

# Trabalho de ETL

## Considere a seguinte situação:

Uma produtora possui um sistema OLTP com um banco de dados transacional para controlar as exibições de filmes em diferentes salas de cinema, distribuídos por diversas cidades no Brasil e no exterior. A empresa também acompanha a quantidade de pessoas que comparece a cada exibição, bem como o valor arrecadado.

A gerência necessita de um ambiente analítico de forma a responder algumas questões sobre o acompanhamento dos negócios. O sistema não possui um **Modelo Entidade-Relacionamento** (MER), então foi entregue uma sugestão de MER criado por um dos colaboradores.

## Pede-se:

### 1) Avalie o MER recebido, levando em conta os dados da fonte.

👉 O modelo deverá ser gerado utilizando o software **BrModelo**, e salvo na pasta "doc".

### 2) Crie uma rotina de ETL utilizando linguagem SQL

👉 Os scripts para extração, transformação e carga se encontram no diretório src.

👉 O cliente deve ser capaz de executar os scripts passando apenas seu caminho relativo como parâmetro para o `sqlite3`. Os scripts sempre serão executados no diretório raiz do projeto. Portanto, deverão funcionar da seguinte forma:

```
sqlite3 < src/<nome_do_script>
```

👉 Deverão ser utilizados os arquivos `.db` presentes no diretórios data (atenção ao caminho passado para a cláusula `ATTACH DATABASE` )

### 3) Crie queries para as seguintes perguntas/necessidades:

- 👉 Acompanhar a evolução semanal da arrecadação
- 👉 Qual a média de público das exibições **por ator**?
- 👉 Qual a média de público das exibições **e por diretor**?
- 👉 Qual foi o total arrecadado por mês em cada estado?

### Observações:

- 👉 Os arquivos necessários ao projeto deverão ser organizados no seguinte formato:

```
projeto_etl
├── data      # Diretório com os arquivos de banco de dados
│   ├── distrib_dw.db
│   ├── distrib_oltp.db
│   └── distrib_stg.db
├── doc       # Enunciado e diagramas criados pelo grupo
│   ├── ...
│   ├── sugestao_olap.JPG
│   ├── sugestao_oltp.JPG
│   └── enunciado.md
└── src       # Scripts de ETL e respostas para as perguntas de negócio
    ├── 1_extração.sql
    ├── 2_transformacao.sql
    ├── 3_carga.sql
    ├── 4_perguntas.sql
    └── cria_stg_dim_tempo.py  # Código python para criação da tabela temp
```

- 👉 É importante que o grupo execute os scripts estando no diretório raiz do projeto e verifique o funcionamento. **Automação é um dos pilares dos projetos de ETL**

### BÔNUS

- 👉 Se o grupo criar um arquivo `.sh` ou `.bat` para executar toda a rotina (a ser executado no diretório raiz)