

Aula 02

Instalação do Postgres

Guiar as pessoas na instalação.

Primeiros passos em análise de dados

Passos da análise de dados:

1. Ouvir e Entender o problema
2. Criar/melhorar o modelo
3. Testar se funciona
4. Voltar ao passo 1

Ouvir e entender o problema

Se eu não sei **qual** problema eu vou resolver, eu não sei como posso resolver, nem se realmente existe um problema.

Um ponto importante é ter sempre em mente o escopo do problema, ou seja, quais os limites definidos para a resolução do problema.

No caso do armazenamento e busca de dados, precisamos entender:

- quais dados precisamos;
- quais dados serão armazenados;
- como estes dados serão armazenados de forma otimizada;
- quais dados serão buscados;
- quais as principais buscas de dados que são feitas.

Com isso podemos chegar ao próximo passo, que o **desenho da solução** ou **modelo**.

Modelar

Um modelo é uma visão restrita do mundo real que nos ajuda a entender uma realidade e nos dá um apoio para resolver um problema.

Um mapa é um modelo de um terreno. Ele nos ajuda a saber como sair de um lugar e ir até o outro, e simplifica o processo usando retas, cores e iconografias. Mas o mapa não é o caminho, ele é apenas uma representação desse caminho.

Com o escopo do problema bem definido, podemos pensar em *qual pedaço do mundo eu preciso que esteja no meu modelo?* Trazer os conceitos do mundo real para o modelo conceitual, e restringir o escopo da solução.

Uma modelagem de banco de dados é um modelo de dados que nos ajuda a armazenar os dados para busca eficiente, armazenamento confiável e distribuição dos dados de acordo com a natureza.

Há um padrão de desenho para modelo de banco de dados, conhecido como MER, ou Modelo Entidade-Relacionamento. Esse padrão usa alguns itens-base comuns, como Entidades (table), Campos (field), Relacionamentos (join) e Cardinalidade.

Dica: falhe rápido, falhe sempre, mas mantenha-se em movimento. Faça algo simples e que pode ser melhorado, e não algo perfeito. E faça as coisas em passos.

Testar

Saber se algo funciona depende de testes, e de uma disciplina que avalie:

- os dados estão armazenados corretamente?
- os dados estão disponíveis?
- as buscas que eu preciso fazer são possíveis/realizáveis?

É importante separar o momento da modelagem do momento do teste, para que o fluxo funcione de forma a trazer resultados. Se isso não for separado, podemos incorrer em erros de análise e depois exigir retrabalho.

Dica: documente os testes, e sempre aplique os mesmos testes pra cada iteração da modelagem. Isso ajuda a saber que problemas foram resolvidos, que problemas apareceram.

Reavaliar

Com os resultados dos testes, você pode tomar duas decisões:

- resolver o problema;

- definir que isso não é parte do escopo.

E daí voltar pro começo.

Links da aula

[CRISP-ML\(Q\)](#)

[What is CRISP DM? - Data Science Process Alliance](#)