

Exercício de Carregamento em Lote

Discente: Asafe Belo Borges

Docente: Allan Gonçalves de Oliveira

Veja esse e outros exercícios em [Asafe Belo](#) no Github

1. Explique como o Postgresql trabalha com processamento batch e bulk load.

O processamento batch é uma abordagem que envolve a execução de operações em grandes volumes de dados em lotes, ao invés de uma operação por vez. Dentre as vantagens do processamento em lote (por batches), incluem: agrupamento das operações, o que trás para si uma maior eficiência do que por meio de um processamento individual e o processamento em segundo plano, sem interação com o usuário.

O bulk load por sua vez se refere a importação ou carregamento de grandes volumes de dados em um banco de dados em um pequeno período de tempo.

No PostgreSQL por sua vez trabalha com o processamento batch e bulk load de duas formas diferentes, a primeira para realizar execução de operações em grandes volumes com o comando INSERT, e com o carregamento de grandes volumes de dados via comando COPY.

2. Mostre o funcionamento de INSERT com várias linhas e COPY. Tem alguma diferença entre os dois?

No PostgreSQL ambos os comandos podem ser usados para realizar operações de carregamento de grandes volumes de dados, mas com certa diferença em termos de opções de uso e eficiência.

Figura I - Insert de várias linhas na tabela tópico

Query	Query History	Notifications
1	INSERT INTO Topico (Codigo, Titulo, QntExercicios)	
2	VALUES	
3	(1, 'Soma de dois números', 5),	
4	(2, 'Polígonos', 5),	
5	(3, 'Material Dourado', 5);	

Este comando INSERT insere três linhas de dados na tabela Tópico. A vantagem de usar um único comando para inserir várias linhas é a redução da sobrecarga associada a múltiplas instruções SQL separadas.

Figura II - importando dados o arquivo alunos.csv para a tabela aluno

Query	Query History	Notifications
1	<code>COPY</code>	<code>aluno (id_aluno, nome)</code>
2	<code>FROM</code>	<code>'/escola/matriculas/alunos.csv'</code>
3	<code>WITH</code>	<code>(FORMAT csv, HEADER true);</code>

O comando COPY é uma maneira eficiente de carregar grandes volumes de dados em uma tabela do PostgreSQL a partir de um arquivo externo. Por exemplo, para importar dados de alunos para a tabela aluno.

2.1. Diferenças entre o comando INSERT e COPY

Em relação ao desempenho, o comando COPY geralmente é mais rápido do que o comando INSERT com várias linhas ao carregar grandes volumes de dados. Isso ocorre porque o COPY é otimizado para leitura direta de arquivos e inserção eficiente de dados.

Para importar dados, o comando COPY é ligeiramente melhor, uma vez que você tem a facilidade e a praticidade de importar os dados de um arquivo externo. O comando INSERT com várias linhas por sua vez é mais apropriado quando você tem os dados já disponíveis como parte da sua instrução SQL.

Já em relação ao controle de erros, em um erro no comando INSERT com várias linhas, a transação falhará e nenhuma linha será inserida. Com o comando COPY, é possível usar opções de controle de erro, como ON ERROR CONTINUE, para lidar com linhas com problemas sem interromper a operação completamente.

3. Considerações finais

Em resumo, o comando COPY é geralmente a melhor para importação em massa de dados de um arquivo externo, enquanto o comando INSERT com várias linhas é mais adequado para inserir múltiplas linhas de dados diretamente em uma instrução SQL. Ambos têm suas próprias vantagens e são úteis em diferentes cenários de carregamento de dados.