



Módulo 2

Organização de dados: banco de dados, estruturação e limpeza de dados



Como vimos no módulo anterior, **coletar, organizar e visualizar dados** é um processo fundamental para obter insights valiosos na análise de dados. Agora, vamos aprofundar esse conceito, explorando como o **SQL** (Structured Query Language) se relaciona com o funcionamento dos **bancos de dados**, uma das principais ferramentas para armazenar e gerenciar essas informações.



Banco de Dados: O que é e Como Funciona

Imagine um **banco de dados** como uma grande tabela, semelhante a uma planilha do Excel. Cada **linha** dessa tabela representa um registro individual, como, por exemplo, as informações de uma pessoa, e cada **coluna** contém atributos específicos sobre esse registro, como nome, idade ou cidade. Portanto, um banco de dados organiza as informações de maneira estruturada, facilitando o armazenamento, a consulta e a manipulação dos dados.

O **SQL** é a linguagem que permite interagir com essa tabela. É a ferramenta que você utiliza para **acessar, inserir, modificar e excluir dados** no banco de dados de maneira eficiente.

Aqui estão alguns exemplos práticos de como o SQL funciona:

1. Buscar dados no banco de dados (Consultar informações):

Suponha que você tenha uma tabela com nomes e dados de clientes e queira saber quais moram em São Paulo. Com o SQL, você pode fazer uma “pergunta” ao banco de dados, que retornará as informações solicitadas:

```
SELECT nome, cidade FROM clientes WHERE cidade = 'São Paulo';
```



Esse comando retorna todos os clientes que têm São Paulo como cidade de residência. Essa consulta pode ser utilizada, por exemplo, para segmentar uma campanha de marketing.

2. Adicionar novos dados ao banco de dados (Inserir informações):

Imagine agora que um novo cliente chamado João se cadastrou e você precisa incluir os dados dele no banco de dados. O SQL facilita essa inserção de forma simples:

```
INSERT INTO clientes (nome, idade, cidade) VALUES ('João', 30, 'Rio de Janeiro');
```

Com esse comando, as informações de João são adicionadas ao banco de dados, tornando-o um novo registro na tabela.



3. Atualizar dados existentes (Modificar informações):

Se João se mudou para São Paulo, você pode atualizar essa informação no banco de dados utilizando SQL:

```
UPDATE clientes SET cidade = 'São Paulo' WHERE nome = 'João';
```

Agora, a cidade de João será alterada de 'Rio de Janeiro' para 'São Paulo' no banco de dados.

4. Apagar dados do banco de dados (Remover informações):

Caso João decida cancelar sua conta e você precise remover seus dados do banco de dados, o SQL também pode executar essa tarefa:

```
DELETE FROM clientes WHERE nome = 'João';
```

Esse comando remove o registro de João e exclui suas informações permanentemente do sistema.





Esse comando remove o registro de João, excluindo suas informações permanentemente do sistema.

Estruturação e Limpeza de Dados

Além de manipular os dados, um passo crucial no processo de análise é a **higienização e organização dos dados**. Isso significa garantir que os dados estejam corretos, consistentes e prontos para análise.



Higienizar dados envolve a remoção de duplicidades, correção de erros e preenchimento de informações faltantes. Por exemplo, em um banco de dados de clientes, você pode encontrar:

- **Dados duplicados:** O mesmo cliente pode ter sido registrado mais de uma vez, o que pode comprometer a precisão das análises.
 - **Nomes escritos de forma incorreta:** Pequenos erros de digitação podem gerar inconsistências, dificultando a consulta e a visualização de dados corretos.
 - **Campos incompletos ou vazios:** Informações importantes, como números de telefone ou endereços, podem estar ausentes, o que prejudica o valor das análises e o alcance de decisões estratégicas baseadas nesses dados.
-



Por isso, a **limpeza dos dados** é essencial. Por exemplo, se você tiver múltiplas entradas para o mesmo cliente, como “João Silva” e “Joaõ Silva”, esses dados precisam ser consolidados e padronizados para garantir uma análise precisa. Além disso, o processo de **padronização** de formatos, como datas, endereços ou números de telefone, também é parte fundamental da limpeza de dados.

Quando os dados estão organizados e limpos, o próximo passo é utilizá-los para gerar insights de valor, permitindo que decisões mais informadas e estratégicas sejam tomadas no ambiente empresarial.

