

Resumo das Transformações

Bronze → Silver Layer

1. Movies (movies_metadata.csv → movies.parquet)

- ☒ **Parse de JSON:** genres, production_companies, countries, languages, collection
- ☒ **Conversão de tipos:** budget, revenue, runtime, ratings para numéricos
- ☒ **Extração de ano:** release_year do release_date
- ☒ **Cálculos derivados:** profit, ROI, flags has_budget/has_revenue
- ☒ **Filtros:** apenas filmes com status='Released' e ano válido
- ☒ **Resultado:** 45.379 filmes válidos, 30 colunas estruturadas

2. Credits (credits.csv → credits.parquet)

- ☒ **Parse de cast:** extrair top 5 atores por ordem de aparição
- ☒ **Parse de crew:** extrair diretor (job='Director')
- ☒ **Resultado:** 45.476 registros, 5 colunas (cast, crew, director)

3. Keywords (keywords.csv → keywords.parquet)


- ☒ **Parse de keywords:** converter JSON em lista de strings
- ☒ **Resultado:** 46.419 registros, 3 colunas (keyword_ids, keyword_names)

4. Ratings (ratings_small.csv → ratings.parquet)



- ☒ **Conversão de timestamp:** Unix timestamp → datetime
- ☒ **Resultado:** 100.004 avaliações válidas

Silver → Gold Layer



1. Movies Enriched (dataset completo)

-  **Merge:** movies + credits + keywords
- ☒ **Resultado:** 46.543 filmes com todas as informações

2. Yearly Analytics (por ano)

-  **Agregações:** COUNT, AVG, SUM por release_year
-  **Métricas:** movie_count, budget, revenue, profit, rating, popularity, runtime
- ☒ **Resultado:** 120 anos analisados (1900-2020)

3. Genre Analytics (por gênero)

-  **Explode:** um filme pode ter múltiplos gêneros
-  **Agregações:** COUNT, AVG, SUM por gênero
- ☒ **Resultado:** 20 gêneros com estatísticas completas

4. Top Movies (rankings)

- 🏆 **Top 100 por receita**
- 💰 **Top 100 por lucro**
- ★ **Top 100 por avaliação** (mínimo 100 votos)
- ☑ **Resultado:** 300 filmes ranqueados

5. Director Analytics (por diretor)

- 📊 **Agregações:** COUNT, AVG, MAX, SUM por diretor
- 🔍 **Filtro:** apenas diretores com 2+ filmes
- ☑ **Resultado:** 4.351 diretores analisados

📊 Resumo de Tamanho

Camada	Arquivos	Tamanho	Compressão
Bronze	7 CSVs	~900 MB	-
Silver	4 Parquets	~50 MB	94%
Gold	5 Parquets	~22 MB	98%

Total de redução: 900 MB → 72 MB ✨