

Cruzando bases com SQL

Transcrição

[00:00] Antes de prosseguir, precisamos aprender mais sobre cruzamentos. Tínhamos começado fazendo um cruzamento em que apareceu uma linha que não queríamos que aparecesse, resolvemos priorizar tabelas, vimos que podemos cruzar as bases de diversas formas. Vamos analisar um pouco as diferentes e diversas formas de cruzar bases.

[00:30] Imagine que temos a base de contatos com telefone e outra com e-mail. Como tínhamos visto, podemos muitas vezes dar apelidos para ficar mais fácil de se referir a cada base. Vamos chamar a de telefone de A e a de e-mail de B. Usando o diagrama de Venn, podemos fazer essa análise. Considerando que o círculo da esquerda são informações da base A e o da esquerda informações da base B, sabemos que uma parte da informação pertence às duas bases, e outra parte não é comum.

[01:48] O primeiro cruzamento que podemos ver é como tínhamos feito na base de clientes com a base de operações. O que queríamos era priorizar, fixar as informações da base de clientes e simplesmente acrescentar o que pudesse ser acrescentado nela que fosse encontrado na outra base.

[02:12] Como aqui estamos querendo fixar a chave da base da esquerda, chamamos de left join. Ou seja, juntando à esquerda, fixando a base da esquerda e colocando informações novas nela.

[02:40] O que as duas bases têm em comum é o fulano. Se fizemos um cruzamento, teríamos fulano e siclano, da base A, com o e-mail do fulano, que está na base da direita. Mas também podemos fazer um cruzamento parecido, que prioriza a base da direita. Dessa base, vou procurar informações que tenho na base da esquerda. Chamamos de right join, porque estamos cruzando a direita. É só uma questão de qual base vamos chamar de que modo.

[03:33] Na base da direita, temos fulano e beltrano, com fulano em comum. Se fizermos o cruzamento, teremos fulano e beltrano, que são os indivíduos da base B, e o telefone do fulano, que é informação puxada da outra base.

[03:55] Também podemos fazer um cruzamento para pegar só o que as duas bases têm em comum. Isso chamamos de inner join, um cruzamento interno. Nessas bases, o que temos em comum é o fulano. Nossa base resultante teria só o fulano, com telefone e e-mail dele.

[04:40] Também podemos pegar tudo de todas as bases. Não queremos priorizar uma ou outra base, ou só pegar o que é comum nas duas. Queremos tudo. Esse tipo de cruzamento é chamado de full join.

[05:17] Cada um desses cruzamentos depende do que você deseja ter no seu resultado. Nós iremos fazer a mesma coisa, que é avaliar quantos jogos cada cliente alugou, só que agora iremos fazer isso usando PROC SQL. O que tínhamos visto de join é muito importante porque o SQL usa essas expressões diretamente.

[06:08] Nós tínhamos primeiro antes sumariado a base de operações para cada CPF, quantos jogos ele alugou naquele mês. Outra operação que fizemos foi criar uma variável nova com o CPF raiz dele para conseguir cruzar as duas bases.

[06:30] Agora, iremos fazer algo parecido. Vamos fazer o cruzamento das bases usando o PROC SQL. Começamos falando create table, depois damos o nome da base. Pedimos para selecionar as variáveis. Na hora de falar a seleção de variáveis, iremos cruzar quais bases? Clientes, que vamos manter à esquerda, e operações sumariadas à direita. Na base de clientes, queremos manter todas as variáveis. Para isso, usamos um asterisco.

[07:55] Também queremos acrescentar uma variável a mais, então separamos com uma vírgula e colocamos a variável. O próximo passo é dizer para selecionar essas variáveis a partir de onde. Depois, cruzamos à esquerda, com left join. Nossa chave de cruzamento vai ser definida com on. Eu quero que a variável CPF raiz dessa primeira base seja igual à variável CPF da minha segunda base.

[09:37] Lembra que no nosso data step podemos dar os apelidos das bases de A e B? Aqui vamos fazer a mesma coisa para simplificar. Por exemplo, na minha base de cadastro de clientes também tem uma variável CPF. Esse CPF sai de onde? Preciso dizer. Iremos simplificar dando um nome para a nossa base que é o as e depois o apelido, por exemplo, A ou B.

[10:30] Da mesma forma que na hora de falar de base, falamos do que está dentro do diretório, aqui quero a base A, com as variáveis que estão dentro dela. E quero a quantidade de jogos que está dentro da base B.

[11:02] Fechamos o SQL com quit. Se executarmos, temos a base com a quantidade de variáveis nela, como tínhamos feito antes também.