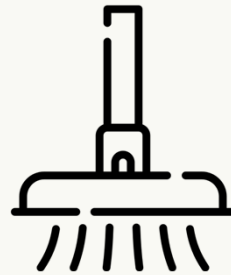


VACUUM



01 PADRÃO DE 7 DIAS

02 EVITE GASTOS EXTRAS

03 ATENÇÃO AO RISCO

04 SQL E PRÁTICAS

VACCUM pra Redução de Custos

Dataset de exemplo

Tabela: pedidos (Delta Lake)

id_pedido	cliente	valor	data_atualizacao
1	Ana	100	01/05
2	Bruno	200	01/05
3	Carla	150	01/05
4	Diego	120	02/05
5	Ana	300	02/05
6	Bruno	80	02/05
7	Carla	60	03/05
8	Diego	220	03/05
...			

Evolução dos arquivos da tabela (com atualizações e deletes)

Commit 1 (01/05) VERSÃO 1
part-0001 (100 MB) part-0002 (100 MB)
Dados iniciais

Commit 2 (02/05) VERSÃO 2
part-0001 (novo) (100 MB) part-0003 (novo) (100 MB)
part-0001 (antigo) (100 MB) part-0002 (antigo) (100 MB)
Arquivos antigos viram obsoletos

Commit 3 (03/05) VERSÃO 3
part-0001 (novo) (100 MB) part-0004 (novo) (100 MB)
part-0001 (versão 2) (antigo) (100 MB) part-0003 (antigo) (100 MB)
Arquivos antigos viram obsoletos

O que acontece sem VACUUM?

- Arquivos obsoletos continuam no storage
- Consumem espaço (armazenamento)
- Aumentam o custo
- Não são usados em queries, mas continuam custando

Problema

Com o tempo, a quantidade de arquivos obsoletos cresce muito, gerando desperdício de storage e custos desnecessários.

1 Como funciona o VACUUM

- Identifica arquivos obsoletos**
Lê o log de transações (Delta Log) e descobre quais arquivos não são mais referenciados pela versão atual.
- Verifica o período de retenção (configuração padrão: 7 dias)**
Arquivos só podem ser removidos se estiverem obsoletos há mais tempo que o período de retenção.
- Remove os arquivos fisicamente do storage**
Apaga os arquivos elegíveis do seu storage (S3, ADLS, GCS, etc.).
- Libera espaço e reduz custos**
Menos dados armazenados = menos custo. Tabela continua funcionando normalmente.

2 Antes e depois do VACUUM

Antes do VACUUM (muitos arquivos obsoletos acumulados)

Depois do VACUUM (obsoletos removidos)

Quando usar?

- Rotineiramente, para controlar o crescimento de dados obsoletos
- Tabelas com muitas atualizações e deletes
- Ambientes com custo baseado em storage (S3, ADLS, GCS)

Total de 8 arquivos físicos = 800 MB (apenas 4 estão ativos)

Total de 6 arquivos físicos = 600 MB (somente ativos)
Economia: 200 MB (25% de redução)

3 Resumo rápido

Aspecto	Sem VACUUM	Com VACUUM
Arquivos obsoletos	Acumulam no storage	Removidos após retenção
Espaço ocupado	Alto	Menor
Custo de storage	Alto	Reduzido
Desempenho de leitura	Pode ser impactado (mais arquivos no listing)	Melhor (menos arquivos)
Segurança de dados	Mantém versões antigas	Remove permanentemente após retenção
Operação	Não remove nada automaticamente	Limpeza controlada e segura

Dica de ouro

O VACUUM respeita o período de retenção (padrão 7 dias) para evitar remoção de arquivos que ainda podem ser necessários para:

- Time travel (consultar versões antigas)
- Rollback (reverter para versões anteriores)

Ajuste com cuidado, especialmente o retention period!

4 Exemplo de comando

- VACUUM com retenção padrão (7 dias)
VACUUM pedidos;
-- Remove arquivos obsoletos com mais de 7 dias
- VACUUM com retenção personalizada (ex.: 2 dias)
VACUUM pedidos RETAIN 2 HOURS;
-- ou
VACUUM pedidos RETAIN 2 DAYS;
-- Remove arquivos obsoletos com mais de 2 dias
- Verificar o que será removido (dry run)
VACUUM pedidos DRY RUN;
-- Lista os arquivos que seriam removidos,
-- mas não apaga nada

Importante: VACUUM remove arquivos fisicamente e não pode ser desfeito. Use apenas quando os arquivos não forem mais necessários para time travel ou rollback.