milhões de linhas ao mesmo tempo.
- Fonte de consulta (BI)
- Convergir informações
- Análise de dados

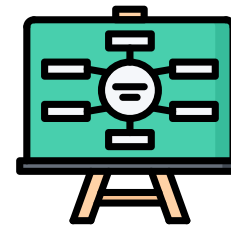
Schema de um data warehouse

Conceitos

- Fact tables que são tabela principal que reúne os atributos das tabelas secundárias mais os seus próprios atributos principais (exemplo: valor total da venda);
- Dimension tables, tabela secundária de características sobre a tabela principal (exemplo: horário que a venda foi realizada).

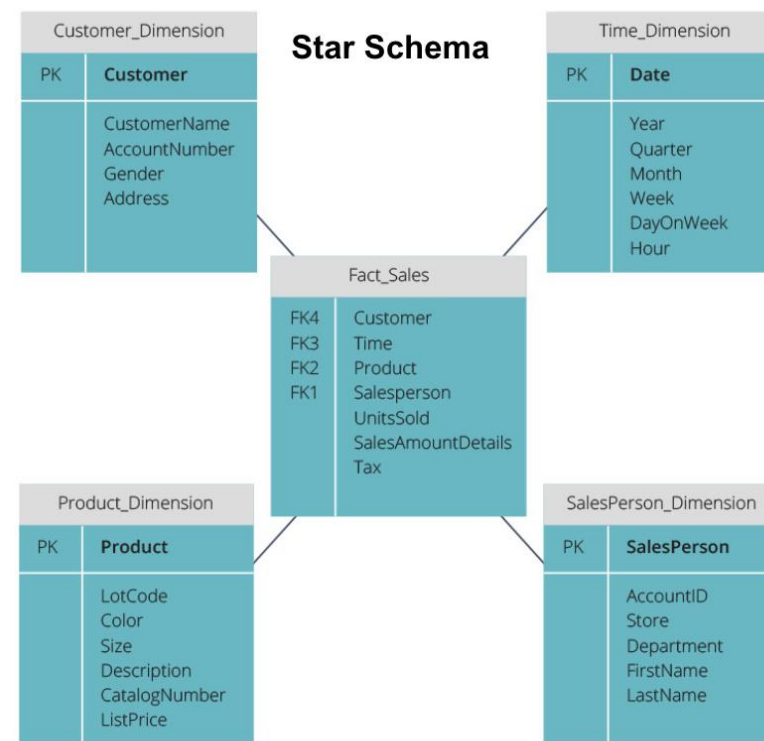


Schema de um data warehouse



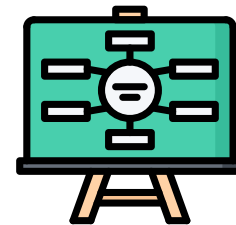
Star Schema

- Dividir fact table em dimension tables desnormalizadas.
- Maior velocidade de consultas por realizar apenas uma junção das dimension tables com a fact table.



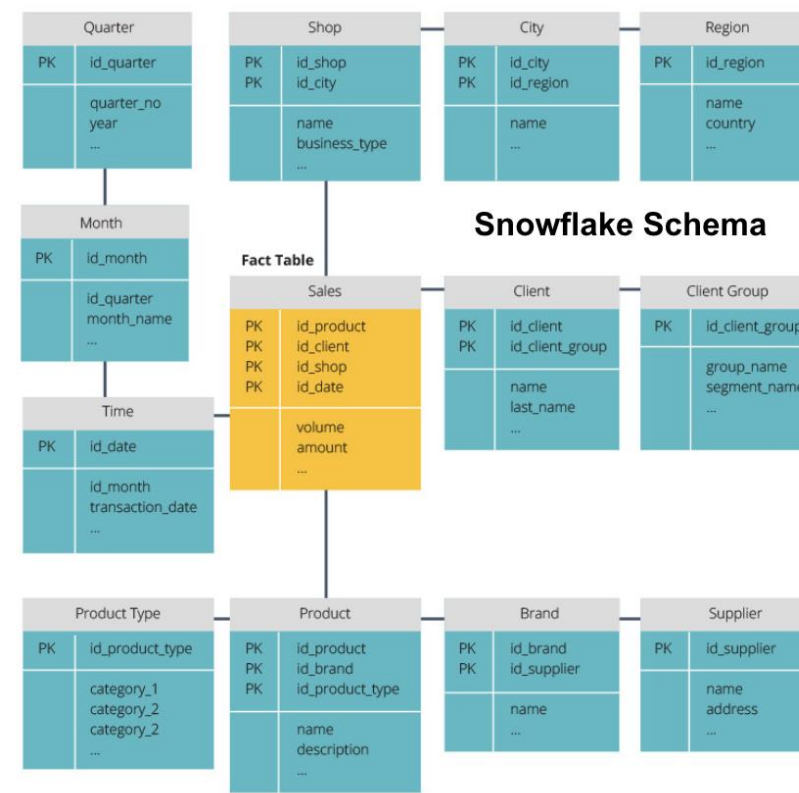
Legend: PK = primary key | FK = foreign key

Schema de um data warehouse



Snowflake Schema

- A fact table é dividida em uma série de dimension tables normalizadas.
- Maior integridade dos dados, porém consultas são mais lentas pela maior quantidade de junções para acessar dados relevantes.



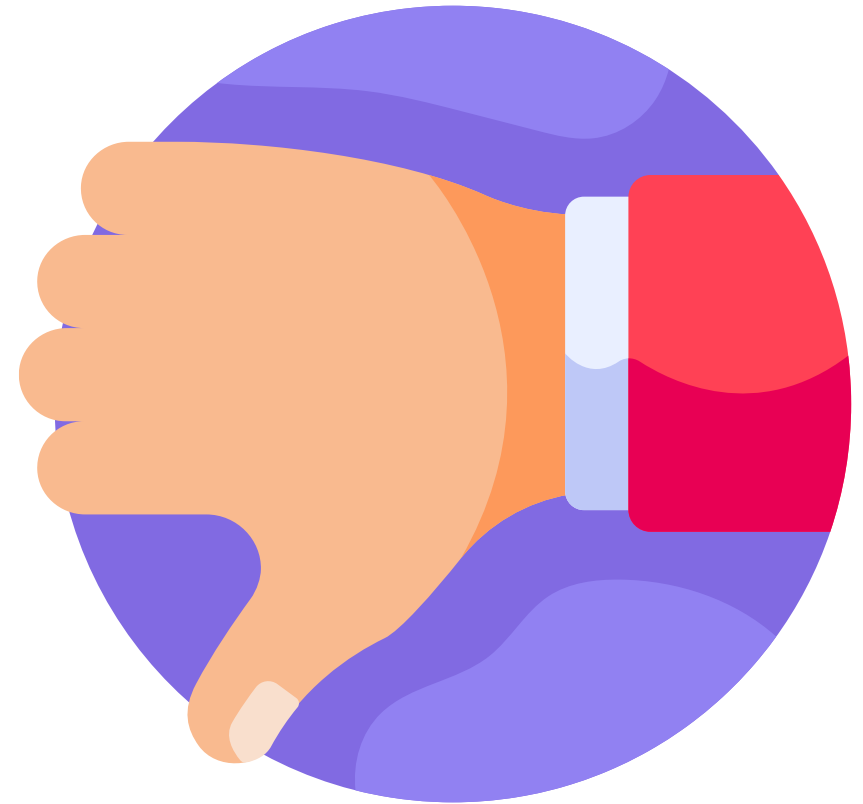
Vantagens

- Integração
- Não volátil
- Variável com o tempo
- Orientado para o assunto



Desvantagens

- Montagem
- Gerenciamento
- Elevado padrão de controle
- Integração de dados é muito desafiadora
- Treinamento da equipe
- Altos custos



Exemplos de Uso

- Vivo (Telefônica)
- Ministério da Justiça do Brasil
- Avon



Tecnologias de cloud data warehouse

- Google BigQuery
- Oracle Autonomous Data Warehouse
- Amazon Redshift



BigQuery

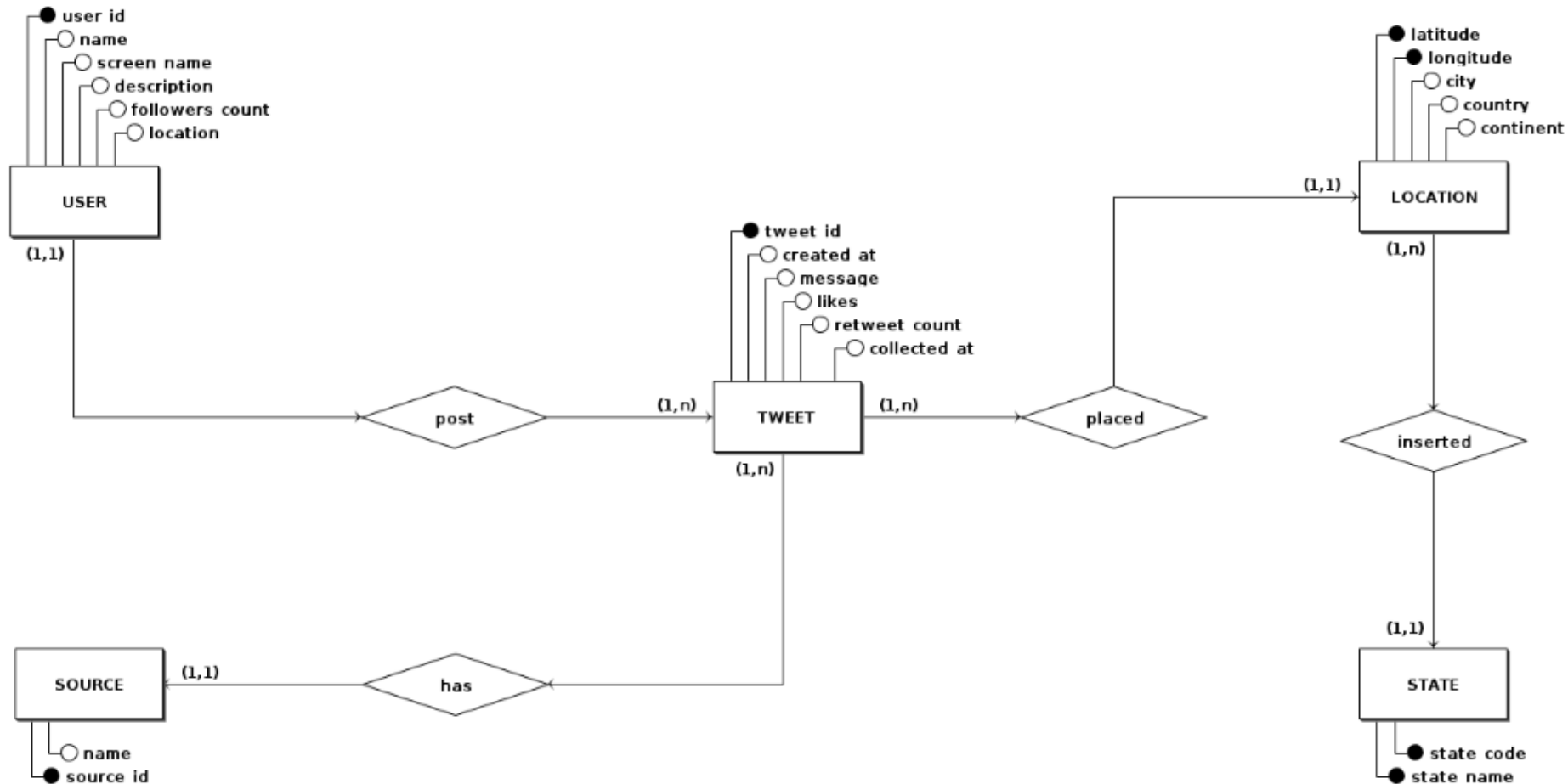


ORACLE
AUTONOMOUS
DATA WAREHOUSE
CLOUD



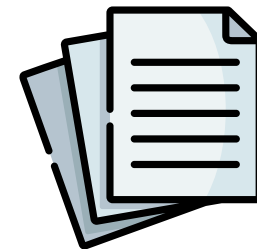
Amazon **Redshift**

Documentação da Base de Dados



Kaggle
Dump

Documentação da Base de Dados



[Kaggle](#)
[Dump](#)

