

**21053 - Fundamentos de Bases de Dados
2013-2014
e-fólio C**

Resolução e Critérios de Correção

**PARA A RESOLUÇÃO DO E-FÓLIO, ACONSELHA-SE QUE LEIA
ATENTAMENTE O SEGUINTE:**

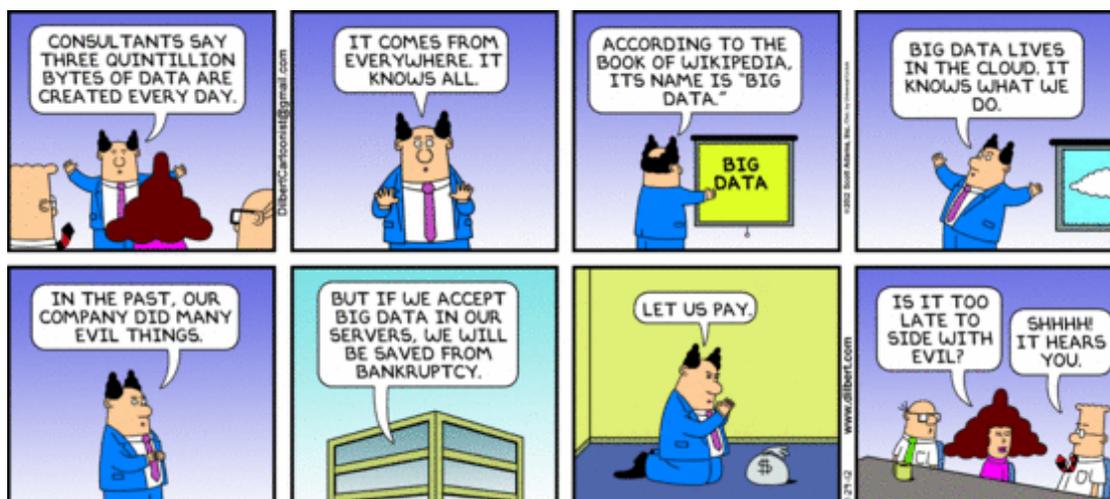
- 1) O e-fólio é constituído por 3 perguntas. A cotação global é de 3 valores.
- 2) O e-fólio deve ser entregue num único ficheiro PDF, não zipado, com fundo branco, com perguntas numeradas e sem necessidade de rodar o texto para o ler. Penalização de 1 a 3 valores.
- 3) Não são aceites e-fólios manuscritos, i.e. tem penalização de 100%.
- 4) O nome do ficheiro deve seguir a normal “eFolioC” + <nº estudante> + <nome estudante com o máximo de 3 palavras>. Penalização de 1 a 3 valores.
- 5) Na primeira página do e-fólio deve constar o nome completo do estudante bem como o seu número. Penalização de 1 a 3 valores.
- 6) Durante a realização do e-fólio, os estudantes devem concentrar-se na resolução do seu trabalho individual, não sendo permitida a colocação de perguntas ao professor ou entre colegas.
- 7) A interpretação das perguntas também faz parte da sua resolução, se encontrar alguma ambiguidade deve indicar claramente como foi resolvida.
- 8) A legibilidade, a objectividade e a clareza nas respostas serão valorizadas, pelo que, a falta destas qualidades serão penalizadas.

A informação da avaliação do estudante está contida no vetor das cotações:

Questão: 1 2.a 2.b 3.1 3.2 3.3

Cotações: 5 5 5 5 5 5 décimas

1) (1/2 valor) Escreva um texto, com 500 palavras, sobre o tema muito atual do “Big Data”. Refira também as relações com o NoSQL.



Resposta:

Com o advento da web 2.0 (a web das pessoas) associada aos dispositivos móveis e à “internet of things”, as clássicas aplicações empresariais foram largamente ultrapassadas em volume de dados.

Num estudo realizado em 2012, o valor estimado de informação no planeta foi de 2,8 ZB (zetabytes, 1021 bytes). A mudança de escala no volume de dados e na sua taxa de atualização deu origem ao que genericamente se chama de Big Data.

Ao nome Big Data está associada a sigla 3V: volume, velocidade de atualização e variedade dos formatos. Alguns autores incluem um quarto V de valor ou veracidade dos dados.

Dos 2,8 ZB existentes, 85% são dados não estruturados, ou seja, medias como o vídeo, fotografia e som. Dos restantes 15%, de dados formatados e de texto, só 3% são analisados. Concluimos assim, que só a pequena percentagem de 0,45% dos dados do planeta são objeto de análise. Tal como na astrofísica onde a matéria negra contém eventualmente várias explicações para a origem do universo, os 99,55% dos dados não analisados são apelidados de “dark data”.

Com o surgimento de novos formatos de dados estruturados surgiu dentro do Big Data, o conceito do NoSQL [Alexandre, Cavique 2013]. O NoSQL, ou Not only SQL, permite o armazenamento, tratamento e consulta de dados de forma muito eficiente. As soluções NoSQL estão divididas em alguns grupos:

- Armazenamento de chave/valor, como Voldemort do LinkedIn
- Armazenamento de super-colunas, como HBase ou Cassandra do Facebook
- Armazenamento de documentos, como XMLdatabase ou MongoDB
- Armazenamento de grafos, como HyperGraphDB ou ArangoDB
- Armazenamento orientado a objetos, como Db4object

Tendo como ponto de partida o modelo relacional e a linguagem declarativa SQL (Structured Query Language), utilizada na maior parte das bases de dados das empresas, o NoSQL apresenta-se como a alternativa para lidar com grandes volumes de dados.

As estruturas das soluções NoSQL foram simplificadas relativamente ao modelo relacional e garantem a consulta da informação de forma muito eficiente, com complexidades algorítmicas de ordem $O(1)$.

Em NoSQL a complexidade máxima deve ser da ordem $O(N)$, sendo N a dimensão do ficheiro. Ao contrário do SQL a operação de junção de tabelas não existe, dada a sua elevada complexidade de ordem $O(N^2)$ para o pior caso.

Para a agregação de dados é utilizado o conceito de MapReduce, implementado em duas fases. A função do operador Map seleciona os dados em subgrupos. A operação Reduce agrega a informação de cada subgrupo. A complexidade algorítmica no pior caso será de duas vezes $O(N)$.

O conceito de Big Data traz um conjunto de novos desafios para lidar com grandes volumes de dados, tanto para as empresas como para a comunidade científica. O desenvolvimento de novos algoritmos é crítico já que as complexidades algorítmicas são de preferência de ordem $O(1)$ e nunca devem exceder a ordem $O(N)$.

Por consequência, o Big Data cria novas oportunidades na tomada de decisão baseada em dados, “data driven decisions”. Tal como refere Peter Norvig, diretor da Google Research, “nós não temos melhores algoritmos; nós temos mais dados” [Davenport 2014].

Critérios de correção:

- discutir os Big Data 0,2 valores
- discutir os NoSQL 0,2 valores
- novos desafios 0,1 valores
- penalização para textos com menos de 500 palavras
- erros e omissões: -2 ou -3 décimas
- redundâncias: -1 a -2 décimas

2) Dependências funcionais e normalização

2.a) (1/2 valor) Para a relação $R(p, q, w, x, y, z)$ e o conjunto de dependências funcionais:

$$F = \{ \begin{array}{l} xy \rightarrow yz, \\ wz \rightarrow p, \\ p \rightarrow z, \\ w \rightarrow pqx, \\ qw \rightarrow pyz \end{array} \}$$

encontre a cobertura canónica (não redundante e reduzido à esquerda) e a redução à direita, utilizando as rotinas da bibliografia de David Maier do Capítulo 5.

Resposta:

Utilizando os símbolos \leq para incluído-ou-igual e $\langle \rangle$ para não incluído, teremos:

No Redundant FD $X \rightarrow Y$

Member($xy \rightarrow yz$)=0;	$yz \langle \rangle$ Closure= xy ;
Member($wz \rightarrow p$)=1;	$p \leq$ Closure= $pqwxyz$;
Member($p \rightarrow z$)=0;	$z \langle \rangle$ Closure= p ;
Member($w \rightarrow pqx$)=0;	$pqx \langle \rangle$ Closure= w ;
Member($qw \rightarrow pyz$)=0;	$pyz \langle \rangle$ Closure= $pqwxyz$;

F não redundante:

$$\begin{array}{l} xy \rightarrow yz \\ p \rightarrow z \\ w \rightarrow pqx \\ qw \rightarrow pyz \end{array}$$

Left Reduced FD $(X-A) \rightarrow Y$

(X-x) Member($y \rightarrow yz$)=0;	$yz \langle \rangle$ Closure= y ;
(X-y) Member($x \rightarrow yz$)=0;	$yz \langle \rangle$ Closure= x ;
(X-p) Member($\rightarrow z$)=0;	$z \langle \rangle$ Closure= $=$;
(X-w) Member($\rightarrow pqx$)=0;	$pqx \langle \rangle$ Closure= $=$;
(X-q) Member($w \rightarrow pyz$)=1;	$pyz \leq$ Closure= $pqwxyz$; (q removed)
(X-w) Member($\rightarrow pyz$)=0;	$pyz \langle \rangle$ Closure= $=$;

Cobertura canónica de F .i.e não redundante e reduzido à esquerda:

$$\begin{array}{l} xy \rightarrow yz \\ p \rightarrow z \\ w \rightarrow pqx \\ w \rightarrow pyz \end{array}$$

Para:

$xy \rightarrow yz$

$p \rightarrow z$

$w \rightarrow pqx$

$w \rightarrow pyz$

Right Reduced FD $X \rightarrow (Y-A)$

(Y-y) Member($xy \rightarrow y$)=1; $y \leq \text{Closure} = xyz$; (y removed)

(Y-z) Member($xy \rightarrow z$)=0; $z \leq \text{Closure} = xy$;

(Y-z) Member($p \rightarrow z$)=0; $z \leq \text{Closure} = p$;

(Y-p) Member($w \rightarrow p$)=1; $p \leq \text{Closure} = pqwxyz$; (p removed)

(Y-q) Member($w \rightarrow q$)=0; $q \leq \text{Closure} = pqwxyz$;

(Y-x) Member($w \rightarrow x$)=0; $x \leq \text{Closure} = pqwyz$;

(Y-p) Member($w \rightarrow p$)=0; $p \leq \text{Closure} = qwxyz$;

(Y-y) Member($w \rightarrow y$)=0; $y \leq \text{Closure} = pqwxz$;

(Y-z) Member($w \rightarrow z$)=1; $z \leq \text{Closure} = pqwxyz$; (z removed)

Cobertura canónica de F e reduzida à direita:

$xy \rightarrow z$

$p \rightarrow z$

$w \rightarrow qx$

$w \rightarrow py$

Solução final:

$xy \rightarrow z$

$p \rightarrow z$

$w \rightarrow pqxy$

Critérios de correção:

- não redundante 0,2 valores
- reduzido esquerda 0,2 valores
- reduzido direita 0,1 valores
- erros e omissões: -2 ou -3 décimas
- redundâncias: -1 a -2 décimas

2.b) (1/2 valor) Para a relação $R(a, b, c, d)$ e o conjunto das respectivas dependências funcionais:

$$F = \{ \begin{array}{l} a \rightarrow bc, \\ b \rightarrow d, \\ d \rightarrow b \end{array} \}$$

determine as chaves candidatas e proponha uma decomposição BCNF.

Resposta:

i) Chaves candidatas

O Fecho(X) representa-se por X^+ .

X é uma super-chave se X^+ contiver todos os atributos de R .

O atributo “c” não pertence ao conjunto de qualquer chave.

$$abd^+ = abcd$$

$$ab^+ = abcd$$

$$ad^+ = abcd$$

$$bd^+ = bd$$

$$a^+ = abcd$$

A chave mínima corresponde ao atributo “a”.

ii) Decomposição BCNF

Uma relação R está em BCNF sse:

- para todas as DF de R não triviais $X \rightarrow Y$, X for super-chave de R .

Basta existir uma DF em que X não seja super-chave, para R não estar em BCNF. Se isso acontecer, diz-se que essa DF viola a condição de BCNF.

Para $a \rightarrow bc$, o fecho $a^+ = abcd$, a^+ contem todos os atributos de R , é super-chave;

Para $b \rightarrow d$, o fecho $b^+ = bd$, não é super-chave;

A decomposição BCNF será: $R_1(a,b,c)$ e $R_2(b,d)$

$$F = \{ a \rightarrow bc, b \rightarrow d, d \rightarrow b \}$$

Critérios de correção:

- chaves candidatas 0,2 valores
- decomposição BCNF 0,3 valores
- erros e omissões: -2 ou -3 décimas
- redundâncias: -1 a -2 décimas

3) Projeto de Bases de Dados

Na biblioteca de uma freguesia existem vários títulos. Cada título tem um código, nome e autor; para cada título podem existir vários livros passíveis de serem requisitados. Os leitores são identificados por um código, nome e contacto. As requisições de livros aos leitores têm uma data de início e uma data de fim. Os títulos estão organizados por descritores (literatura, arte, informática, gestão, etc...) e por cotas conforme a figura.

biblioteca		Listagem das Principais Cotas			
		OBRAS DE REFERÊNCIA			
		LITERATURA HISTÓRIA GEOGRAFIA			
		030	Dicionários	36	Ciências Sociais em Geral
		(03)	Dicionários	31...	Estadística. Ponto de Acesso do INE
		(043)	Teses	316...	Sociologia das Organizações
		811...	Línguas	32...	Política
		831	Literatura	32...	Economia
		908...	Estudos de Área	330-101-541	Macroeconomia
		908...	História Local	330-101-542	Microeconomia
		911	Geografia	330-322	Investimentos
928	Geografia	330-324	Desenvolvimento Económico		
94...	História	330-4	Economia Matemática		
GENERALIDADES FILOSOFIA RELIGIÃO		CIÊNCIAS SOCIAIS			
0...	Generalidades. Documentação. Culture	331	Trabalho - Emprego		
061.1	UE - União Europeia	336	Finanças		
1...	Filosofia	336.2	Revelia Pública		
2...	Religião	336.7	Sistema Monetário - Bancos		
INFORMÁTICA AUTOMAÇÃO		336.76	Merc. Financeiro - Merc. de Capitais		
061...	Ciência	336	Política Económica		
	Conhecimento e Metodologia Científica	336.45	Posição Industrial		
	Organização do Trabalho Intelectual	336.50	Comércio, Economia Mundial		
		336.50	Organização Económica Internacional		
004	Informática	34	Direito		
004.2	Hardware	341	Direito Internacional		
004.3	Software	342	Direito Constitucional - Estado Administrativo		
004.4	Programação de Computadores	347	Direito Civil		
004.41	Programação de Computadores	348.2	Direito de Trabalho		
004.42	Programação de Computadores	348.48	Direito da Arquitectura		
004.43	Programação de Computadores	348.6	Direito de Autoridade		
004.431	Sistemas Operativos	35.07	Administração Pública		
004.432	Sistemas Operativos	35.17.13	Impostos		
004.433	Sistemas Operativos	36...	Assistência Social - Seguros		
004.434	Sistemas Operativos	37...	Educação - Paragem		
004.435	Sistemas Operativos	90	Ciências Puras - Generalidades		
004.436	Sistemas Operativos	CIÊNCIAS ENGENHARIA TECNOLOGIA			
004.437	Sistemas Operativos	91	Matemática		
004.438	Sistemas Operativos	912	Álgebra		
004.439	Sistemas Operativos	917...	Análise Matemática		
004.44	Redes de Computadores	919	Matemática Computacional		
004.45	Redes de Computadores	919.2	Estatística Matemática		
004.46	Redes de Computadores	919.8	Pesquisa Operacional		
004.47	Redes de Computadores	52	Astronomia		
004.48	Redes de Computadores	53	Física		
004.49	Redes de Computadores	53.3	Mecânica de Fluidos		
004.5	Inteligência Artificial	536	Óptica		
004.50	Inteligência Artificial	537	Óptica - Termodinâmica		
004.51	Inteligência Artificial	537	Electronagnetro		
004.52	Inteligência Artificial	54	Química		
004.53	Inteligência Artificial	54.7	Química Analítica		
004.54	Inteligência Artificial	548	Química Orgânica		
004.55	Inteligência Artificial	547	Química Orgânica		
004.56	Inteligência Artificial	547.8	Polímeros		
004.57	Inteligência Artificial	614.8	Segurança no Trabalho		
004.58	Inteligência Artificial	62...	Engenharia em Geral		
004.59	Inteligência Artificial	620.1	Engenharia em Geral		
004.60	Inteligência Artificial	620.1	Engenharia em Geral		
004.61	Inteligência Artificial	621	Engenharia Mecânica		
004.62	Inteligência Artificial	621.1	Máquinas - Ferramentas em Geral		
004.63	Inteligência Artificial	621.3	Eng. Eléctric. Electroléctrica		
004.64	Inteligência Artificial	621.34	Circuitos		
004.65	Inteligência Artificial	621.313	Máquinas Eléctricas		
004.66	Inteligência Artificial	621.316	Instalações Eléctricas		
004.67	Inteligência Artificial	621.38	Electrónica		
004.68	Inteligência Artificial	621.39	Processamento de Sinal		
004.69	Inteligência Artificial	621.5	Pneumática/Hidráulica		
004.7	Inteligência Artificial	621.6	Elementos de Máq. Desenho Mecânico		
004.8	Inteligência Artificial	621.9	Máquinas Ferramentas		
004.9	Inteligência Artificial	628.1	Água		
004.90	Inteligência Artificial	628.26	Eng. Sanitária - Trat. de Efluentes		
004.91	Inteligência Artificial	CONTABILIDADE GESTÃO PRODUÇÃO			
004.92	Inteligência Artificial	607	Contabilidade		
004.93	Inteligência Artificial	607.4	Cotas		
004.94	Inteligência Artificial	607.41.05	Contabilidade Financeira		
004.95	Inteligência Artificial	607.47	Contabilidade de Gestão		
004.96	Inteligência Artificial	607.6	Auditoria		
004.97	Inteligência Artificial	608	Gestão de Empresas		
004.98	Inteligência Artificial	608.01	Técnicas e Métodos de Gestão		
004.99	Inteligência Artificial	608.15	Gestão Financeira		
005	Informática	608.2	Gestão de Processos		
005.1	Informática	608.5	Gestão da Produção		
005.2	Informática	608.56	Gestão de Qualidade		
005.3	Informática	608.59	Manutenção		
005.4	Informática	608.7	Gestão de Stock		
005.5	Informática	608.8	Gestão de Vendas - Marketing		
005.6	Informática	608.9	Publicidade - Gestão da Informação		
005.7	Informática	609.2	Publicidade - Gestão da Informação		
005.8	Informática	ENGENHARIA TÊXTIL			
005.9	Informática	607	Engenharia Têxtil		
005.10	Informática	677.01	Análise de Materiais		
005.11	Informática	677.02	Processos de Indústria Têxtil		
005.12	Informática	677.022...	Têxteis - Estampagem - Acabamentos		
005.13	Informática	677.1/5	Fibras Têxteis		
005.14	Informática	687	Confecção		
005.15	Informática	ARTE ARQUITECTURA DESIGN			
005.16	Informática	69...	Construção Civil		
005.17	Informática	691...	Materiais de Construção		
005.18	Informática	692...	Equipamentos - Instalações		
005.19	Informática	693...	Arquitetura		
005.20	Informática	694...	Equipamentos - Instalações		
005.21	Informática	695...	Climatização - Ventilação - Ar Condicionado		
005.22	Informática	7...	Arte		
005.23	Informática	7.01	Estética e Teoria da Arte		
005.24	Informática	7.03	História da Arte		
005.25	Informática	711-316	Sociologia Urbana		
005.26	Informática	711	Arquitetura Pedagógica - Urbanismo		
005.27	Informática	711.4	Planeamento Urbano		
005.28	Informática	72	AZ-Arquitectos		
005.29	Informática	72	Arquitetura		
005.30	Informática	72.01	Estética e Teoria da Arquitectura		
005.31	Informática	72.011	Projecto		
005.32	Informática	72.013	História da Arquitectura		
005.33	Informática	72...	Estil. Rádicos. Contempor. Industriais		
005.34	Informática	728...	Habitación - Edifícios Residenciais		
005.35	Informática	74	Design		
005.36	Informática	74.03	História do Design		
005.37	Informática	742	Perspectiva - Geometria Descritiva		
005.38	Informática	744	Desenho Técnico		
005.39	Informática	Arquitetura - Arquivo de Perifoneos			
005.40	Informática	72(05) Periódicos de Arquitectura			
005.41	Informática	ARQUIVO DE PERIÓDICOS			

3.1) (1/2 valor) Comece por identificar as diferentes entidades do Modelo Entidade-Relação. De seguida, identifique os relacionamentos entre as entidades do Modelo Entidade-Relação. Dê um nome ao relacionamento. Classifique cada relacionamento quanto à cardinalidade (1:1, 1:N, N:N). De seguida, desenhe o diagrama do Modelo Entidade-Relação. Identifique os relacionamentos e os atributos das entidades.

Resposta:

Entidades:

títulos (id_título, nome, autor)

livros(id_livro, id_título, id_cota)

leitores(id_leitor, nome, contacto)

requisições(id_requisição, id_leitor, id_livro, data_início, data_fim)

descritores(id_descritor, nome)

cotas(id_cota, nome, id_descritor)

Relações:

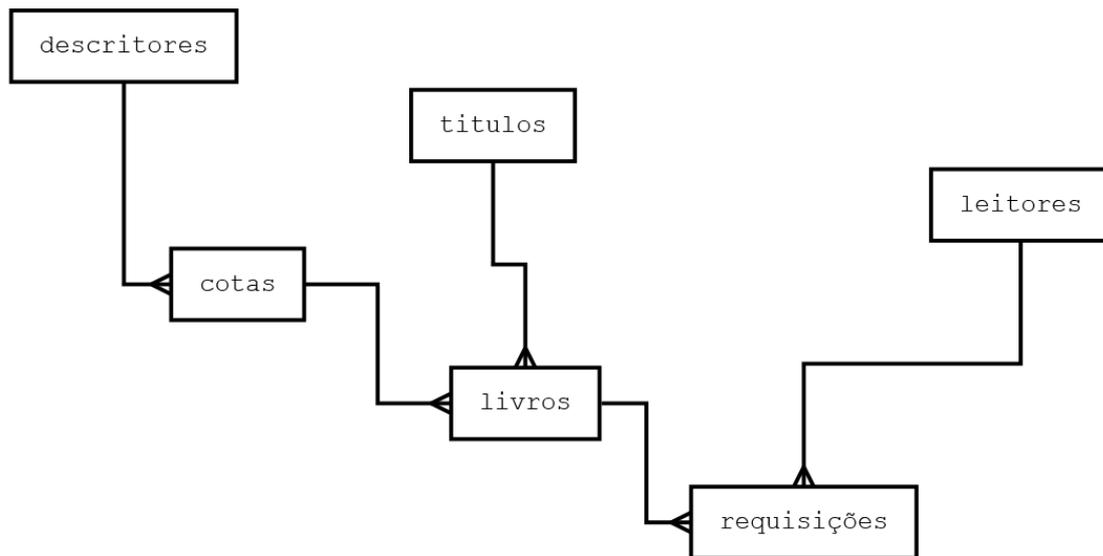
títulos – livros, 1:N, têm

requisições – leitores, N:N, têm

requisições – livros, N:N, têm

descritores – cotas, 1:N, têm

cotas – livros, 1:N, têm

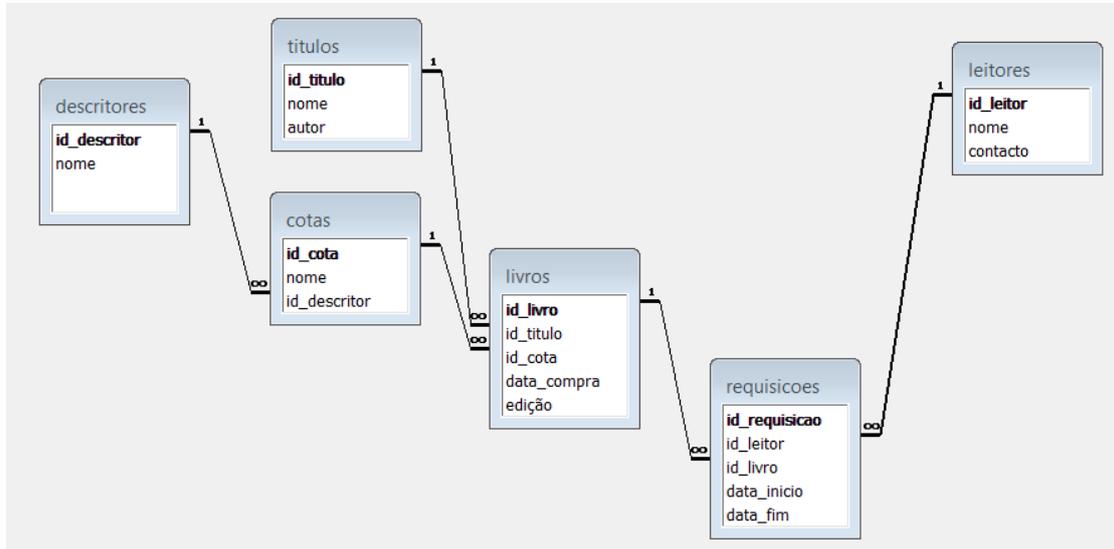


No modelo ER foi utilizada uma representação alternativa com os pés-de-galinha.

Critérios de correção:

- entidades e relações 0,2 valores
- modelo ER 0,3 valores
- erros nas cardinalidades -2 décimas
- erros e omissões: -2 ou -3 décimas
- redundâncias: -1 a -2 décimas

2) (1/2 valor) Desenhe a base de dados relacional correspondente ao modelo anterior, em que nas ligações de 1:N, a tabela com uma única linha é desenhada em cima e da tabela com várias linhas é desenhada por baixo. A base de dados não deve exceder as 7 tabelas. Se utilizar ligações que denotem transitividade, deve justificá-la convenientemente.



CrITÉRIOS de correção:

- tabelas 0,3 valores
- chaves estrangeiras 0,2 valores
- erros e omissões: -2 ou -3 décimas
- redundâncias: -1 a -2 décimas

3.3) (1/2 valor) Identifique dois fatores críticos de sucesso e escreva as respectivas consultas em SQL.

i) quais os leitores com livros em atraso?

```
SELECT data_fim, id_requisicao, leitor, id_livro
FROM requisicoes
WHERE data_fim < Date()
ORDER BY data_fim
```

ii) quais os livros mais lidos?

```
SELECT id_livro, COUNT(id_requisicao)
FROM requisicoes
GROUP BY id_livro
HAVING COUNT(id_requisicao) >=100
ORDER BY COUNT(id_requisicao) DESC
```

Critérios de correção:

- FCS 1 0,2 valores
- FCS 2 0,2 valores
- legibilidade SQL 0,1 valores
- erros e omissões: -2 ou -3 décimas
- redundâncias: -1 a -2 décimas