



Pipeline para ingestão de dados e processamento (batch)

1. Dados são obtidos do ambiente on-premise e sincronizados para o S3 via AWS Cli.
2. A "Transient Zone" é a primeira camada de ingestão de dados, onde ocorre a catalogação das origens e tipos de dados. Essa é uma zona temporária e os arquivos são apagados após o carregamento na zona de "Raw Data".
3. O EventBridge Scheduler está configurado para iniciar o pipeline.
4. Na zona de "Raw Data", os dados são armazenados de forma "bruta", sem qualquer modificação.
5. O AWS Step Functions é responsável pela orquestração de jobs.
6. Uma lambda function é destinada para o catalogamento dos arquivos que chegam ao bucket S3 (Transient Zone).
7. Em seguida, outra lambda é acionada. Iniciando um cluster EMR, posteriormente, jobs são adicionados para o processamento dos arquivos.
8. Novas tabelas são geradas e armazenadas no formato Parquet no bucket (S3) Trusted Zone.
9. Os dados presentes nessa área já passaram por processamentos e transformações necessárias para serem utilizados, assegurando sua qualidade e tornando-os precisos e confiáveis.
10. Consultas SQL realizadas pelo Athena.
11. Dados são carregados para o Power BI, para a criação de dashboards.
12. Dashboards e relatórios são acessados por outras áreas/clientes.

