

# Análise da popularidade de jogos e sua previsibilidade dentro da plataforma Steam

Aluno: José Ronaldo Pereira Diniz Neto  
Email: [Ronaldoml2016@hotmail.com](mailto:Ronaldoml2016@hotmail.com)

Fonte dos Dados:

<<https://www.kaggle.com/datasets/artermiloff/steam-games-dataset/data>>

Aplicação OLAP:

<[https://1drv.ms/f/c/cd829de0553489b7/Esl\\_04X\\_st9OqYrF0aAV2UoBqMSZ3PDKNOiHlfCQQ5n-yA?e=rNwfxk](https://1drv.ms/f/c/cd829de0553489b7/Esl_04X_st9OqYrF0aAV2UoBqMSZ3PDKNOiHlfCQQ5n-yA?e=rNwfxk)>

Github: <<https://github.com/RonaldoPD/DataMart-Steam>>



# ETAPA 1 - PLANEJAMENTO



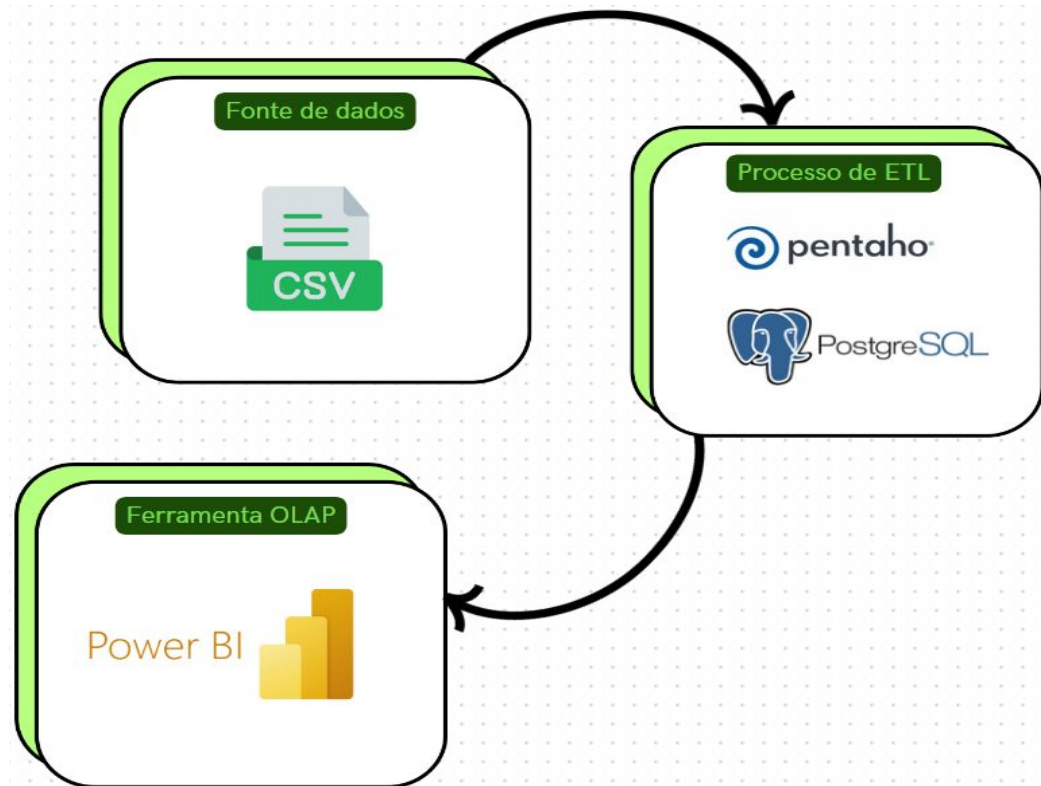
# 1. Contextualização

O Data Mart proposto tem como foco o setor de distribuição de jogos digitais, com ênfase na plataforma Steam. O objetivo principal é oferecer suporte analítico para a análise da popularidade dos jogos e a previsibilidade de seu desempenho, com base em indicadores históricos e comportamentais dos usuários.

## 2. Escopo/objetivo do Data Mart

O data mart desse projeto será desenvolvido com o objetivo de auxiliar demandas de setores como o marketing digital, tomadas de decisões acerca no setor de e-commerce e business intelligence. Por meio de análises diversas de dados como preço, reviews positivas e negativas, estimativa da quantidade de donos de determinado jogo, categorias, tags e tempo médio de campanha é esperado que seja possível alcançar insights sobre o desempenho de vendas de jogos, tendências do mercado, preferências do usuário, ajustes de preços e campanhas direcionadas utilizando dados de jogos lançados de 1997 até 2025.

### 3. Arquitetura Tecnológica



## 4. Processo

1. Planejamento do data mart
2. Levantamento das necessidades
3. Modelagem dimensional do data mart
4. Projeto físico do BD's
5. ETC do dados
6. Desenvolvimento e aplicações OLAP

## 4. Abordagem

A construção do data mart será baseada na abordagem bottom-up juntamente com o star schema onde a tabela central(tabela fato) se conectará com outras tabelas relacionadas(dimensões) via foreign key.

## 5. Usuários

Os possíveis usuários do data mart serão empregados do setor de marketing, desenvolvedores de jogos e executivos que atuam nas tomadas de decisões.





## ETAPA 2 - LEVANTAMENTO DAS NECESSIDADES



## 6. Consultas de Apoio à Decisão

- Jogos mais recomendados em determinado período de tempo.
- Jogos mais populares baseados na maior quantidade de jogadores simultâneos.
- Faixa de preço dos jogos por categoria?
- Comparativo entre tempo de jogatina desde o lançamento e o tempo de jogatina recente dos jogos
- Gênero de jogos com maior quantidade de donos estimados por ano.
- Evolução da quantidade de jogadores simultâneos ao decorrer dos anos.
- Quantidade de jogatina total por mês em um ano definido.

## 7. Indicadores do <negócio>

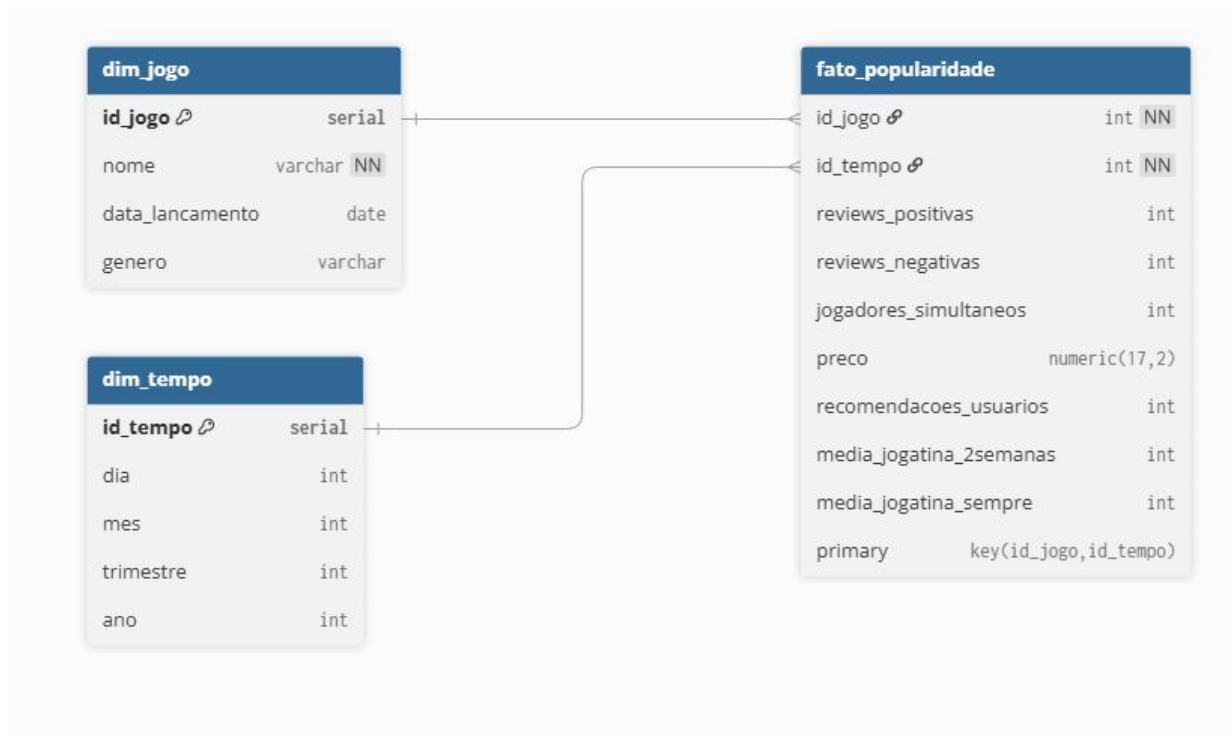
- Quantidade de recomendações dos jogos pelos usuários.
- Reviews positivas ou negativas disponíveis apenas para compradores na página do jogo na Steam.
- Quantidade total de jogadores simultâneos alcançadas por determinado jogo.
- Histórico de descontos e variação de preços dos jogos.



## ETAPA 3 - MODELAGEM



# 9. Modelo Relacional



[https://docs.google.com/document/d/1xf6cjT2c\\_xT\\_\\_9ZSI3FoPPuP5Pb3MgSITgHrqk0ltYo/edit?usp=sharing](https://docs.google.com/document/d/1xf6cjT2c_xT__9ZSI3FoPPuP5Pb3MgSITgHrqk0ltYo/edit?usp=sharing)

# 10. Modelo Dimensional

- Área de Negócios
  - Indústria dos jogos/distribuição de jogos digitais
- Processo
  - Analisar dados acerca dos jogos lançados na plataforma Steam.
- Granularidade
  - Jogo x Tempo

# 10. Modelo Dimensional

## D. Atributos e Hierarquia das Dimensões

(dim\_tempo: dia -> mês -> trimestre -> ano)

(dim\_jogo: nome, data\_lançamento, genero, estimativa\_donos)

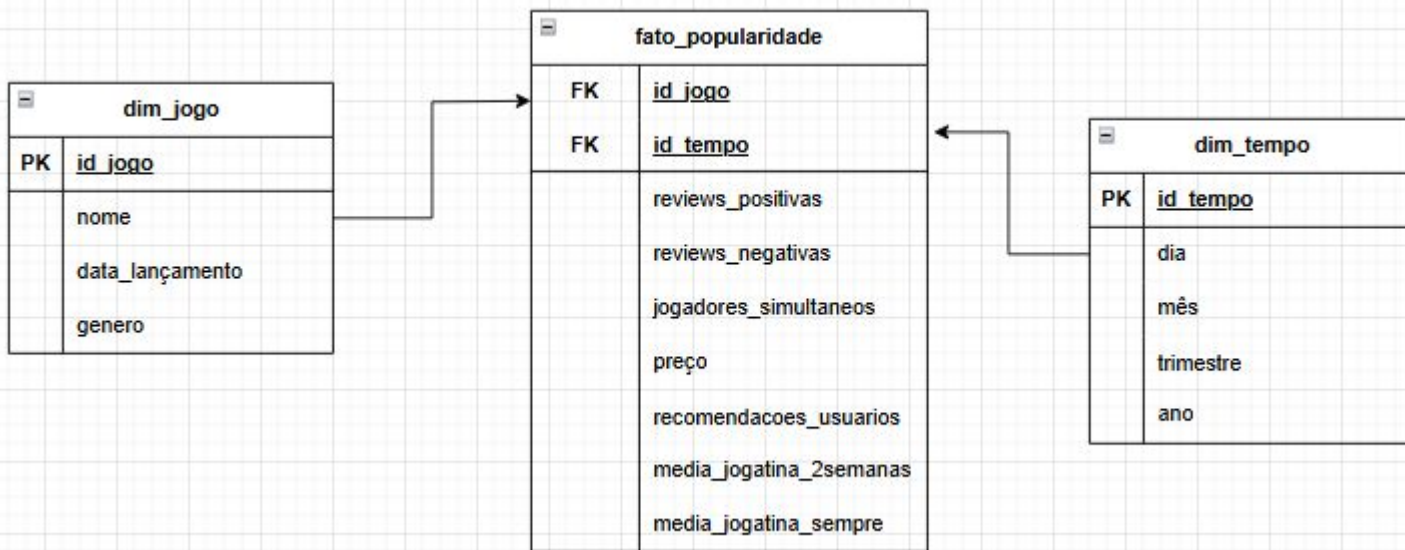
# 10. Modelo Dimensional

- E. Aditivas: jogadores\_simultaneos, reviews\_positivas, reviews\_negativas, recomendacoes\_usuarios
- F. Semi-aditivas: media\_jogatina\_2semanas, media\_jogatina\_sempre
- G. Não aditiva: Preço



# 10. Modelo Dimensional

## F. Esquema Estrela



# 10. Modelo Dimensional

G. Simulação de inserção de 10 “fatos”

id_jogo	id_tempo	reviews_positivas	reviews_negativas	jogadores_simultaneos	preço	recomendacoes_usuarios	media_jogatina_2semanas	media_jogatina_sempre
70	1	12000	800	50000	49.99	9500	60	32000
90	2	8500	1200	35000	39.99	7000	150	15000
120	3	15000	400	62000	59.99	14000	1200	1400
170	4	2200	300	8000	19.99	1800	360	800
190	5	4600	1500	12000	29.99	3900	400	9900
210	6	300	120	600	14.99	250	700	5601
230	7	7500	1300	27000	34.99	6900	790	991
315	8	9800	900	41000	44.99	8500	4500	8750
490	9	660	95	1300	9.99	590	856	1700
675	10	5400	500	15000	24.99	4900	740	7800

## 10. Modelo Dimensional do Data Mart (lógico)

### H. Estimativa de espaço

Supondo que a steam adicione em média 40 jogos por dia e há uma perspectiva de armazenamento para 10 anos, logo  $40 \times 365 \times 10 = 146000$  novos registros

A tabela fato do data mart possui 9 chaves onde cada uma possui 4 bytes, assim cada linha irá ocupar aproximadamente 36 bytes. Conclui-se que a estimativa para a tabela fato será  $146000 \times 36 = 5256000$  bytes(4,9mb)

Por fim, cada dimensão ocupa cerca de 25% da tabela fato, como o data mart do projeto possui 2 dimensões, estima-se que cada uma 1,3mb totalizando 2,6mb

Estimativa final: 7,5mb



# ETAPA 4 - PROJETO FÍSICO DO BD



# 11. Modelo Relacional do Data Mart (físico)

```
1 CREATE TABLE dim_tempo(  
2     id_tempo SERIAL PRIMARY KEY,  
3     dia_mes INT,  
4     mes INT,  
5     trimestre INT,  
6     ano INT,  
7     data_completa DATE,  
8     dia_semana INT,  
9     mes_ano VARCHAR(40),  
10    data_mes_dia mes_ano VARCHAR(40),  
11    mes_descricao mes_ano VARCHAR(40),  
12    dia_descricao mes_ano VARCHAR(40)  
13 );
```

```
CREATE TABLE dim_jogo (  
    id_jogo SERIAL PRIMARY KEY,  
    nome VARCHAR(400) NOT NULL,  
    data_lancamento DATE,  
    genero VARCHAR(40)  
);
```

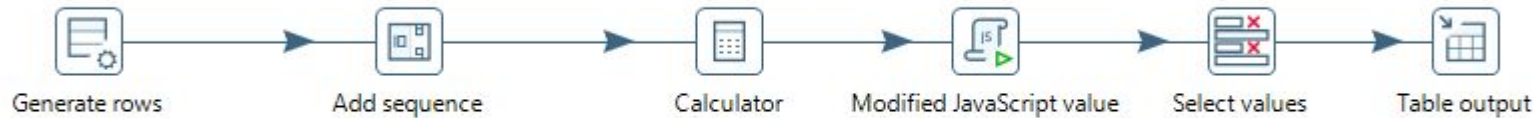
```
CREATE TABLE fato_popularidade (  
    id_jogo INT NOT NULL,  
    id_tempo INT NOT NULL,  
    reviews_positivas INT,  
    reviews_negativas INT,  
    jogadores_simultaneos INT,  
    preco NUMERIC(10,2),  
    recomendacoes_usuarios INT,  
    media_jogatina_2semanas INT,  
    media_jogatina_sempre INT,  
    PRIMARY KEY (id_jogo, id_tempo),  
    FOREIGN KEY (id_jogo) REFERENCES dim_jogo(id_jogo),  
    FOREIGN KEY (id_tempo) REFERENCES dim_tempo(id_tempo)  
);
```



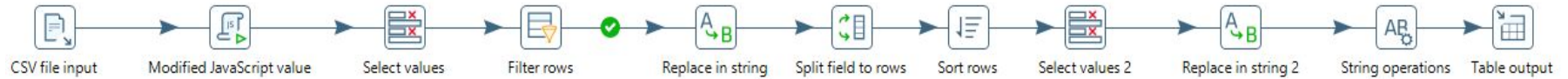
# ETAPA 5 - EXTRAÇÃO, TRANSFORMAÇÃO E CARGA



## 12. Plano de Carga da Dimensão Tempo



# 13. Plano de Carga da Dimensão Jogo





## 15. Plano de Carga da Fato





# ETAPA 6 - APLICAÇÃO OLAP e PAINEL DE BORDO



# 16. Consulta OLAP 1

Principais métricas exibidas ao usuário por meio da consulta, varia de acordo com o tempo selecionado.



## 16. Consulta OLAP 2

Informações sobre as métricas de reviews exibidas ao usuário, separadas em positivas e negativas, filtradas para cada gênero. Varia de acordo com o tempo selecionado.

Total de reviews		
59.016.836		
Reviews positivas x Reviews negativos(%)		
Action	85,74%	14,26%
Adventure	93,47%	6,53%
Animation & Modeling	93,83%	6,17%
Audio Production	77,92%	22,08%

# 16. Consulta OLAP 3



# 16. Painel de Bordo XXXX (ponto extra)



# Referências

1. <https://www.kaggle.com/datasets/artermiloff/steam-games-dataset/data>