

BANCOS DE DADOS E MODELAGEM RELACIONAL APLICADA AO ARMAZENAMENTO DE DADOS AMBIENTAIS

Cezar Augusto Parladore¹; Letícia Moura²

RESUMO

O armazenamento de dados ambientais comuns costuma ser feito em diferentes meios. Neste resumo, é feita uma breve revisão da teoria entidade relacionamento e de bancos de dados relacionais a fim de elencar os problemas que a abordagem se propõe a resolver e mostrar como a aplicação da modelagem relacional a um conjunto exemplo de dados ambientais pode trazer benefícios ao armazenamento dos dados visando a redução de redundância de informações. Como resultado, têm-se um exemplo de modelagem que pode ser aplicado para gerar um conjunto de tabelas relacionadas e seus atributos. Conclui-se que o armazenamento dos dados em um banco relacional traz vantagens quando se comparado a outras formas de armazenamento como planilhas.

ABSTRACT

The storage of common environmental data is usually done in different media. In this summary, a brief review of the entity relationship theory and relational databases is made in order to list the problems that the approach aims to solve and show how the application of relational modeling to an example set of environmental data can bring benefits to the data storage aiming at reducing information redundancy. As a result, there is an example of modeling that can be applied to generate a set of related tables and their attributes. It is concluded that storing the data in a relational database has advantages when compared to other forms of storage such as spreadsheets.

¹ Water Services & Technologies, Av. Luiz Boiteux Piazza, 1302 - 3º andar sala 19 - Cachoeira do Bom Jesus, Florianópolis - SC, 88056-000; cezar.parladore@waterservicestech.com.

² Water Services & Technologies, Av. Luiz Boiteux Piazza, 1302 - 3º andar sala 19 - Cachoeira do Bom Jesus, Florianópolis - SC, 88056-000; leticia.moura@waterservicestech.com.

1 - INTRODUÇÃO

Nos últimos anos, o Meio Ambiente tem sido uma pauta cada vez mais recorrente. A Zion Market Research estimou em 2021 um crescimento anual de 7,5% para o mercado de remediação de áreas contaminadas, esse crescimento tem refletido em um aumento considerável no volume de dados ambientais. A análise desses dados pressupõe que sejam armazenados de forma consistente e confiável, porém o que se observa em setores da indústria é que, por fatores como falta de planejamento, falta de recursos financeiros ou técnicos, entre outros, os dados obtidos são armazenados de forma pouco estruturada em formatos como documentos impressos, imagens e planilhas. Isso se torna contraditório quando se considera que o primeiro modelo de banco de dados relacional foi proposto ainda em 1970 por E. F. Codd com objetivo de solucionar problemas dos modelos de armazenamento de dados então disponíveis como inserção de dados anômalos ou contraditórios em tabelas ou armazenamento de informação redundante. A utilização de um banco de dados relacional ou mesmo a utilização de princípios da teoria relacional traz benefícios importantes para o armazenamento de dados ambientais.

2 - OBJETIVOS

Têm-se como objetivos comparar um cenário exemplo de coleta de dados ambientais e aplicar conceitos da teoria de bancos de dados relacionais para mostrar os benefícios que podem ser obtidos no armazenamento, confiabilidade e posterior acesso aos dados para análises.

3 - METODOLOGIA

Para este trabalho, considerou-se a revisão da teoria de entidades e relacionamentos de Peter Chen, da teoria de bancos de dados relacionais proposta por E. F. Codd e dos princípios utilizados em sistemas de banco de dados relacionais. Partindo-se de um modelo simplificado de dados no contexto ambiental de subsuperfície representado na Figura 1 a seguir, aplicaram-se os princípios de modelagem relacional numa base de dados armazenada em uma planilha a fim de se obter um modelo que evidencie quais benefícios esperados de um modelo relacional foram atingidos.

Local de coleta da amostra															
Informações da amostra	Nome do Ponto	Latitude	Longitude	Responsável pela Coleta	Endereço do Responsável	Nome da Amostra	Data e Hora de Coleta	Matriz da Amostra	Teste Analítico	Laboratório	Data da Análise	Parâmetro	Resultado	Unidade de Medida	Grupo do Parâmetro
Testes aplicados na amostra															
Resultados obtidos															

Figura 1: exemplo de um caso comum de coleta de dados ambientais para análise laboratorial.

4 - DISCUSSÃO E RESULTADOS

O modelo de bancos de dados relacionais proposto por E. F. Codd se baseia na teoria entidade-relacionamento de Peter Chen. No modelo, cada entidade é representada por uma tabela e os atributos (características) da entidade em colunas; cada linha na tabela representa um registro da entidade. Dentro de cada tabela, o subconjunto de atributos que identificam unicamente cada registro é chamado de chave primária. As relações entre entidades ou tabelas são representadas por meio de chaves estrangeiras, que são mapeamento entre colunas ou atributos que referenciam outras entidades. Além disso, a modelagem relacional lança mão das chamadas “formas normais”, regras que podem exigir a remoção de atributos ou divisão de uma tabela em duas ou mais para se prevenir

problemas como: inserção desnecessária de dados repetidos em atributos que são sempre dependentes do valor assumido por outro atributo, inserção de dados inconsistentes pois são consequência direta de outro atributo na tabela e não arbitrários, volume de dados armazenado, entre outros.

Como exemplo, partindo-se do caso ilustrado na Figura 1 e admitindo que cada registro pode ser unicamente identificado pelos campos [nome do ponto, nome da amostra, data da coleta, parâmetro], a coluna “grupo do parâmetro” representa uma possível violação de uma forma normal, uma vez que o grupo do parâmetro pode ser obtido sabendo o nome do próprio parâmetro. Além disso, o grupo do parâmetro depende apenas do parâmetro em si, não estando relacionado a qual ponto a amostra foi coletada ou a data da coleta. Logo, essa informação pode ser guardada em uma tabela separada onde ocorre apenas o nome do parâmetro e quais são os atributos que dele dependem, como o grupo, evitando-se assim que uma coluna inteira seja preenchida com valores repetidos desnecessariamente ou que seja preenchido um grupo de parâmetro que não se aplica ao parâmetro em questão que foi analisado. O mesmo ocorre entre as colunas “responsável pela coleta” e “endereço do responsável” por razões similares. O diagrama a seguir ilustra o resultado esquemático de uma aplicação da modelagem relacional sobre o modelo proposto na Figura 1.

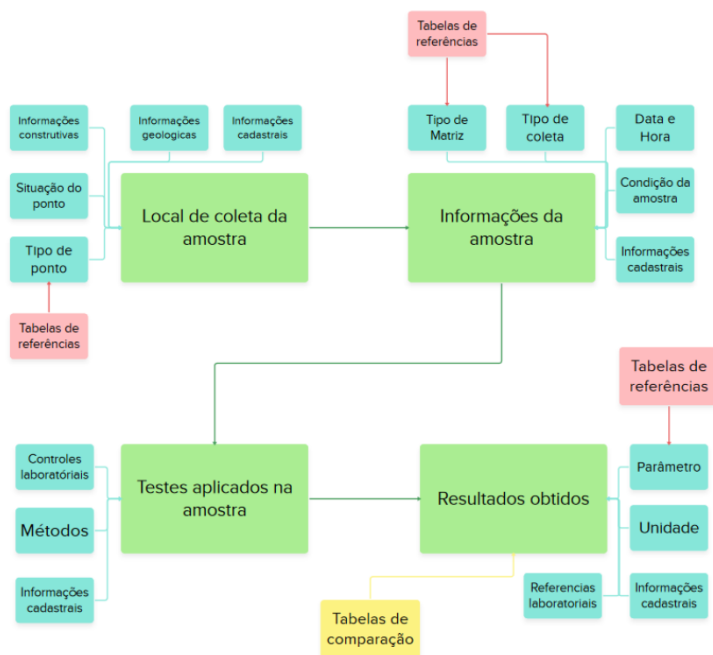


Figura 2: diagrama ilustrando a modelagem relacional ao exemplo utilizado mostrando atributos de cada entidade e possíveis tabelas relacionadas que reduziriam alguns atributos das tabelas originais.

5 - CONCLUSÃO

Conclui-se que o armazenamento de dados ambientais em uma estrutura relacional disfruta de benefícios que não ocorrem no armazenamento dos dados de formas não estruturadas ou compiladas. Exemplos de ferramentas para tal são o Microsoft SQL Server e MySQL ou *softwares* de prateleira como o Hydro GeoAnalyst da Waterloo Hydrogeologic.

6 - REFERÊNCIAS

Machado, F. N. R. Banco de Dados – Projetos e Implementação. Editora Érica, 4ª edição, São Paulo, 31 de janeiro de 2020.

Environmental Data Management Systems. Disponível em: <https://edm-1.itrcweb.org/environmental-data-management-systems/>. Acesso em: 21 jul. 2023.