AWS Data Pipeline and ML Architecture

logs/métricas

Amazon S3 Bucket (Bronze)
raw_data_customers.csv
modelo_churn_vX.tar.gz

AWS Glue Data Catalog
churn_db (Base de datos)

AWS Glue Jobs

Job 1: Bronze → Silver (Limpieza)

Job 2: Silver → Gold (Features)

Amazon EventBridge (Scheduling)
schedule
AWS Step Functions (Orquestación del pipeline)

Amazon CloudWatch (Logs y Métricas)

orquesta

Amazon S3 Bucket (Silver)
clientes_limpios/

Amazon S3 Bucket (Gold)
clientes_modelado/

Training Job (scikit-learn)

Amazon S3 Bucket (Modelos)
modelo_churn_vX.tar.gz

Amazon SageMaker

Endpoint (Inferencia)

Model Registry

orquesta

logs/métricas