Introdução

Neste trabalho dei continuidade ao desenvolvimento do projeto relativo a unidade curricular Introdução e Processamento Analítico de Informação que se trata da análise de avaliações de filmes por base na média de estrelas.

Desnvolvimento do trabalho

Prompt para o Chat GPT “Boa Tarde, para definir um processo de negócios seguro e que dê valiosos insights, preciso que me ajude, denominamos o processo de negócio como "Avaliação de Filme por média de estrelas" referente a plataforma IMDB que nos forneceu os dados/as dimensões dos Filmes, Atores, Realizadores... criamos a dimensão Data. Neste momento já criamos o modelo dimensional mas não sabemos até que ponto falta agregar informação para termos uma tabela de factos suficientemente boa para responder às questões que possam surgir ao gestor da plataforma.”

Adição da dimensão demografia e da dimensão de géneros em modelo em floco de neve que está diretamente relacionada com a dimensão filme e a criação de uma tabela ponte Filme Género que permite essa ligação e a agregação da informação de que 1 filme pode ter mais de um género que é representado desta forma pela Tabela Filme Género assim o nosso projeto consegue responder a perguntas como :

* **"Qual gênero tem as melhores avaliações médias?"**
* **"Quais gêneros são mais populares ao longo dos anos?"**
* **"O público avalia melhor filmes de um único gênero ou de múltiplos gêneros?"**

Para a construção da nossa matriz em bus referenciamos alguns processos de negócio a seguir apresentados:

1. **Avaliação de Filme** 📽️  
    ➡ Média de avaliação dos filmes (exemplo: 4.5 estrelas no IMDb).
2. **Avaliação por Realizador** 🎬  
    ➡ Qual a média de avaliações dos filmes de cada realizador (diretor)?
3. **Avaliação por Ator** 🎭  
    ➡ Qual a média de avaliações dos filmes estrelados por cada ator?
4. **Avaliação por Gênero** 🍿  
    ➡ Quais gêneros têm as melhores avaliações?
5. **Avaliação por Demografia** 🌍  
    ➡ Como a avaliação de filmes varia entre diferentes países, faixas etárias e gêneros?
6. **Avaliação ao Longo do Tempo** 🕒  
    ➡ Como as avaliações mudaram ao longo dos anos e meses?

Neste momento já garantimos o modelo dimensional e o matriz em bus com informações importantes para mais tarde poder definir as hierarquias adequadas para, por exemplo, a dimensão data.

Depois de termos as seguintes hierarquias já devidamente criadas:

* Hierarquia Temporal
  + Ano
  + Trimestre
  + Mês do Ano
  + Semana do Ano
  + Dia da Semana
  + Data Completa
* Hierarquia Demográfica
  + País
  + Região
  + Faixa Etária
* Hierarquia Filme
  + Género
  + Título Principal (Filme)

TAREFAS:

* Aplicar Regras para a denominação dos atributos das dimensões (Pág. 160) acrescentar o nome da dimensão no atributo em casos que possam causar confusão para a interpretação ;
* Na dimensão Género - diferenciar um atributo com "Nome do Género em Inglês" e uma com "Nome do Género em Português”

[Tableau help](https://help.tableau.com/current/pro/desktop/pt-br/datasource_multitable_normalized.htm?source=productlink)

Ótimo! Como você está utilizando o Tableau, o próximo passo é garantir que suas hierarquias de dimensões estejam bem definidas para facilitar a navegação nos dados.

### **Passos para Definir Hierarquias no Tableau**

1. **Abrir o Tableau e Conectar à Fonte de Dados**
   * Importe suas tabelas dimensionais e de fatos.
   * Confirme que os relacionamentos entre elas estão corretamente definidos.
2. **Criar uma Hierarquia na Dimensão Data**
   * No painel de **Dados**, localize a dimensão **Data**.
   * Clique com o botão direito sobre um dos campos e selecione **Criar Hierarquia**.
   * Dê um nome, como "Hierarquia Temporal".
   * Arraste os seguintes campos para dentro da hierarquia, na ordem:  
     + **Ano**
     + **Trimestre**
     + **Mês do Ano**
     + **Semana do Ano**
     + **Dia da Semana**
     + **Data Completa**
3. **Validar a Hierarquia**
   * Crie uma visualização simples, arrastando **Ano** para as colunas e **Número de Avaliações** para as linhas.
   * Expanda a hierarquia clicando no ícone “+” para verificar se a navegação entre os níveis funciona corretamente.
4. **Ajustes Finais**
   * Verifique se os valores agregados são calculados corretamente ao mudar de um nível para outro.
   * Caso necessário, ajuste os tipos de agregação para cada nível.

Depois de validar a hierarquia na dimensão Data, você pode replicar o processo para outras dimensões, como **Filme** (com níveis como Gênero > Filme) e **Demografia** (País > Região > Faixa Etária).

Caso tenha dúvidas sobre a implementação no Tableau, posso fornecer capturas de tela simuladas ou mais instruções detalhadas. Quer seguir assim?

Vamos lá! Mesmo que as etapas anteriores (ETL, criação de tabelas, etc.) ainda não estejam completamente implementadas, você pode planejar e até simular a construção do seu Cubo OLAP com base no modelo dimensional já definido. Vou explicar passo a passo, de forma simples, como você pode avançar nessa direção:

## **1. O que é um Cubo OLAP?**

Imagine o cubo OLAP como um **bloco de gelo** com várias faces. Cada face representa uma **dimensão** (por exemplo, Data, Filme, Realizador, etc.) e, no interior, você tem os **fatos** (como a média de avaliação e número de avaliações). Esse cubo permite que você “gire” os dados para ver diferentes ângulos – como ver a avaliação dos filmes ao longo do tempo, por gênero ou por região.

## **2. Planejando o Cubo OLAP**

Mesmo sem ter todos os dados fisicamente carregados, você pode definir:

* **Dimensões:** – Filme  
   – Data  
   – Realizador  
   – Ator  
   – Gênero  
   – Plataforma  
   – Demografia
* **Medidas (Fatos):** – Média de Avaliação  
   – Número de Avaliações  
   – (Outras medidas, se necessário: desvio padrão, soma das avaliações, etc.)

## **3. Passos para Implementar o Cubo OLAP**

### **Passo 3.1: Selecionar a Ferramenta OLAP**

Você pode escolher uma ferramenta que permita construir cubos OLAP. Alguns exemplos são:

* **Microsoft SQL Server Analysis Services (SSAS)**
* **Pentaho Mondrian**
* **Oracle OLAP**
* **Apache Kylin**

Essas ferramentas permitem criar cubos a partir das tabelas dimensionais e de fatos do seu Data Warehouse.

### **Passo 3.2: Definir a Estrutura do Cubo(Este no Tableau)**

Você deve configurar o cubo para que ele “saiba” quais são as dimensões e medidas. Por exemplo:

* **Defina as dimensões**:  
   – Crie hierarquias, como na dimensão Data: Dia → Mês → Ano; ou na dimensão Filme: Filme → Gênero, etc.
* **Defina as medidas**:  
   – A medida “Média de Avaliação” e “Número de Avaliações” são agregadas – você pode usar funções como AVG() ou SUM().

### **Passo 3.3: Modelar o Cubo**

Na ferramenta escolhida, você vai:

1. **Importar o Modelo Dimensional:** – Conectar-se ao seu banco de dados e selecionar as tabelas de fatos e dimensões.
2. **Configurar as Relações:** – Estabelecer os relacionamentos entre a tabela de fatos e as dimensões por meio das chaves.
3. **Criar Hierarquias:** – Exemplo: Na dimensão Data, crie uma hierarquia "Ano > Trimestre > Mês > Dia".
4. **Definir Agregações:** – Especifique como as medidas serão agregadas: média, soma, contagem etc.

### **Passo 3.4: Processar o Cubo**

Após modelar, a ferramenta precisará “processar” o cubo. Isso significa que ela vai ler os dados das tabelas, calcular as agregações e armazenar os resultados para que as consultas sejam rápidas.

### **Passo 3.5: Testar o Cubo**

* **Consulta e Visualização:** – Use a ferramenta OLAP ou uma ferramenta de visualização (como o Power BI ou Tableau) para conectar-se ao cubo e testar consultas.  
   – Por exemplo, veja a média de avaliação dos filmes por ano ou por gênero.
* **Verifique se os resultados estão coerentes** com o esperado. Caso contrário, revise as definições do cubo e os relacionamentos.

## **4. Exemplo Simplificado (Ilustração Visual)**

Imagine o cubo OLAP como um bloco onde cada lado é uma dimensão:

[Dimensão Data]

↑

| (hierarquia: Dia → Mês → Ano)

|

[Dimensão Filme] ←→ [Tabela de Fatos: Avaliações] ←→ [Dimensão Realizador]

(Informações do filme) (Média, Número de Avaliações) (Diretor)

↑

| (outras dimensões: Ator, Gênero, Plataforma, Demografia)

Cada “fatia” do cubo pode ser explorada em diferentes combinações – por exemplo, visualizar a média de avaliação dos filmes (fato) filtrando por ano e gênero (dimensões).

## **5. Resumo**

1. **Selecione a Ferramenta OLAP:** Escolha a tecnologia que irá usar (SSAS, Mondrian, etc.).
2. **Defina as Dimensões e Medidas:** Utilize seu modelo dimensional existente.
3. **Configure o Cubo:** Importe as tabelas, defina hierarquias, relações e agregações.
4. **Processe e Teste:** Gere o cubo e valide as consultas para garantir que ele responda às perguntas do gestor.

Esse é o caminho para implementar o Cubo OLAP, que faz parte da etapa de análise e facilita consultas multidimensionais rápidas e interativas. Se precisar de mais detalhes sobre algum dos passos ou de exemplos específicos (como código SQL ou configurações em uma ferramenta específica), estou à disposição!

Ótima abordagem! Vamos focar **apenas** na **Matriz em Bus** agora e garantir que tudo esteja bem definido antes de avançarmos para a próxima etapa. Vou te guiar **passo a passo** para construir a matriz corretamente.

# **📌 Passos para Criar a Matriz em Bus do Data Warehouse**

### **🔷 O que é a Matriz em Bus?**

A **Matriz em Bus** é um **planejamento estratégico** do Data Warehouse. Ela ajuda a organizar **os processos de negócio** e as **dimensões** envolvidas, garantindo que os dados sejam reutilizáveis e bem estruturados.

## **✅ Passo 1: Definir os Processos de Negócio**

Os **processos de negócio** representam as análises que os gestores querem fazer.  
 Para o nosso Data Warehouse de **avaliação de filmes**, os principais processos são:

1️⃣ **Avaliação de Filme** 📽️  
 ➡ Média de avaliação dos filmes (exemplo: 4.5 estrelas no IMDb).

2️⃣ **Avaliação por Realizador** 🎬  
 ➡ Qual a média de avaliações dos filmes de cada realizador (diretor)?

3️⃣ **Avaliação por Ator** 🎭  
 ➡ Qual a média de avaliações dos filmes estrelados por cada ator?

4️⃣ **Avaliação por Gênero** 🍿  
 ➡ Quais gêneros têm as melhores avaliações?

5️⃣ **Avaliação por Plataforma** 💻  
 ➡ Comparação da média de avaliações em diferentes plataformas (exemplo: Netflix vs. Amazon Prime).

6️⃣ **Avaliação por Demografia** 🌍  
 ➡ Como a avaliação de filmes varia entre diferentes países, faixas etárias e gêneros?

7️⃣ **Avaliação ao Longo do Tempo** 🕒  
 ➡ Como as avaliações mudaram ao longo dos anos e meses?

🎯 **Objetivo:** Definir **todos os processos principais** que serão analisados.

## **✅ Passo 2: Listar as Dimensões**

As **dimensões** representam os **atributos descritivos** dos fatos e ajudam a organizar os dados.

### **Dimensões para nosso Data Warehouse:**

📌 **Filme** 🎬 ➝ Nome, Ano de Lançamento, Duração, Orçamento, Receita  
 📌 **Realizador (Diretor)** 👨‍🎬 ➝ Nome, Nacionalidade, Idade  
 📌 **Ator** 🎭 ➝ Nome, Nacionalidade, Idade  
 📌 **Data** 🕒 ➝ Dia, Mês, Ano, Trimestre, Feriado, Estação do Ano  
 📌 **Gênero** 🍿 ➝ Ação, Comédia, Drama, etc.  
 📌 **Plataforma** 💻 ➝ Netflix, Disney+, Prime Video, etc.  
 📌 **Demografia** 🌍 ➝ País, Região, Faixa Etária, Gênero, Idioma

🎯 **Objetivo:** Ter um conjunto de **dimensões reutilizáveis** que possam ser usadas em vários processos de negócio.

## **✅ Passo 3: Criar a Matriz em Bus**

Agora vamos construir a **tabela cruzada** da Matriz em Bus. Ela indica **quais dimensões** estão ligadas a cada **processo de negócio**.

| **Processo de Negócio** | **Data** | **Filme** | **Realizador** | **Ator** | **Gênero** | **Plataforma** | **Demografia** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Avaliação de Filme | ✅ | ✅ | ✅ | ✅ | ✅ | ✅ | ✅ |
| Avaliação por Realizador | ✅ | ✅ | ✅ | ❌ | ❌ | ✅ | ✅ |
| Avaliação por Ator | ✅ | ✅ | ❌ | ✅ | ❌ | ✅ | ✅ |
| Avaliação por Gênero | ✅ | ✅ | ❌ | ❌ | ✅ | ✅ | ✅ |
| Avaliação por Plataforma | ✅ | ✅ | ❌ | ❌ | ❌ | ✅ | ✅ |
| Avaliação por Demografia | ✅ | ✅ | ❌ | ❌ | ❌ | ✅ | ✅ |
| Avaliação ao Longo do Tempo | ✅ | ✅ | ✅ | ✅ | ✅ | ✅ | ✅ |

🔹 **Como ler essa matriz?**

* Se um **✅** está presente, significa que a dimensão é usada naquele processo.
* Se um **❌** está presente, significa que a dimensão **não é necessária** para aquele processo.

🎯 **Objetivo:** Criar **um padrão para a modelagem** do Data Warehouse, garantindo **consistência e reutilização de dimensões**.

## **✅ Passo 4: Validar e Ajustar**

Depois de construir a matriz, precisamos validá-la e garantir que:

✔ **Todos os processos de negócio importantes estão cobertos**.  
 ✔ **As dimensões são reutilizadas corretamente** sem redundâncias.  
 ✔ **Nenhuma dimensão essencial foi esquecida**.

Se encontrarmos um **problema**, podemos ajustar a matriz antes de avançar para a modelagem dimensional.

## **📌 Resumo da Etapa da Matriz em Bus**

✅ **Definimos os processos de negócio** (o que queremos analisar).  
 ✅ **Listamos as dimensões** (como descreveremos os dados).  
 ✅ **Construímos a matriz em Bus** para estruturar o Data Warehouse.  
 ✅ **Validamos e ajustamos** para evitar erros futuros.

🎯 **Agora a modelagem do Data Warehouse está bem planejada e podemos avançar para a implementação das tabelas dimensionais e de fatos!** 🚀

## **⏭️ Próximo Passo**

Agora que temos a Matriz em Bus bem definida, podemos avançar para a **implementação do modelo dimensional**.

🔹 **Você quer revisar ou ajustar algo na Matriz em Bus antes de seguirmos para a próxima etapa?** 😊

INFORMAÇÕES DO CHAT GPT

Imagine que sua tabela de fatos é como um grande caderno onde você anota as avaliações dos filmes com duas chaves principais: uma que identifica o filme e outra que identifica a data da avaliação. Essa tabela tem colunas como:

* **Chave do Filme** (referenciando a dimensão Filme)
* **Chave da Data** (referenciando a dimensão Data)
* **Avaliação** (a média de estrelas, por exemplo)
* **Número de Avaliações** (quantas pessoas avaliaram o filme naquele período)

Agora, para incluir outras informações importantes como as dos **Realizadores**, **Atores** e **Demografia**, você pode fazer o seguinte:

### **1. Inclusão de Chaves Estrangeiras na Tabela de Fatos**

Você pode ampliar a tabela de fatos para incluir colunas adicionais que são chaves estrangeiras para as dimensões correspondentes. Por exemplo, sua tabela de fatos poderia ficar assim:

* **Chave do Filme**
* **Chave da Data**
* **Chave do Realizador**
* **Chave do Ator Principal**
* **Chave da Demografia**
* **Avaliação**
* **Número de Avaliações**

Com essas chaves adicionais, cada registro da tabela de fatos pode ser “ligado” às informações detalhadas de cada dimensão. Por exemplo:

* **Chave do Realizador:** permite saber quem dirigiu o filme e agrupar as avaliações por realizador.
* **Chave do Ator Principal:** permite analisar se filmes com certos atores têm avaliações melhores ou piores.
* **Chave da Demografia:** se você tiver informações sobre os avaliadores (como faixa etária, localização ou outros dados demográficos), poderá entender como diferentes grupos avaliam os filmes.

### **2. Vantagens Dessa Abordagem**

* **Facilidade na Análise:** Ao incluir essas chaves, você pode fazer consultas que cruzam informações de várias dimensões. Por exemplo, "Qual realizador tem os filmes mais bem avaliados entre o público jovem?" ou "Filmes com determinado ator recebem avaliações diferentes em várias regiões?"
* **Flexibilidade:** Essa abordagem permite que você combine e agregue dados conforme a necessidade do gestor. Você pode filtrar por data, gênero do filme, região dos avaliadores, etc.

### **3. Implementação**

* **Crie as Tabelas de Dimensões Separadas:**
  + **Dimensão Filme:** com atributos como título, gênero, ano de lançamento, etc.
  + **Dimensão Realizador:** com informações sobre o diretor.
  + **Dimensão Ator:** com dados do ator principal (ou até de todo o elenco, se for o caso).
  + **Dimensão Demografia:** com atributos como faixa etária, gênero do avaliador, localização, etc.
  + **Dimensão Data:** com todas as informações detalhadas do calendário.
* **Crie a Tabela de Fatos Expandida:**
  + Adicione colunas para cada uma das chaves (ID) dessas dimensões. Assim, o registro que representa uma avaliação de um filme conterá não só a média de estrelas e o número de avaliações, mas também referências a quem dirigiu, qual ator estrelou e qual é o perfil dos avaliadores.

### **Resumindo com um Exemplo Simples**

Imagine que você tem um registro na tabela de fatos assim:

| **Chave Filme** | **Chave Data** | **Chave Realizador** | **Chave Ator** | **Chave Demografia** | **Avaliação** | **Número de Avaliações** |
| --- | --- | --- | --- | --- | --- | --- |
| 101 | 365 | 55 | 200 | 12 | 4.5 | 1000 |

Isso quer dizer que:

* O filme com **ID 101** foi avaliado.
* A avaliação aconteceu na data com **ID 365** (por exemplo, 31 de dezembro de 2010).
* O filme foi dirigido pelo realizador com **ID 55**.
* O ator principal tem **ID 200**.
* Os avaliadores pertencem à demografia com **ID 12** (por exemplo, jovens adultos de um certo país).
* A média de estrelas foi 4.5, com 1000 votos.

Com essa estrutura, quando o gestor perguntar "Qual diretor tem filmes com melhor avaliação entre jovens adultos?" você pode cruzar a Tabela de Fatos com a Dimensão Realizador e a Dimensão Demografia para responder.

Essa é uma forma simples de como incluir as dimensões adicionais e como elas se conectam à tabela de fatos, permitindo análises ricas e úteis para a tomada de decisão.