

# Fase 2 - Matriz em Bus

## Licenciatura em Engenharia Informática

Integração e Processamento Analítico de Informação  
2023/2024



UNIVERSIDADE DE ÉVORA  
ESCOLA DE CIÊNCIAS E TECNOLOGIA

DEPARTAMENTO DE INFORMÁTICA



---

**Bárbara Loureiro (48469) e Gustavo Gomes (48392)**

Docente Carlos Pampulim Caldeira

*2ª Fase do Trabalho - Matriz em Bus, desenvolvido no âmbito da unidade curricular de Integração e Processamento Analítico de Informação da Licenciatura em Engenharia Informática.*

Évora, 17 de abril de 2024

---

# Tabela de conteúdos

<b>1 Introdução.....</b>	<b>1</b>
<b>2 Desenho do Modelo Dimensional.....</b>	<b>2</b>
2.1 Escolha do Processo de Negócio.....	2
2.1.1 Perguntas Analíticas.....	2
2.2 Declaração do Grão do Processo de Negócio.....	3
2.3 Escolha das Dimensões.....	3
2.3.1 Dimensão Ator.....	4
2.3.2 Dimensão Realizador.....	4
2.3.3 Dimensão Filme.....	4
2.3.4 Dimensão Género.....	5
2.3.4 Dimensão Papel.....	5
2.3.1 Dimensão Data.....	6
2.4 Identificação dos Factos e Medidas.....	6
2.5 Outras Dimensões.....	8
2.5.1 Dimensões Role-Playing.....	8
2.5.2 Dimensões em Floco de Neve e Dimensão Outrigger.....	10
<b>3 A Arquitetura em Bus.....</b>	<b>11</b>
3.1 Matriz em Bus.....	11
3.2 Matriz em Bus Analítica.....	12
<b>4 Conclusão.....</b>	<b>13</b>
<b>Referências.....</b>	<b>14</b>

---

# 1 Introdução

Os dados da indústria cinematográfica são importantes para perceber melhor esse meio, e perceber como podem ser uma ferramenta valiosa para o negócio da indústria.

O trabalho desta unidade curricular envolve o desenho e a construção de uma *data warehouse* funcional e eficiente, no âmbito dessa mesma indústria .

As quatro decisões chave feitas para o desenho do modelo dimensional, baseadas nas técnicas de modelação dimensional de Kimball, foram: [1]

1. Escolha do Processo de Negócio
2. Declaração do Grão do Processo de Negócio
3. Escolha das dimensões
4. Identificação dos Factos e Medidas

Inicialmente, a escolha do processo de negócio foi estabelecida para delinear o objetivo da análise, com perguntas analíticas específicas formuladas para direcionar o desenvolvimento do modelo. Com o processo de negócio definido, a declaração do grão foi determinada para orientar o que cada linha da tabela de factos representa, garantindo a consistência entre dimensões e factos.

Após definir o grão, as dimensões foram seleccionadas para fornecer o contexto necessário ("quem, o quê, onde, quando, por que e como") para cada evento de negócio presente na tabela de factos. A tabela de factos e as tabelas de dimensões foram apresentadas num esquema em estrela, a abordagem de modelação pedida na Fase 1 do trabalho.

O critério de sucesso no nosso trabalho, se um filme tem sucesso ou não na indústria, baseia-se no ranking do IMDB. Não vamos fazer distinção entre avaliação da comunidade em geral e avaliação de especialistas da área. Também, como referido mais tarde na 2.3.3 Dimensão Filme, neste trabalho vamos supor que um filme só pode ter um género.

No desenho do nosso modelo dimensional, verificámos também a existência de dimensões *Role-Playing*, e de possíveis dimensões *Outrigger* caso adoptado as dimensões em floco de neve. Por outro lado, não identificamos nenhuma mini dimensões, nem dimensões lixo.

A arquitetura em *bus* foi detalhada por meio de duas versões da matriz em *bus* e de uma matriz em *bus* analítica. Onde, os processos de negócio foram alinhados com as dimensões, facilitando o desenho e a comunicação da arquitetura do *data warehouse*.

Tal como na Fase 1, o desenvolvimento das tabelas e agora das matrizes foi realizado no Google Sheets, disponível em [48392\\_48469\\_IPAI](#)

---

## 2 Desenho do Modelo Dimensional

Para o desenho do Modelo Dimensional é necessário haver uma compreensão das necessidades do negócio, bem como da origem dos dados, isso permite uma melhor tomada de decisão no seu desenvolvimento. Num caso real, teríamos de ser bastante minuciosos com a avaliação da viabilidade dos dados. [1]

No nosso caso de estudo, os dados são fictícios, criados manualmente. Apesar de o negócio se basear na realidade da indústria cinematográfica, o nosso conhecimento sobre a indústria é limitado. Assim, o nosso objetivo no desenho do modelo dimensional é ser o mais simples e de fácil compreensão, tanto para nós desenvolvedores, como para possíveis utilizadores.

### 2.1 Escolha do Processo de Negócio

Processos de Negócios são atividades operacionais conduzidas pelas organizações. A sua escolha é importante pois define um objetivo específico e permite que o grão do negócio, as dimensões e os factos sejam declarados. [1]

Todos sabemos, que na indústria cinematográfica, o sucesso de um filme não é linear. Muitas vezes um roteiro é muito bom, mas os atores do elenco são pouco conhecidos e o filme acaba por não ser um sucesso ou precisamente o oposto, os atores são bastantes conhecidos e o filme acaba por ser um sucesso. Ou seja, os atores acabam por ser um dos principais agentes neste negócio

O foco do nosso processo de negócio vai abranger exatamente isso, o **acompanhamento do desempenho do ator** na indústria cinematográfica. A análise deste processo de negócio consiste nomeadamente em estudar o que acontece nos diferentes parâmetros, por exemplo tipo de papel do ator, duração e género do filme em que o ator participa.

#### 2.1.1 Perguntas Analíticas

Apontamos algumas perguntas para nos ajudar a compreender melhor o nosso processo de negócio:

- Quantos filmes de comédia o ator "X" já fez?
- Em quantas curtas metragens o ator "Y" já participou? E em quantas longas?
- Qual é o total de filmes que a atriz "B" realizou durante o seu primeiro ano de carreira?
- Qual é a média das avaliações do IMDB da atriz "Z" quando o seu papel é de atriz principal de um filme?
- Quantos filmes o ator "P" fez que foram realizados pelo realizador "M"?

---

## 2.2 Declaração do Grão do Processo de Negócio

A declaração do Grão do Processo de Negócio é uma etapa fundamental para o desenho do modelo dimensional. O grão estabelece o que cada linha da tabela de factos representa. Deve ser declarado antes de escolher dimensões ou factos porque cada dimensão ou facto candidato deve ser consistente com a granularidade. [1]

O grão do nosso processo de negócio é **a participação de cada ator em cada filme**. Os factos com este grão podem ser utilizados para responder às perguntas realizadas no 2.1.1 e podem ter níveis de detalhe diferentes, dependendo dos parâmetros a serem estudados:

*“Quantos filmes de comédia o ator “X” já fez?”*

*“Qual é a média das avaliações do IMDB da atriz “Z” quando o seu papel é de atriz principal de um filme?”*

Respostas hipotéticas às perguntas, respetivamente:

*“O ator “X” já fez 40 filmes de comédia.”*

*“A atriz “Z” tem uma média de avaliação de 7.2 no IMDB nos seus filmes, quando é a atriz principal.”*

Ou seja, podemos identificar por exemplo que atores estão mais ativos ou têm mais sucesso num determinado género de filme, ou, identificar atores que estão cada vez mais em ascensão na indústria, onde os seus filmes estão a ter bastante sucesso por consequência.

## 2.3 Escolha das Dimensões

As dimensões fornecem o contexto “quem, o quê, onde, quando, por que e como” que envolve um evento de um processo de negócio. As tabelas de dimensões contêm os atributos descritivos usados pelas aplicações de BI (*Business Intelligence*) para filtrar e agrupar os factos. [1]

O Ator, o Realizador, o Filme, o Género e a Data, foram as cinco dimensões identificadas, de acordo com os atributos definidos para a tabela de factos.

---

### 2.3.1 Dimensão Ator

Como podemos observar na Tabela 1, a Chave e os Atributos da Dimensão Ator e a descrição de cada um deles.

Coluna	Descrição
Chave do Ator	Identificador único para cada ator
Nome do Ator	Nome completo do ator
Nome Artístico	Nome artístico do ator
Data de Nascimento	Data de nascimento do ator
Data de Falecimento	Data de falecimento do ator
Sexo	Sexo do ator
Nacionalidade	Nacionalidade do ator
Ano de Início de Carreira	Ano em que o ator iniciou sua carreira no cinema

Tabela 1 - Dimensão Ator

### 2.3.2 Dimensão Realizador

Como podemos observar na Tabela 2, a Chave e os Atributos da Dimensão Realizador e a descrição de cada um deles.

Coluna	Descrição
Chave do Realizador	Identificador único para cada realizador
Nome do Realizador	Nome completo do realizador
Data de Nascimento	Data de nascimento do realizador
Data de Falecimento	Data de falecimento do realizador
Sexo	Sexo do realizador
Nacionalidade	Nacionalidade do realizador
Ano de Início de Carreira	Ano em que o diretor iniciou sua realizador

Tabela 2 - Dimensão Realizador

### 2.3.3 Dimensão Filme

Como podemos observar na Tabela 3, a Chave e os Atributos da Dimensão Filme e a descrição de cada um deles.

Na dimensão Filme, a Duração (em minutos) divide os filmes em duas categorias. As curtas metragens, se a Duração for inferior a 60 minutos e longas metragens, se a Duração for igual ou superior a 60 minutos.

---

No nosso caso de estudo, apesar de não corresponder à realidade, vamos assumir que um filme só pode ter um género.

Coluna	Descrição
Chave do Filme	Identificador único para cada filme
Título do Filme	Nome do filme
Ano de Lançamento	Ano em que o filme
Duração	Duração do filme em minutos
Classificação Indicativa	Idade mínima recomendada para o público
Idioma Original	Idioma original do filme
Género	Tipo de género do filme
Realizador Principal	Nome do realizador principal do filme
Elenco Principal	Nomes dos atores principais do filme
Avaliação do IMDB	Avaliação do filme no site da IMDB

Tabela 3 - Dimensão Filme

### 2.3.4 Dimensão Género

Como podemos observar na Tabela 4, a Chave e os Atributos da Dimensão Género e a descrição de cada um deles.

Coluna	Descrição
Chave do Género	Identificador único para cada género
Nome do Género	Nome completo do género
Descrição do Género	Breve descrição do tipo de género do filme

Tabela 4 - Dimensão Género

### 2.3.4 Dimensão Papel

Como podemos observar na Tabela 5, a Chave e os Atributos da Dimensão Papel e a descrição de cada um deles.

A Dimensão Papel é referente aos papéis que os atores desempenham em cada filme que participam, assim é possível detalhar o tipo de participação de um determinado ator, num determinado filme.

Coluna	Descrição
Chave do Papel	Identificador único para cada tipo de papel
Nome do Papel	Nome indicador da importância do papel do ator no filme, como principal, secundário, figurante, etc...
Descrição do Papel	Descrição do papel, características da personagem

Tabela 5 - Dimensão Papel

---

---

### 2.3.1 Dimensão Data

Como podemos observar na Tabela 6, a Chave e os Atributos da Dimensão Data e a descrição de cada um deles.

Coluna	Descrição
Chave da Data	Identificador único para cada data
Data Completa	Data no formato AAAA-MM-DD
Ano	Ano da data
Semestre	Semestre (1 ou 2)
Trimestre	Trimestre (1 a 4)
Mês	Mês (1 a 12)
Nome do Mês	Nome do mês (Janeiro a Dezembro)
Dia	Dia do mês (1 a 31)
Dia da Semana	Dia da semana (Segunda a Domingo)
Nome do Dia	Nome do dia da semana
Semana do Ano	Semana do ano (1 a 52)
Dia do Ano	Dia do ano (1 a 365/366)
Feriado	Indica se é um feriado
É Final de Semana	Indica se é final de semana
É Dia Útil	Indica se é dia útil
Quinzena	Quinzena do mês (1 ou 2)
Época do Ano	Época do ano (Primavera, Verão, Outono, Inverno)
Ano Fiscal	Ano fiscal
Trimestre Fiscal	Trimestre fiscal (1 a 4)
Mês Fiscal	Mês fiscal (1 a 12)
Semana Fiscal	Semana fiscal (1 a 52)
Início da Semana	Data de início da semana
Fim da Semana	Data de fim da semana
Início do Mês	Data de início do mês
Fim do Mês	Data de fim do mês
Início do Trimestre	Data de início do trimestre
Fim do Trimestre	Data de fim do trimestre
Início do Ano	Data de início do ano
Fim do Ano	Data de fim do ano

Tabela 6 - Dimensão Data

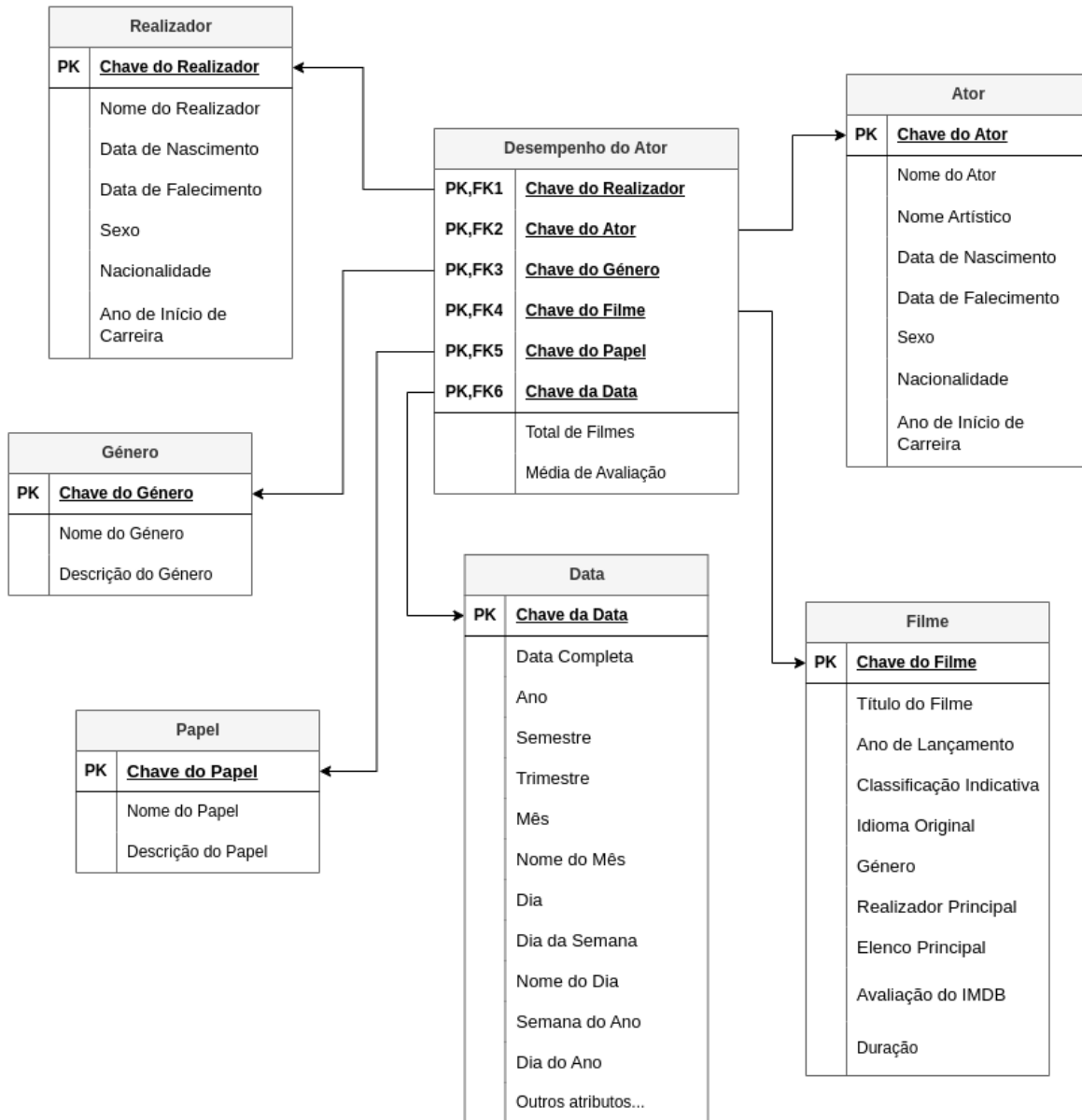
### 2.4 Identificação dos Factos e Medidas

Os factos são as medidas que resultam de um evento de um processo de negócio e são quase sempre numéricos. Uma única linha da tabela de factos tem um relacionamento um-para-um com um evento de medição, conforme descrito pela granularidade da tabela de factos. [1]

As tabelas de factos dividem-se em três tipos, as transaccionais, as de sumarização periódica e as de sumarização acumulada. Os factos que são medidos em vários pontos discretos ao longo de um período de tempo podem ser registados em tabelas transaccionais. Como é o reservatório central de cada *data mart*, este tipo de tabelas é a mais comum no modelo dimensional. [2]



O Esquema em Estrela demonstrado no Esquema 1, liga as tabelas de dimensão do no nosso processo de negócio à tabela de factos do mesmo, o Desempenho do Ator.



Esquema 1 - Tabelas de dimensão e factos num Esquema em Estrela para análise do Desempenho do Ator

---

Os elementos principais desta tabela de factos, pertencente ao Esquema 1, são:

- **Chave de Realizador:** Uma chave estrangeira que liga a tabela de factos à dimensão Realizador.
- **Chave do Ator:** Uma chave estrangeira que liga a tabela de factos à dimensão Ator.
- **Chave do Género:** Uma chave estrangeira que liga a tabela de factos à dimensão Género, permitindo análises baseadas no tipo de filme.
- **Chave do Filme:** Uma chave estrangeira que conecta cada registo de desempenho a um filme específico.
- **Chave do Papel:** Uma chave estrangeira que indica o papel específico desempenhado pelo ator no filme.
- **Chave da Data:** Uma chave estrangeira que fornece a data específica relacionada ao desempenho, que pode ser a data de lançamento do filme ou a data de início de carreira do ator.
- **Total de Filmes:** Uma métrica que indica o número total de filmes em que um ator participou, pode ser filtrado por algum tipo de género de filme ou produtor do filme.
- **Média de Avaliação:** Uma métrica que reflete a avaliação média dos filmes do ator de acordo com o IMDB, pode ser filtrado também por algum tipo de género de filme ou por um papel específico que o ator tenha desempenhado nesses filmes.

O facto Total de Filmes é aditivo e o facto Média de Avaliação é semi-aditivo. Os factos aditivos podem ser somados em relação a todas as dimensões de um esquema em estrela e, os semi-aditivos só podem ser somados em relação a algumas dimensões. [2]

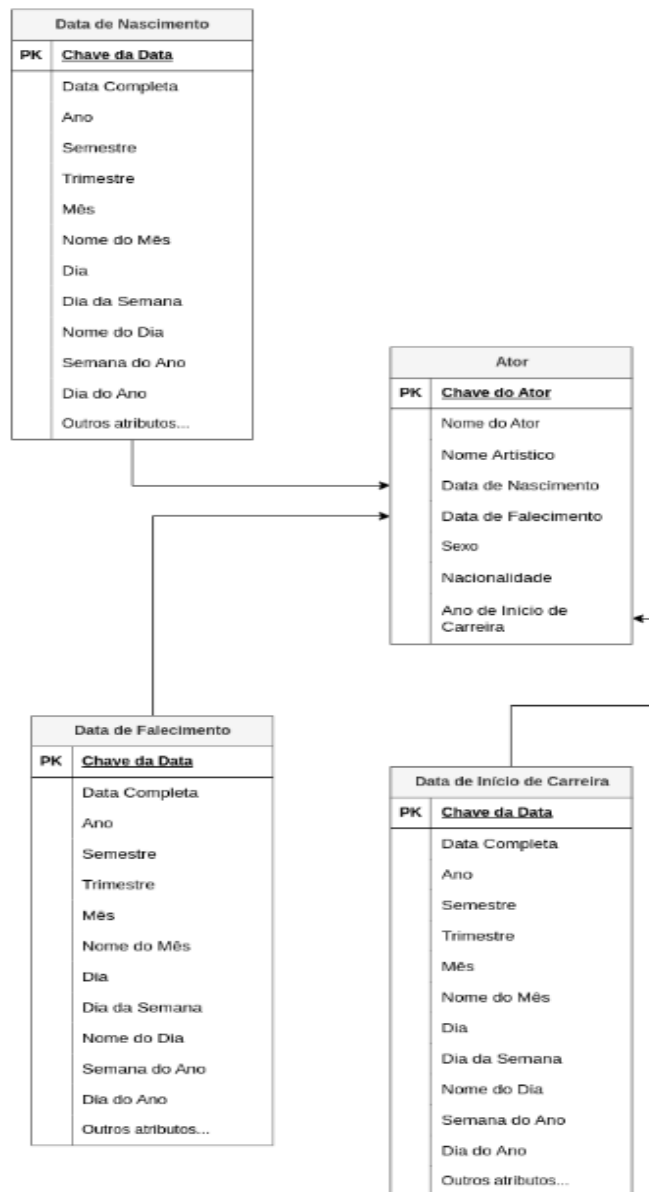
Cada linha na tabela representa um evento único de desempenho, como uma participação num filme, e as métricas associadas a esse evento. Logo podemos concluir que se trata de uma tabela de facto transaccional pois armazena uma linha por cada acontecimento.

## 2.5 Outras Dimensões

### 2.3.1 Dimensões Role-Playing

Uma dimensão *role-playing* é uma dimensão que é usada várias vezes num modelo dimensional, mas com referências diferentes. Por exemplo, uma tabela de factos, pode incluir várias datas, cada uma representada por uma chave estrangeira pela dimensão Data. Para garantir independência entre referências, cada chave estrangeira deve estar relacionada com uma representação diferente da dimensão Data. [1]

Como acontece no Esquema 1, a dimensão Data está a ser usada para diferentes propósitos (como a data de lançamento de um filme e a data de início de carreira de um ator, data de nascimento de um realizador, entre outras...), logo ela seria uma dimensão *role-playing*. No Esquema 2 podemos ver uma possível representação da dimensão *role-playing* da Data.



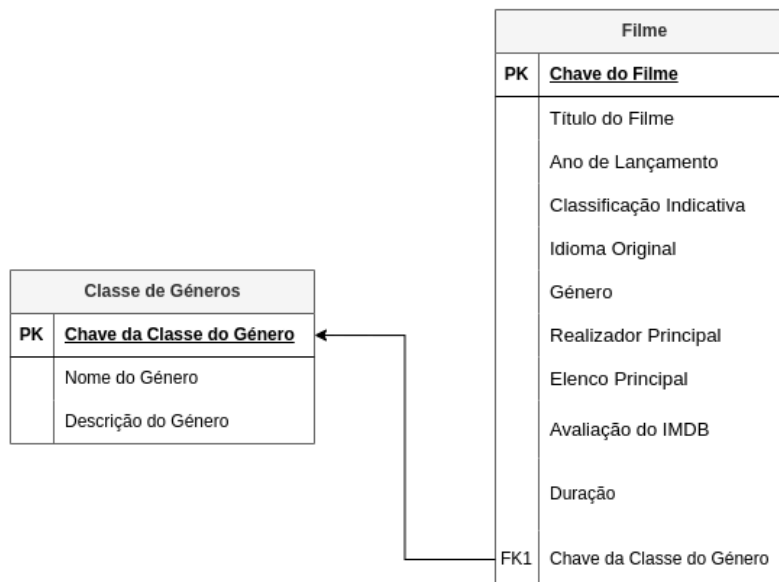
Esquema 2 - Possível representação da dimensão *role-playing* da Data

---

## 2.5.2 Dimensões em Floco de Neve e Dimensão Outrigger

Ao ato de normalizar o modelo dimensional de dados dá-se o termo *snowflaking*. Quando uma hierarquia numa tabela de dimensão é normalizada, os atributos de baixa cardinalidade aparecem como tabelas secundárias conectadas à tabela de dimensão base através dos seus atributos. Apesar do modelo em floco de neve apresentar hierarquias com precisão, é um modelo a evitar, pois tende a ser mais complexo, mais difícil de usar e por consequência com uma manutenção mais complexa. [1,2]

Num esquema em floco de neve, a dimensão Género poderia ser usada como *outrigger*. Visto que, na 2.3.3 Dimensão Filme, assumimos que um filme só teria apenas um género. Assim, essa dimensão pode ser normalizada fora da dimensão Filme, como podemos verificar no Esquema 3.



Esquema 3 - A tabela Filme ligada à dimensão *outrigger* que possibilita a repartição automática dos filmes por classes de géneros

---

## 3 A Arquitetura em Bus

A arquitetura em *bus* do *data warehouse* empresarial fornece uma abordagem incremental para construir o sistema DW/BI empresarial. Esta arquitetura decompõe o processo de planeamento do sistema em partes que se possam gerir, focam-se nos processos de negócios e, ao mesmo tempo, fornecem integração por meio de dimensões padrão que são utilizadas ao longo de todos os processos. [1]

### 3.1 Matriz em Bus

A matriz em *bus* do *data warehouse* empresarial é a ferramenta essencial para desenhar e comunicar a arquitetura em *bus* do *data warehouse* empresarial. As linhas da matriz são os processos de negócio e as colunas as dimensões. As células assinaladas da matriz indicam se uma dimensão está associada a um processo de negócio. [1]

Um dos principais objetivos do *data warehouse* é ser compreensível e simples aos seus utilizadores, como podemos ver na Figura 1, a matriz em bus sobre o processo Desempenho do Ator cumpre esse requisito.

Dimensões							
	Ator	Realizador	Género	Filme	Data	Papel	
Processos							Dimensionalidade
Desempenho do Ator	✓	✓	✓	✓	✓	✓	6

Figura 1 - Matriz em *bus* sobre o processo Desempenho do Ator

Enquanto a Figura 1 nos mostra uma matriz inicial, a Figura 2 apresenta-se como uma matriz mais detalhada e com uma análise mais profunda do negócio. O processo de negócio apesar de ser só um, está sujeito a vários tipos diferentes de análise. Estas análises permitem-nos responder às perguntas lançadas inicialmente em 2.1.1 Perguntas Analíticas.

Dimensões	Ator	Realizador	Género	Filme	Data	Papel	
Processos							Dimensionalidade
Participação do Ator num Filme	✓	✓	✓	✓	✓	✓	6
Participação do Ator num Filme de Comédia	✓	✓	✓	✓	✓	✓	6
Participação do Ator num Filme com o papel de ator principal	✓	✓	✓	✓	✓	✓	6

Figura 2 - Matriz em *bus* com linhas dedicadas à granularidade das tabelas de factos

### 3.2 Matriz em Bus Analítica

A matriz em bus analítica, apresentada na Figura 3, é uma matriz com informação mais extensa, inclui o tipo de tabela e o respetivo grão. No entanto, esta matriz não deve substituir a matriz elementar devido à sua maior complexidade. Fica condicionada à fase de planeamento da construção física dos *data marts*. [2]

Dimensões			Ator	Realizador	Género	Filme	Data	Papel	
Processos	Tipo	Grão							Dimensionalidade
<b>Desempenho do Ator</b>									
Participação do Ator num Filme	Transaccional	Participação: ator; filme; data	✓	✓	✓	✓	✓	✓	6
Participação do Ator num Filme de Comédia	Transaccional	Participação: ator; filme; género; data	✓	✓	✓	✓	✓	✓	6
Participação do Ator num Filme com o papel de ator principal	Transaccional	Participação: ator; filme; papel; data	✓	✓	✓	✓	✓	✓	6

Figura 3 - Matriz em *bus* analítica sobre o processo Desempenho do Ator

---

## 4 Conclusão

Nesta segunda fase o nosso foco foi discriminar melhor o processo de negócio e por consequência o seu grão. Isso permitiu-nos perceber que tínhamos uma dimensão em falta: a dimensão Papel, essencial para o nosso processo de negócio. Todas as dimensões acabaram por sofrer melhorias, o que nos possibilitou esquematizar com uma facilidade maior o esquema em estrela.

Apesar de o processo de negócio estar bem detalhado, o mais desafiante neste trabalho foi detalhar os processos nas matrizes em *bus*.

Contudo, tiramos um aproveitamento positivo do trabalho pois conseguimos perceber a gravidade e a diversidade do impacto que Desempenho do Ator tem na indústria cinematográfica.

---

## Referências

- [1] [Kimball Dimensional Modeling Techniques](#) [Online; accessed 21-abril-2024].
- [2] Carlos Pampulim Caldeira. Data Warehousing. Conceitos e modelos. Edições Sílabo - 3ª Edição, 2023.