



**Um pipeline executado
100 vezes com o mesmo
resultado vale mais do
que um mais rápido
que falha no replay.**

Desliza →

Por que importa a idempotência?

Reprocessamentos, correções e reconciliações chegam a consumir 20–30% do tempo das equipas de engenharia de dados — e põem em risco relatórios e processos críticos.

Quatro princípios essenciais

Identificação única, deduplicação determinística, escrita transaccional e gestão explícita do estado são a base para pipelines reentrantes e previsíveis.

Padrão no Microsoft Fabric

Ingestão bruta para staging em OneLake →
transformação com Spark Notebook → MERGE
para upserts em Delta → actualização de tabela
de controlo com last_offset e job_run_id.

MERGE: upsert e logical deletes

O MERGE em Delta permite inserir, actualizar e marcar deletes lógicos numa única operação atómica; combine com um dedup por (order_id, source_sequence) antes do MERGE.

Tabelas de controlo que evitam reaplicações

Tabela com (pipeline_name, last_processed_offset, last_run_at, last_job_run_id, last_row_count). Antes de aplicar um batch, verifique se o source_batch_id já foi processado.

Testes e monitorização

Inclua testes de reproprocessamento (3–5 execuções) no CI, e métricas: taxa de duplicados, diferenças nas contagens, latência do MERGE e throughput.

Mini-caso: retalho digital

120 colaboradores, 2M eventos/dia. Antes: reprocessamento de 6h, 80 compute units, 4% duplicados e €14k/mês de impacto. Depois: 45min (-88%), 12 compute units e <0,02% duplicados.

CONTINUA NO BLOG

Como planeia garantir que um reproprocessamento do seu pipeline não cause mais problemas do que resolve?

www.bconcepts.pt/insights