**21.103 Sistemas de Gestão de Bases de Dados**

**Atividade Formativa**

Nesta atividade e na seguinte, pretende-se exemplificar a construção de um Data Warehouse e realizar consultas OLAP.

Considere a texto de apoio em anexo para a fase de desnormalização de uma base de dados e de ETL no Data Warehouse.

1) Considere o seguinte esquema de base de dados, com as tabelas:

Pessoa = (id\_pessoa-> idade, género)

Questionário = (questionário-> detalhes)

Pergunta = (id\_pergunta-> questionário, texto)

Pessoa-Questionário= (id\_pessoa, questionário-> detalhes)

Pessoa-Pergunta= (id\_pessoa, id\_pergunta-> detalhes)

e com as ligações/chaves-estrangeiras:

Pessoa-Questionário.id\_pessoa ⊆ Pessoa.id\_pessoa

Pessoa-Pergunta. Id\_pessoa ⊆ Pessoa.id\_pessoa

Pessoa-Questionario.questionario ⊆ Questionario.questionario

Pergunta.questionario ⊆ Questionario.questionario

Pessoa-Pergunta.id\_pergunta ⊆ Pergunta.id\_pergunta

Carregue a base de dados e execute as seguintes consultas:

Q1: ∏ id\_pessoa, questionário (Pessoa |><| Pessoa-Questionário |><| Questionário)

Q2: ∏ id\_pessoa, questionário (Pessoa |><| Pessoa-Pergunta |><| Pergunta |><| Questionário)

O que pode concluir acerca das consultas com caminhos de junções diferentes?

Resposta:

O esquema de base de dados relacional é apresentado na figura 1, onde, para facilitar a leitura, houve o cuidado de colocar as tabelas com chave estrangeira ligeiramente abaixo da tabela com chave principal, produzindo ligações inclinadas de 1 para ∞.

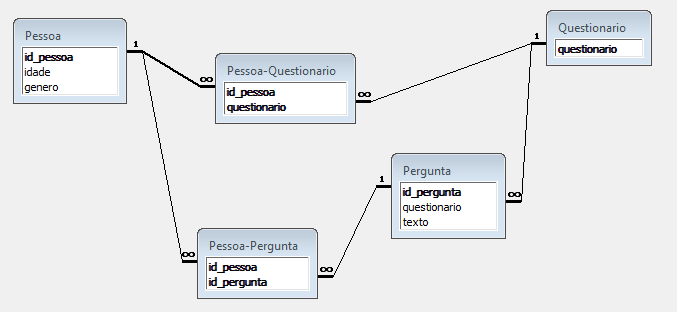


Fig. 1. Esquema da base de dados relacional

Depois de definir o esquema da base de dados não acíclica, vamos definir duas consultas, já que na figura podemos identificar dois caminhos entre as tabelas Pessoa e Questionário.

Nas respostas às consultas, pretende-se saber quais os questionários que as pessoas realizaram. Para o efeito, vamos utilizar como operador de junção o símbolo |><| e o símbolo ∏ para a projeção.

Q1: ∏ id\_pessoa, questionário (Pessoa |><| Pessoa-Questionário |><| Questionário)

Q2: ∏ id\_pessoa, questionário(Pessoa|><|Pessoa-Pergunta|><|Pergunta|><|Questionário)

Em SQL teremos para Q1:

SELECT Pessoa.id\_pessoa, Questionario.questionario

FROM Questionario

INNER JOIN (Pessoa

INNER JOIN [Pessoa-Questionario]

ON Pessoa.id\_pessoa = [Pessoa-Questionario].id\_pessoa)

ON Questionario.questionario = [Pessoa-Questionario].questionario;

e para Q2:

SELECT Pessoa.id\_pessoa, Questionario.questionario

FROM (Questionario

INNER JOIN Pergunta

ON Questionario.questionario = Pergunta.questionario)

INNER JOIN (Pessoa

INNER JOIN [Pessoa-Pergunta]

ON Pessoa.id\_pessoa = [Pessoa-Pergunta].id\_pessoa)

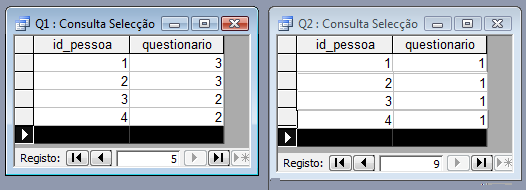
ON Pergunta.id\_pergunta = [Pessoa-Pergunta].id\_pergunta;

Fig. 2. As consultas Q1 e Q2 têm resultados diferentes

Um exemplo de resultado das duas consultas, está apresentado na figura 2, mostrando uma total inconsistência nas respostas. Mostra-se assim, que para consultas com caminhos de junções diferentes é possível obter resultados diferentes.

O Problema das Consultas por Múltiplos Caminhos, “Multiple Access Path Problem”, tratado em [Wald, Sorenson 1984] [Ullman 82] procura traduzir uma frase numa consulta não ambígua na base de dados. Este problema tem como solução imediata a utilização de bases de dados onde existe um só caminho entre duas tabelas.

2) Com base nas “Lecture Notes: Extraction”, aplique as regras de desnormalização. Quantas tabelas de factos encontra? O que as distingue?

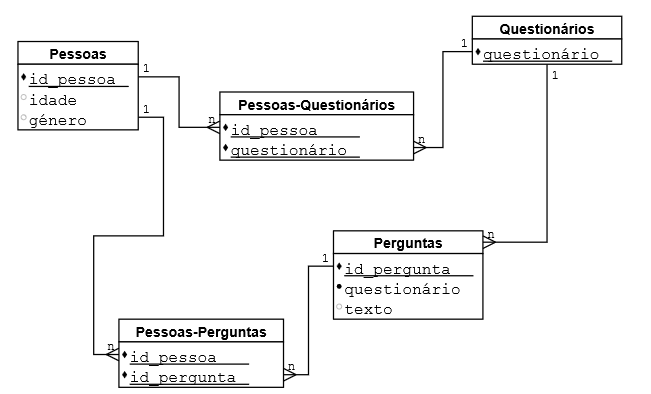
Resposta:

Segundo as Notas de Leitura temos 2 Formas para a Desnormalização:

1FD – obter uma poli-árvore, replicando as tabelas de ‘lookup’ que forem necessárias

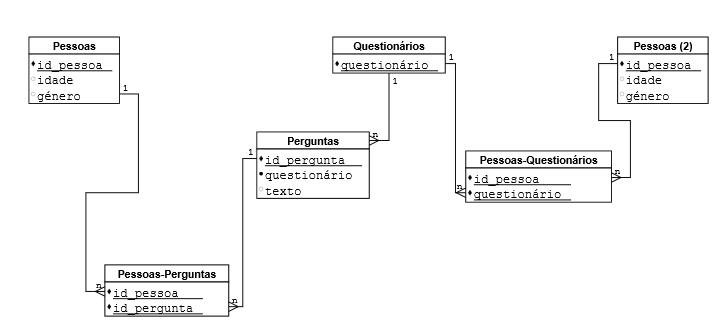
2FD – obter árvores separadas, replicando as tabelas de ‘lookup’ que forem necessárias

Tabela original



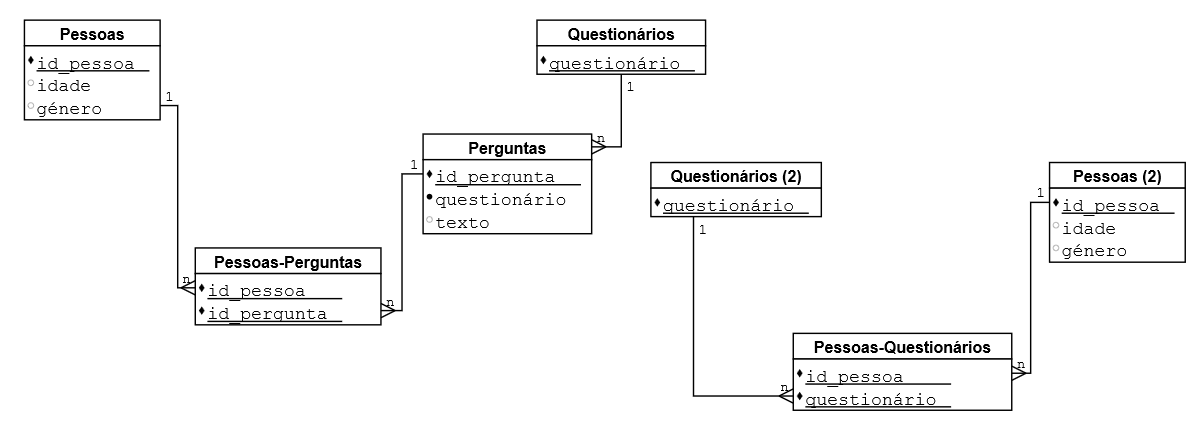
1FD -1ª forma desnormalizada: em poli-árvore

A tabela Pessoas foi replicada



2FD – 2ª forma desnormalizada: em árvores separadas ou em tabelas de factos

A tabela Questionários foi replicada



Finalmente:

Quantas tabelas de factos encontra? Obtemos 2 tabelas de factos

O que as distingue? Ambas as tabelas de factos, dizem-se tabelas de factos sem factos, visto que não têm dados aditivos.