

INSIGHTS · DATA ENGINEERING

Canary deployments: validar alterações em pipelines sem pôr a produção em risco

Desliza →

O risco em produção

Promover sem testar pode afectar KPIs críticos.
Numa área com 10M registos/dia, reduzir regressões de 5% para 0,5% equivale a poupanças significativas de tempo e custo.

O que é um canary

É uma implementação incremental que processa uma fracção dos dados ou corre em paralelo (shadow), permitindo validar alterações com recuperabilidade rápida.

Modelos práticos

Porcentagem (ex.: 5% de 200k eventos/h → 10k ev/h), amostra determinística (hash de IDs) e shadow runs (execução paralela com custo extra).

Passos no Microsoft Fabric

Versione transforms no Git, use parâmetros canary_selector, escreva em partições temporárias e orchestre com Pipelines; prefira Lakehouses com commits ACID.

Métricas para decidir

Automatize contagens, checksums, percentagem de nulos e percentis (P50/P90/P99). Use thresholds objetivos (ex.: contagens >1% ou chaves não casadas >0,5%).

Rollback seguro

Opções: swap de views atômica ou dual writes com toggle. Tenha scripts de rollback e playbooks testados para recuperação rápida.

Monitorização em tempo real

Combine métricas de infra (CPU, I/O, latência por job) com métricas de negócio (KPIs, taxa de exceções) para identificar causas reais.

CONTINUA NO BLOG

Por que usar canary deployments em pipelines de dados?

www.bconcepts.pt/insights