

Empilhando bases

Transcrição

[00:00] Se prestarmos mais atenção no nosso diretório, veremos que não temos só a base de setembro de 2017. Na verdade, temos o restante do ano inteiro. É isso que vamos fazer agora, uma análise para a Alura Play em que vamos observar quantos jogos cada cliente alugou no quarto trimestre de 2017 como um todo, considerando os três meses, e também a análise individual de cada mês, da mesma forma que tínhamos feito para setembro, mas para outubro, novembro e dezembro.

[00:50] Vamos começar dando uma olhada nas nossas bases com o PROCCONTENTS. Vou copiar e colar, porque vai ser basicamente a mesma análise olhando as três bases ao mesmo tempo.

[01:52] Para voltarmos para o código e fazermos essa análise das bases consolidadas, seria mais fácil juntar todas de uma vez só. Vamos ver agora que juntar informações similares é diferente de juntar observações similares. Quando tínhamos no caso do cruzamento duas bases, em uma tínhamos fulano e na outra também, tínhamos o beltrano e o beltrano, mas cada uma das bases tinham informações diferentes sobre os mesmos clientes. O que queríamos era juntar todas as informações dos clientes numa mesma base.

[03:04] No cruzamento, fazíamos uma junção lateral das bases, tentando colocar todas numa mesma base fazendo a paridade das observações, para assim ter uma base completa com todas as informações dos clientes. Agora, vamos fazer um cruzamento juntando informações similares. Temos as bases de operações em que as variáveis são todas iguais, mas o que vemos em cada mês é diferente. Imagine que tivéssemos diversos dispositivos em que conseguíssemos colocar nossa agenda de contatos, mas sincronizado. Agora queremos juntar todas as agendas em uma só.

[03:47] O que vemos na nossa base é que todas as listas de contato têm as mesmas informações, nome, telefone e e-mail. O que muda são as pessoas, as observações de cada uma das bases. O que precisamos fazer é acrescentar as observações, as pessoas que estão em cada uma. Vamos fazer uma forma de juntar bases diferentes. Vamos fazer a junção vertical. Isso muitas vezes chamamos de concatenar ou empilhar bases, porque quando vemos no resultado final, temos exatamente uma base com outra concatenada.

[05:02] De volta ao código, vamos empilhar as bases do quarto trimestre de 2017. Vamos começar analisando como fazer isso usando o próprio SQL. Começamos de forma que já conhecemos, criando uma tabela como, já que não queremos fazer nenhuma alteração no jeito como ela é. Depois falamos para selecionar todas as variáveis da base que está na Alura.

[07:26] Pegamos a base do jeito que está e carregamos no PROCSQL. No próximo comando vamos falar para unir essa base com outra, usando union. Essa outra base é basicamente o mesmo comando que usamos em cima. E precisamos juntar também com a base de dezembro, fazendo a mesma coisa.

[08:30] Temos uma base com 1.097 linhas e todas as variáveis. Conseguimos empilhar corretamente nossas bases usando SQL. Só temos que tomar cuidado com o comando de união, porque ele faz uma espécie de junção. Se ele encontrar linhas iguais, ele vai manter só as que são únicas e jogar fora as repetidas.

[09:24] Para garantir que isso não aconteça, temos que mudar de union normal para union all. Vamos unir tudo, independente das semelhanças.

[10:06] Além dessa opção, também podemos usar o data step. Para usar ele, usamos o comando set, puxando as bases que queremos. Fechamos o data step com run. Se executarmos, temos a mesma base resultante, com as variáveis e as 1.097 linhas.

[11:36] Eu pessoalmente, para concatenação, acho o comando do data step mais prático, mas temos diversas opções para resolver o problema.

[12:00] Como fazer agora a análise que tínhamos feito para a Alura Play? É só usar o mesmo código usado antes, mas ao invés de usar a base de setembro de 2017, vamos usar a base do quarto trimestre. É o mesmo código que tínhamos usado anteriormente, trocando as bases. Se olharmos os contratos agora, temos um número muito maior, porque estamos olhando três meses de uma vez.