

## ENGENHARIA REVERSA DE ARQUIVOS

Existem situações no ciclo de vida de um sistema nas quais um modelo conceitual é de grande valia:

- A manutenção rotineira do software de um sistema. Neste caso, o modelo conceitual pode ser usado como documentação abstrata dos dados.
- A existência de um modelo conceitual permite que pessoas que não conheçam o sistema possam aprender rapidamente seu funcionamento.
- Na migração do banco de dados o modelo conceitual pode acelerar o processo de migração, pois, permite encurtar a etapa de modelagem de dados do novo banco de dados.

O primeiro passo no processo de engenharia reversa de arquivos é a representação, do arquivo, na forma de uma tabela relacional não-normalizada.

A seguir, a tabela não-normalizada passa pelo processo de normalização, através do qual é obtido um modelo relacional, contendo as descrições das tabelas correspondentes ao arquivo em questão.

A partir do modelo relacional obtido, usando as regras de engenharia reversa (de um modelo relacional), chegamos no modelo ER do arquivo.

O objetivo do processo de normalização é:

- reagrupar informações de forma a eliminar redundâncias de dados que possam existir nos arquivos;
- reagrupar informações de uma forma que permita a obtenção de um modelo ER.

### PROBLEMAS / ANOMALIAS

- Redundância de dados: Repetição de dados.
- Inconsistência de dados:
- Inserção de dados: Não é possível inserir sem que exista outro
- Remoção de dados: Remover algo significa remover outras que não estão vinculadas.
- Alteração de dados:

Anomalias de inclusão, alteração e exclusão:

- Grupos de repetição;
- Dependências parciais de chave;
- Redundâncias de dados;
- Perdas acidentais de informação;
- Dificuldades de representações;
- Dependências transitivas entre atributos.

Anomalias devem ser evitadas porque violam a semântica da informação representada pelos itens de dados das entidades.

O primeiro processo consta em transformar a descrição do documento ou arquivo a ser normalizado em um esquema de uma tabela relacional.

Obtido o esquema relacional correspondente ao documento, passa-se ao processo de normalização. Este processo baseia-se no conceito de forma normal. Uma forma normal é uma regra que deve ser obedecida por uma tabela para que esta seja considerada “bem projetada”.

**1º Forma Normal** = Quando não contém tabelas aninhadas.

Passos:

- É criada uma tabela na 1FN, referente à tabela não-normalizada, sem as tabelas aninhadas. A chave primária da tabela na 1FN é idêntica à chave da tabela não-normalizada.
- Para cada tabela aninhada, é criada uma tabela na 1FN composta pelas seguintes colunas:
  - A chave primária de cada uma das tabelas nas quais a tabela em questão está aninhada;
  - As colunas da própria tabela aninhada.
- São definidas as chaves primárias das tabelas na 1FN que correspondem a tabelas aninhadas da seguinte forma:
  - Tomar como parte da chave primária da tabela na 1FN a chave primária da tabela ÑÑ.
  - Verificar se esta chave primária é suficiente para identificar as linhas da tabela na 1FN.
    - Caso seja suficiente, a chave primária da tabela na 1FN é a mesma que a da tabela aninhada na forma ÑÑ.
    - Caso contrário, deve-se determinar quais são as colunas necessárias para identificar as linhas da tabela na 1FN, compondo assim a chave primária na 1FN.

**2º Forma Normal** = Quando não existe dependência parcial em relação a chave primária composta.

Passos:

- Copiar para a 2FN cada tabela que tenha chave primária simples ou que não tenha colunas além da chave.
- Para cada tabela com chave primária composta e com pelo menos uma coluna não chave:

- Criar, na 2FN, uma tabela com as chaves primárias da tabela na 1FN.
- Para cada coluna não chave, pergunta-se “A coluna depende de toda a chave ou de apenas parte dela?”
  - Caso a coluna dependa de toda a chave:
    - Criar a coluna correspondente na tabela com a chave completa na 2FN.
  - Caso, a coluna dependa apenas de parte da chave:
    - Criar, caso ainda não existir, uma tabela na 2FN que tenha como chave primária a parte da chave que é determinante da coluna em questão.
    - Criar a coluna dependente dentro da tabela na 2FN.

**3º Forma Normal** = Não contém dependência transitiva.

Passos:

- Copiar para o esquema na 3FN cada tabela que tenha menos que duas colunas não-chave.
- Para tabelas com duas ou mais colunas não chave:
  - Criar uma tabela no esquema da 3FN com a chave primária da tabela em questão.
  - Para cada coluna não-chave fazer a seguinte pergunta “Existe dependência transitiva nesta coluna não chave?”
    - Caso não existir:
      - Copiar a coluna para a tabela na 3FN
    - Senão:
      - Criar, caso não exista, uma tabela no esquema na 3FN que tenha como chave primária a coluna da qual há a dependência transitiva.
      - Copiar a coluna dependente para a tabela criada.
      - A coluna determinante deve permanecer também na tabela original.

**4º Forma Normal** = Não contém dependência multivalorada.

Dependência multivalorada: Data uma relação R com atributos A, B, C, existe uma dependência multivalorada do atributo A no atributo B ( $A \twoheadrightarrow B$ ) se um valor de A é associado a uma

**coleção específica de valores de B**, independente de quaisquer valores de C.

- Dada  $R(A, B, C)$ , a DMV  $A \rightarrow B$  existe se  $A \rightarrow C$  também existir.
  - Notação:  $A \rightarrow B|C$