

INSIGHTS · DATA ENGINEERING

Instrumentar pipelines: métricas que evitam crises

Desliza →

Porquê observabilidade?

Passar de «funciona» para «sabe-se que funciona» evita picos de incidentes e reprocessamentos urgentes.

Métricas essenciais

Comece com volume_in/volume_out, latency_ms (P50/P95/P99), error_count, freshness_ts e null_rate por coluna.

Deteção de anomalias com números

Se esperava ~120 000 linhas e o volume_in cai para 80 000, isso sinaliza impacto no downstream.

Telemetria por run

Cada job gera um `run_id` e escreve uma linha atômica com `start/end`, `status`, `volumes` e `schema_signature`.

Design da tabela de métricas

Particione por `start_ts` e `pipeline_id`; armazene `null_rate` e `tags` como JSON para flexibilidade e consultas rápidas.

Regras e alertas simples

Defina thresholds dinâmicos: por exemplo, alertar se volume_in cair $>30\%$ face à mediana das últimas 7 execuções.

Retenção e tendências

Mantenha detalhe 30–90 dias e tabelas agregadas para tendências a 1/7/30 dias, para suporte a decisões arquitectónicas.

CONTINUA NO BLOG

Por que tornar os pipelines observáveis?

www.bconcepts.pt/insights