



INSIGHTS · DATA ENGINEERING

Sem contratos de dados, os pipelines são promessas; com contratos, são acordos executáveis.

Desliza →

O problema: expectativas invisíveis

Colunas que desaparecem ou formatos que mudam geram erros silenciosos. Em ambientes grandes, regressões podem levar horas ou dias a ser detetadas.

O que é um data contract?

Um acordo executável sobre esquema, tipos, invariantes e SLAs. Deve estar versionado e ter um dono claro para evitar deriva.

Conteúdo mínimo de um contrato

Identificador, esquema esperado, regras de obrigatoriedade, invariantes de negócio, métricas operacionais e ações em caso de violação.

Operacionalizar no Microsoft Fabric

Defina contratos em YAML/JSON no Git, valide com Jobs Spark antes de escrita em Delta e coloque lotes falhados em 'quarantine'.

Testes e promoção (CI/CD)

Fluxo sugerido: Dev → PR → testes de contrato → Staging com amostras (1-5%) → Prod só se os testes e SLAs passarem.

Monitorização e alertas acionáveis

Métricas essenciais: taxa de sucesso dos testes, % de nulos por coluna, variação de volume e MTTR. Dashboards em Power BI para visibilidade.

Resultados mensuráveis

Mini-caso: incidentes de 18 → 5/mês (-72%),
MTTR 4h → 45m e custos operativos reduzidos
para ~2.2k€/m.

CONTINUA NO BLOG

Os seus pipelines têm data contracts que previnem regressões — ou continuam à espera que os relatórios revelem os problemas?

www.bconcepts.pt/insights