

ANÁLISE DE DADOS E BIG DATA

Aran Bey Tcholakian Morales

Saulo Popov Zambiasi

Richard Henrique de Souza

AVALIAÇÕES LEMBRETE

- A1 : 30 pontos
 - Data provável: 09/05
- A2 : 30 pontos
 - Data provável: 13/06
- A3 : 40 pontos
 - Data de entrega: 19/06

DICA

- Cursos
 - Trilha: Banco de dados
 - <https://www.ev.org.br/trilhas-de-conhecimento/banco-de-dados>
- Modelagem de Dados
- Implementando Banco de Dados
- Administrando Banco de Dados

fundação bradesco | escola virtual

Inicio Cursos Parcerias Sobre Ajuda

Acessar Cadastr

Banco de Dados

Um **banco de dados** é uma coleção organizada de informações estruturadas, normalmente armazenadas eletronicamente em um sistema de computador.

E por falar em **informação**, todos sabemos que ela tem um peso essencial hoje em dia, pois auxilia pessoas e empresas na tomada de decisões.

Duração **38h** | Cursos **3**

Mostrar mais

Quero me matricular

Conteúdo Detalhes

Administrando Banco de Dados

Neste curso, você verá uma introdução de como administrar um banco de dados e, desta forma, conseguir realizar...

Duração 15h | Nível Avançado

Implementando Banco de Dados

Neste curso, você verá uma introdução sobre a implementação de banco de dados, compreendendo suas...

Duração 15h | Nível Intermediário

Modelagem de Dados

O assunto modelagem de dados é bem relevante, principalmente, nos dias de hoje, pois o armazenamento e a...

Duração 8h | Nível Iniciante

BUSCA ATIVA

- Leitura Recomendada
- Ulife → Bibliotecas online → Minha Biblioteca
 - Vida, Edinilson da, S. et al. Data warehouse.
 - Disponível em: Minha Biblioteca, Grupo A, 2021.

BIG DATA

DATA WAREHOUSE

Edinilson da Silva Vida
Nicoli Souza Pires Alves
Rafael Gastão Coimbra Ferreira
Douglas Campos de Souza
Fabrício Felipe Melo Barbosa
Hygo Sousa de Oliveira
Luarte de Marque
Luis Gustavo Meschietto
Priscila de Fátima Gonçalves

sa
GAH



INSTITUTO
DE PESQUISA
E INOVAÇÃO

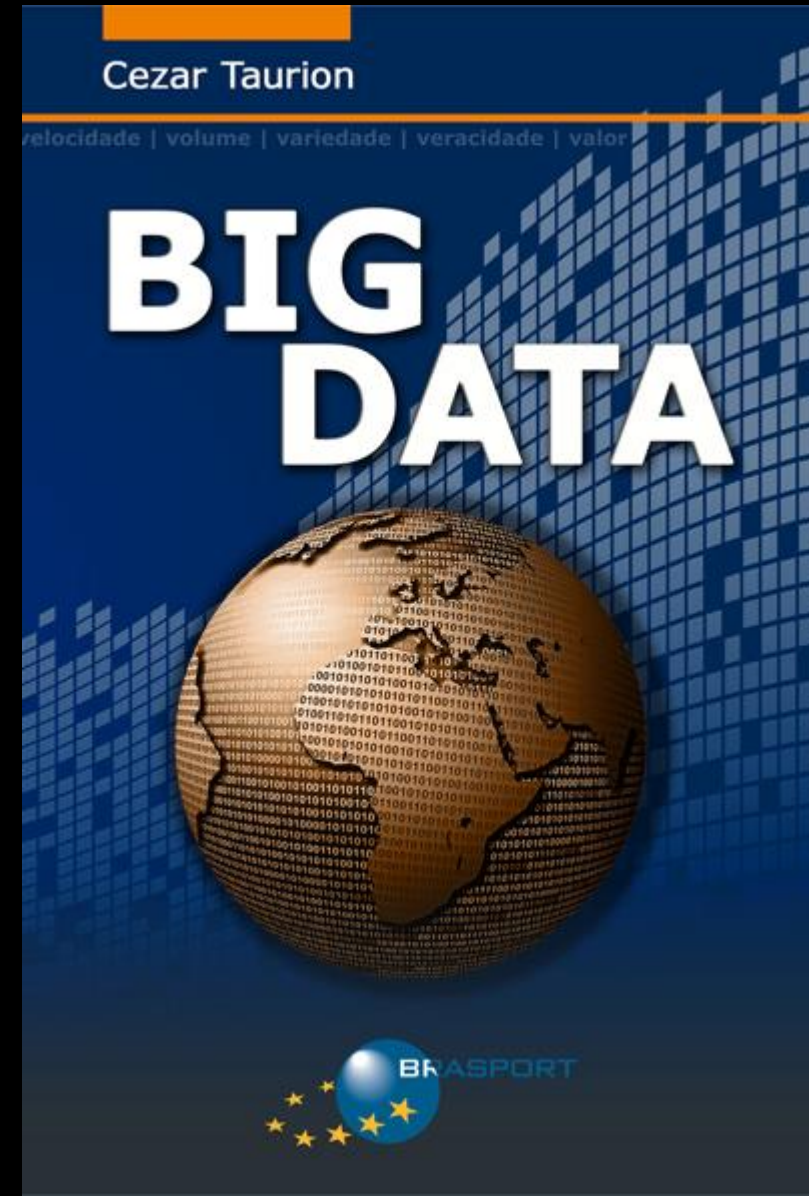
BUSCA ATIVA

- Leitura Recomendada
 - Ulife → Bibliotecas online → Biblioteca Pearson
 - Big data / Douglas Eduardo Basso.2020 (Basso, Big Data)



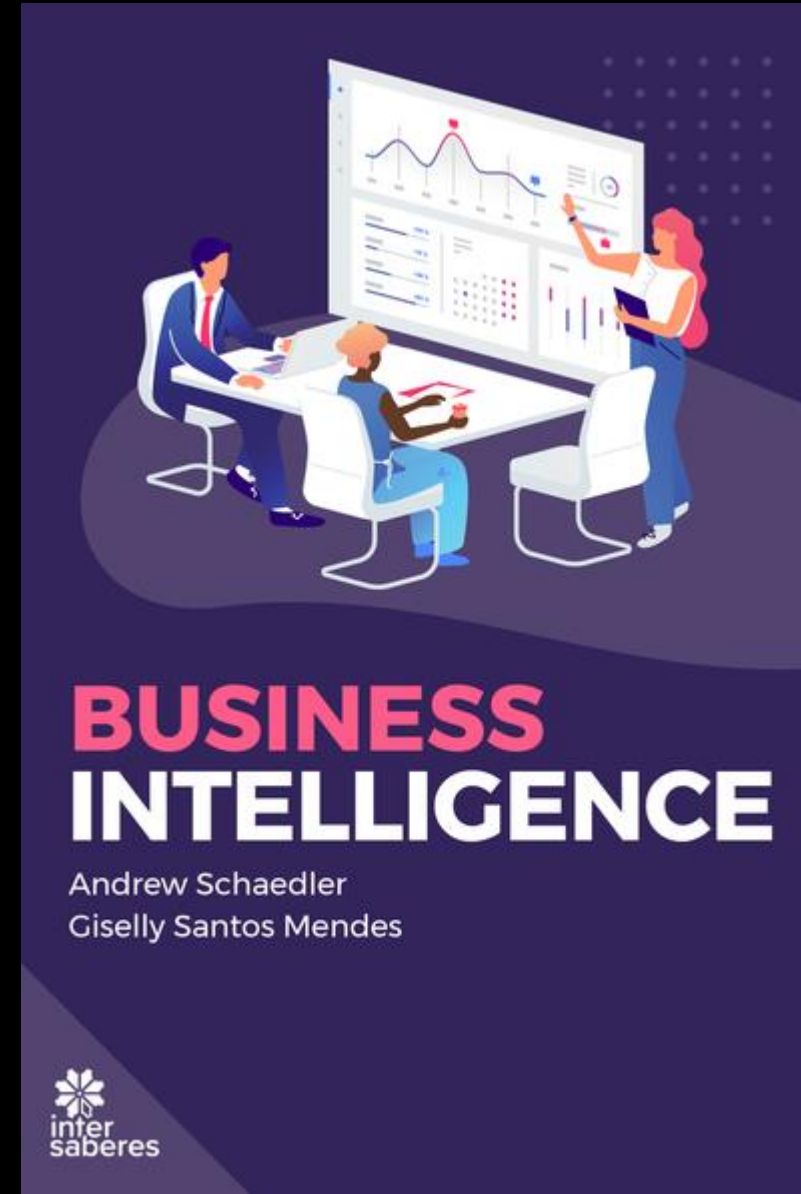
BUSCA ATIVA

- Leitura Recomendada
 - Ulife → Bibliotecas online → Biblioteca Pearson
 - aurion, CezarBig Data / Cezar Taurion - Rio de Janeiro: Brasport, 2013.ePUBISBN: 978-85-7452-608-91. Big Data 2. Negócios I. Título. (Cezar Taurion, , Big Data)



BUSCA ATIVA

- Leitura Recomendada
 - Ulife → Bibliotecas online → Biblioteca Pearson
 - Business intelligence. Giselly Santos Mendes, Andrew Schaedler. 2021.





**BIG
DATA**



EXERCÍCIO 1

- Uma empresa administradora de planos de saúde decidiu construir um sistema de BI para auxiliar os analistas de negócios na gestão da empresa. A empresa administra um total de 10 clínicas, 100 médicos e 1000 clientes. No levantamento de requisitos foram identificadas as necessidades de informação para a construção de um modelo dimensional.

O modelo consiste em 4 dimensões (tempo, médico, clínica e paciente), e das medidas *nro_consultas* (número de consultas realizadas pelo médico) e *vlr_consultas* (quantia cobrada pelo médico no número de consultas). A dimensão tempo deve ter o mês, trimestre, semestre e ano; a dimensão “médico” o nome do médico e a especialidade médica; a dimensão paciente o sexo e a faixa etária do paciente (de 0 a 17 anos; de 18 a 39 anos; de 40 a 59 anos; 60 ou mais anos), e a dimensão clínica, o nome da clínica.

Desenhe o modelo dimensional (tabela de fatos, medidas e dimensões).

DI_CARATERISTICA_PACIENTE

SEQ_CARATERISTICA_PACIENTE: INTEGER [PK]

SEXO: VARCHAR(10)

FAIXA_ANO_NASC: VARCHAR(20)

ANO_INICIO_FAIXA: INTEGER

ANO_FIM_FAIXA: INTEGER

DI_TEMPO

SEQ_TEMPO: INTEGER [PK]

MES: VARCHAR(20)

TRIMESTRE: INTEGER

SEMESTRE: INTEGER

ANO: INTEGER

FT_CONSULTAS

SEQ_TEMPO: INTEGER [PFK]

SEQ_MEDICO: INTEGER [PFK]

SEQ_CLINICA: INTEGER [PFK]

SEQ_CARATERISTICA_PACIENTE: INTEGER [PFK]

NRO_CONSULTAS: INTEGER

VRL_CONSULTAS: DOUBLE

DI_MEDICO

SEQ_MEDICO: INTEGER [PK]

NME_MEDICO: VARCHAR(60)

ESPECIALIDADE: VARCHAR(100)

DI_CLINICA

SEQ_CLINICA: INTEGER [PK]

NME_CLINICA: VARCHAR(40)



Novo Projeto

Modelo de Dados



Nova Tabela (Atalho T)



EXERCÍCIO 1

Propriedades da tabela

Nome da Tabela Lógica New_Table|

Nome da Tabela Física New_Table

Nome da chave primária New_Table_pk

Comentários

Cor da tabela

Cor do texto

Linhas tracejadas

Cantos arredondados

OK Cancelar

Propriedades da tabela

Nome da Tabela Lógica FT_CONSULTAS

Nome da Tabela Física FT_CONSULTAS|

Nome da chave primária New_Table_pk

Comentários

Cor da tabela

Cor do texto

Linhas tracejadas

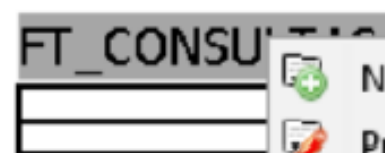
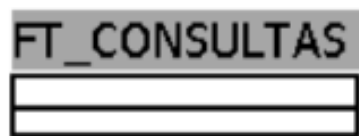
Cantos arredondados

OK Cancelar



Novo Projeto

- Modelo de Dados
 - FT_CONSULTAS (TABLE)



- Novo Índice... I
- Propriedades do índice >
- Nova Coluna... C**
- Propriedades da coluna
- Alinhar tabelas >
- Aparência da Tabela >
- Propriedades da Tabela...
- Trazer para frente
- Enviar para trás
- Excluir selecionado Excluir

Inserir Coluna (Atalho C)

Fonte para mapeamento ETL

Nada especificado

Nome Lógico

NRO_CONSULTAS

Nome Físico

NRO_CONSULTAS

 Chave primária**Tipo**

INTEGER

Data Types

- ARRAY
- BIGINT
- BINARY
- BIT
- BLOB
- BOOLEAN
- CHAR
- CLOB
- DATALINK
- DATE
- DECIMAL
- DISTINCT
- DOUBLE
- FLOAT
- INTEGER**
- JAVA_OBJECT
- LONGNVARCHAR
- LONGVARBINARY

Escala

0



0

Sequência (Sequência disponível para plataformas que usam sequências)

AS_seq

OK

Cancelar

FT_CONSULTAS

NRO_CONSULTAS: INTEGER NOT NULL

EXERCÍCIO 1

The screenshot shows the SQL Power Architect interface. The title bar reads "Novo Projeto - SQL Power Architect". The menu bar includes "Arquivo", "Editar", "Conexões", "ETL", "OLAP", "Enterprise", "Ferramentas", "Perfil", "Janela", and "Ajuda". The toolbar contains icons for file operations, refresh, print, SQL, and other functions.

The left sidebar shows a tree view with the following structure:

- Novo Projeto
- Modelo de Dados
 - FT_CONSULTAS (TABLE)**
 - Columns folder for FT_CONSULTAS
 - NRO_CONSULTAS: INTEGER
 - VRL_CONSULTAS: DOUBLE**
 - Exported keys folder for FT_CONSULTAS
 - Imported keys folder for FT_CONSULTAS
 - Indices folder for FT_CONSULTAS

The main workspace displays the table definition for **FT_CONSULTAS** in a structured format:

FT_CONSULTAS	
NRO_CONSULTAS: INTEGER	NOT NULL
VRL_CONSULTAS: DOUBLE	NOT NULL

DI_MEDICO

NME_MEDICO: VARCHAR(100) NOT NULL
ESPECIALIDADE: VARCHAR(50) NOT NULL

FT_CONSULTAS

NRO_CONSULTAS: INTEGER NOT NULL
VRL_CONSULTAS: DOUBLE NOT NULL

Propriedades da coluna New Column

Fonte para mapeamento ETL

Nada especificado

Nome Lógico

SEQ_MEDICO

Nome Físico

SEQ_MEDICO

Chave primária

Tipo

INTEGER

Precisão

0

Escala

0

Permite nulos

No

Auto Incremento

Yes

Valor Padrão

Nome da Sequência (Apenas aplicável para plataformas que usam sequências)

DI_MEDICO_SEQ_MEDICO_seq

Comentários

OK

Cancelar

EXERCÍCIO 1

DI_MEDICO

SEQ_MEDICO: INTEGER NOT NULL [PK]

NME_MEDICO: VARCHAR(100) NOT NULL

ESPECIALIDADE: VARCHAR(50) NOT NULL

FT_CONSULTAS

NRO_CONSULTAS: INTEGER NOT NULL

VRL_CONSULTAS: DOUBLE NOT NULL

Novo relacionamento identificado

DI_MEDICO

SEQ_MEDICO: INTEGER NOT NULL [PK]

NME_MEDICO: VARCHAR(100) NOT NULL

ESPECIALIDADE: VARCHAR(50) NOT NULL



FT_CONSULTAS

SEQ_MEDICO: INTEGER NOT NULL [PFK]

NRO_CONSULTAS: INTEGER NOT NULL

VRL_CONSULTAS: DOUBLE NOT NULL

DI_CARATERISTICA_PACIENTE

SEQ_CARATERISTICA_PACIENTE: INTEGER [PK]

SEXO: VARCHAR(10)

FAIXA_ANO_NASC: VARCHAR(20)

ANO_INICIO_FAIXA: INTEGER

ANO_FIM_FAIXA: INTEGER

DI_TEMPO

SEQ_TEMPO: INTEGER [PK]

MES: VARCHAR(20)

TRIMESTRE: INTEGER

SEMESTRE: INTEGER

ANO: INTEGER

FT_CONSULTAS

SEQ_TEMPO: INTEGER [PFK]

SEQ_MEDICO: INTEGER [PFK]

SEQ_CLINICA: INTEGER [PFK]

SEQ_CARATERISTICA_PACIENTE: INTEGER [PFK]

NRO_CONSULTAS: INTEGER

VRL_CONSULTAS: DOUBLE

DI_MEDICO

SEQ_MEDICO: INTEGER [PK]

NME_MEDICO: VARCHAR(60)

ESPECIALIDADE: VARCHAR(100)

DI_CLINICA

SEQ_CLINICA: INTEGER [PK]

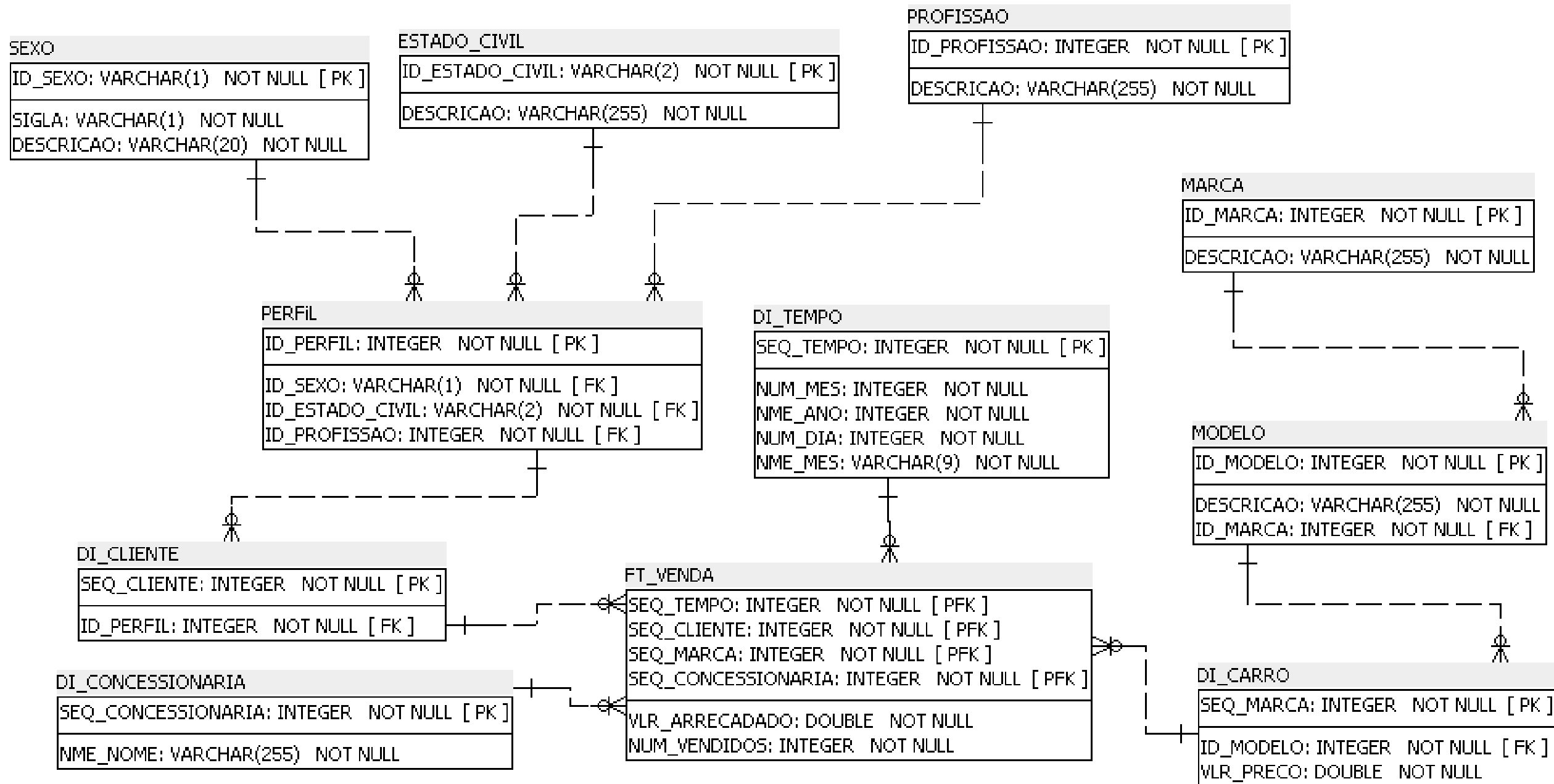
NME_CLINICA: VARCHAR(40)

EXERCÍCIO 2

Nas entrevistas de levantamento de requisitos, para a construção de um modelo dimensional de uma rede de concessionárias de carros foram identificadas as seguintes necessidades de informação:

- Acompanhar a evolução mensal das vendas (número de carros vendidos) e do valor total arrecadado por concessionária;
- Identificar o perfil dos clientes pelo: sexo, estado civil (solteiro, casado, viúvo, outro) e a profissão (empregado privado, funcionário público, profissional liberal, negócio próprio);
- Identificar qual é a marca (exemplo, Fiat, etc.) e o modelo (exemplo, Palio, Siena, Argo, Cronos, etc.) de carros mais vendidos pelas concessionárias.

Desenhe o modelo dimensional (tabela de fatos, medidas e dimensões) que atenda os requisitos de informação.



EXERCÍCIO 3

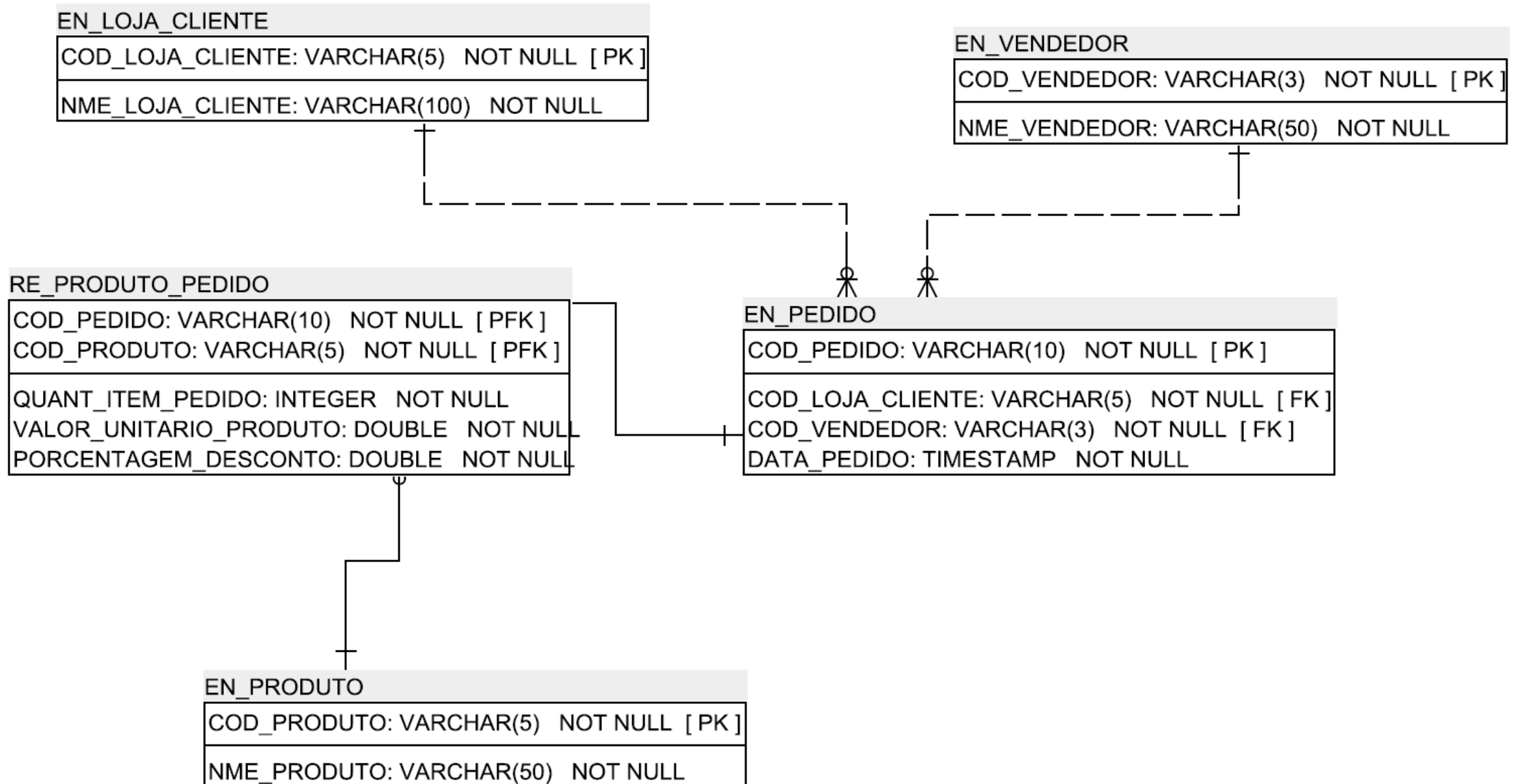
Uma importadora de TVs vende seus produtos, para as lojas da cidade. O esquema “relacional_pedidos” (ver próxima página) representa os pedidos recebidos pela importadora (ver figura a seguir). O esquema, leva em consideração qual é a loja que solicitou o pedido, quem foi o vendedor que atendeu ao pedido, a data da solicitação, assim como a quantidade de itens solicitados e os descontos outorgados pela importadora.

Nas entrevistas de levantamento de requisitos de informação, para a construção de um modelo dimensional, se identificou as seguintes necessidades:

- Deseja-se saber por mês, por produto e por loja:
 - a quantidade de itens solicitados;
 - o preço unitário médio (já que um produto pode ter diferentes preços);
 - o valor total vendido e o valor total vendido com desconto.

Desenhe o modelo dimensional (tabela de fatos, medidas e dimensões).

EXERCÍCIO 3



EXERCÍCIO 4

Desenhe um modelo dimensional (fatos, medidas e dimensões) para um Instituto de Meteorologia que possui 1.000 estações meteorológicas, classificadas em terrestres, marítimas e espaciais, as quais recolhem dados básicos sobre as condições atmosféricas a cada hora, incluindo pressão atmosférica, temperatura e precipitação.

Com o modelo dimensional construído, responder:

- Qual é o tamanho das tabelas (número de registros) em 1 ano?
- Altere o modelo para ter as condições meteorológicas por minuto. Qual é agora o tamanho das tabelas em 1 ano?

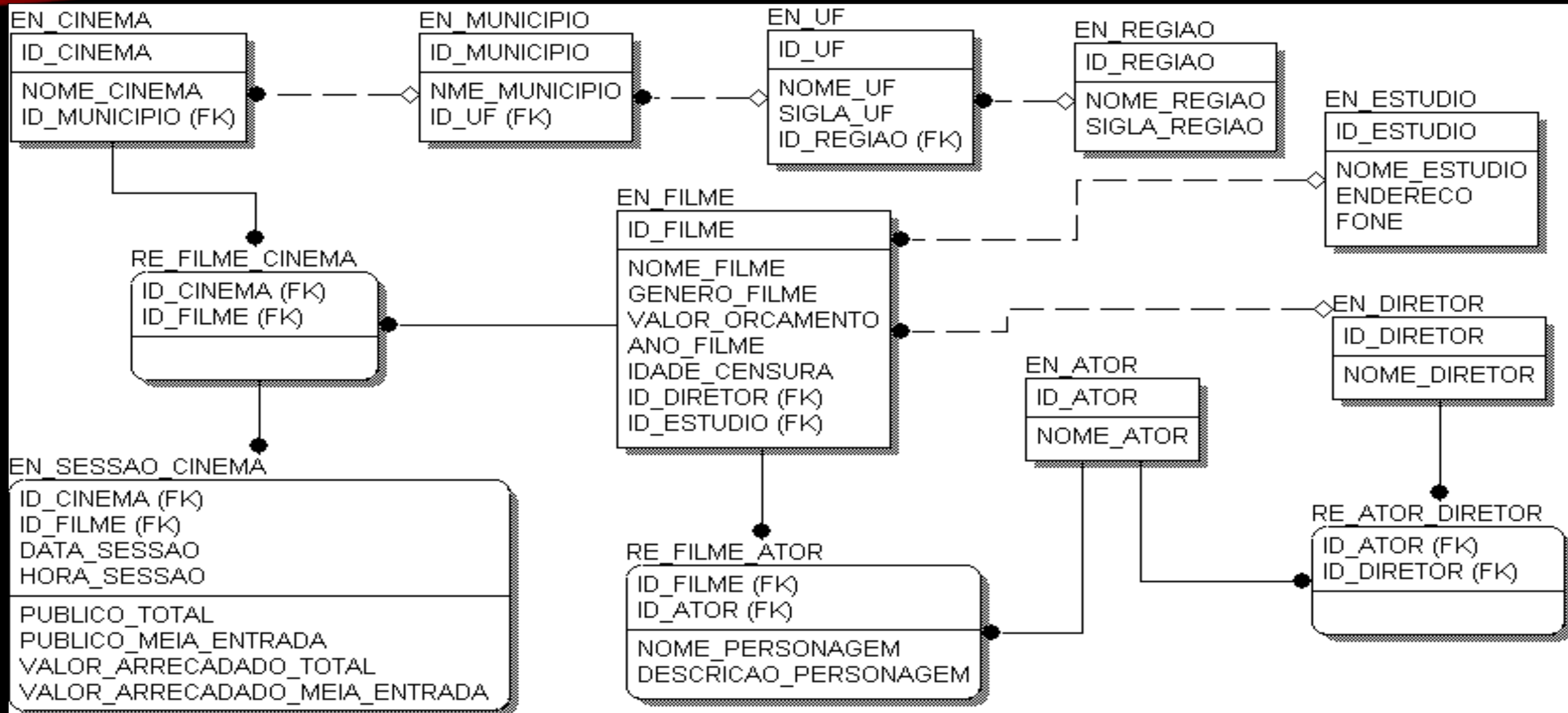
EXERCÍCIO 5

Uma distribuidora de filmes possui um sistema que permite o controle das salas de cinema (identificado pelo nome_cinema) que exibem os filmes de uma distribuidora. Algumas das informações que o sistema tem são:

- capacidade de lotação e localização de cada sala (cinema),
- os registros de bilheteria e o valor arrecadado em cada sessão diária de exibição dos filmes em cada sala,
- os atores e diretores que participam do filme,
- a classificação dos filmes por gênero,
- entre outras informações.

EXERCÍCIO 5

Entidade-Relacionamento do sistema de controle da distribuidora de filmes



EXERCÍCIO 5

Nas entrevistas de levantamento de requisitos de informação, para a construção de um modelo dimensional, se identificou as seguintes necessidades:

a. Acompanhar a evolução mensal de público e de arrecadação (total e quem pagou meia entrada), por região, UF e cidade, classificados por sala, por gênero e faixa etária (censura) do filme. Construir o modelo dimensional que atenda aos requisitos.

b. Qual seriam as alterações necessárias no modelo dimensional para calcular a arrecadação mensal por Shopping Center, e por diretor do filme?

c. Qual seria a alteração do modelo para saber o total de público por filme.

d. A distribuidora dos filmes é responsável pela propaganda nas diferentes mídias (jornais, revistas, tv, internet, etc), dos filmes que distribui. Deseja-se saber os valores investidos mensalmente em propaganda nas diferentes mídias para cada filme. A informação de investimento, não está no sistema de operação da distribuidora, é obtida dos contratos da distribuidora com as empresas prestadoras do serviço de propaganda. Altere o modelo para atender este requisito.

e. A distribuidora deseja saber se a relação entre os gastos em propaganda e valor arrecadado estão correlacionados, isto é, maior investimento em propaganda, maior valor arrecadado, por mês e por filme. Altere o modelo para atender este requisito.

BOM ESTUDO!

**BIG
DATA**



WIKI DA UC ANÁLISE DE DADOS E BIG DATA

- https://saulo.arisa.com.br/wiki/index.php/An%C3%A1lise_de_Dados_e_Big_Data
- <https://drive.google.com/drive/folders/1dIS-xdfukSV4LrBBvVMQeXLoeCN1q764?usp=sharing>

