



INSIGHTS · DATA ENGINEERING

CDC em Microsoft Fabric — pipelines robustos e práticos para dados em tempo quase real

Desliza →

Por que CDC importa

Decisões operacionais e relatórios exigem frescura. CDC reduz latência para minutos e pode cortar 10–30% do esforço humano em correcções.

Escolha de ingestão

Micro-batch (1–5 min) é simples e económico; streaming é para latência sub-minuto e picos altos.

Padrão comum no Fabric

Origem com CDC → Debezium/Event Hubs →
Spark Structured Streaming → Delta Lake na
Lakehouse.

Idempotência e ordenação

Use MERGE por key usando sequence/LSN e aceite eventos tardios numa janela de ≈ 15 minutos para evitar duplicados e reordenações.

Deletes e tombstones

Trate deletes como tombstones (is_deleted + change_ts) e aplique políticas de retenção e expurgo após auditoria.

Compactação e small files

Apontar ficheiros Parquet entre 128–512 MB; micro-batches de 2 min e compaction diária reduzem latência de leitura 30–70%.

Observabilidade e SLAs

Monitore throughput, lag ($P95 < 5 \text{ min}$), taxa de duplicados $< 0.05\%$ e alerte quando os limites são excedidos.

CONTINUA NO BLOG

Como desenhar pipelines CDC fiáveis no Microsoft Fabric?

www.bconcepts.pt/insights