

Implementando regras de validação com PySpark



Validação de Schema



Uso de flags no fluxo



Código Reutilizável

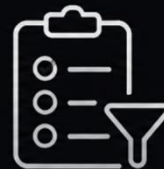
Estratégias Avançadas de Deduplicação no Spark



Drop Duplicates



Window Functions



Critérios de Seleção



Trade-offs

Fluxo de qualidade na Arquitetura Medallion



Decisão de Dados

Gold (Aggregated)

Silver (Cleaned)

Bronze (Raw)

Testes Automatizados

50%

Queda nos Registros

100%

Testes Unicidade

99%

Conformidade Range

95%

Pipelines Validados

