TelcoPulse: Real-Time Network Metrics Dashboard

A real-time data pipeline to monitor and analyze network performance metrics using AWS-native services and display the results in a dynamic dashboard.
A streaming data pipeline that captures metrics like signal strength, network status, and GPS precision from various providers, processes the data in near real time, and delivers actionable insights to management via a live streamlit dashboard.

OBJECTIVES:

- Ingests streaming data from a Kinesis data stream.
- Processes and transforms the data using Apache Spark on AWS Glue.
- Stores the processed results in an S3-based data lake.
- Uses Glue Crawlers and AWS Athena to make the data queryable.
- Displays metrics in a Streamlit dashboard hosted on AWS ECS.

TOOLS:

- Amazon Kinesis - for ingesting real-time network metric streams
- AWS Glue (Spark Streaming) - for real-time data transformation
- Amazon S3 - as the destination data lake
- AWS Glue Crawlers - for cataloging transformed data
- Amazon Athena - for running SQL queries on the S3 data.
- Amazon ECS - to deploy a real-time Streamlit dashboard

KPIS REQUIREMENTS

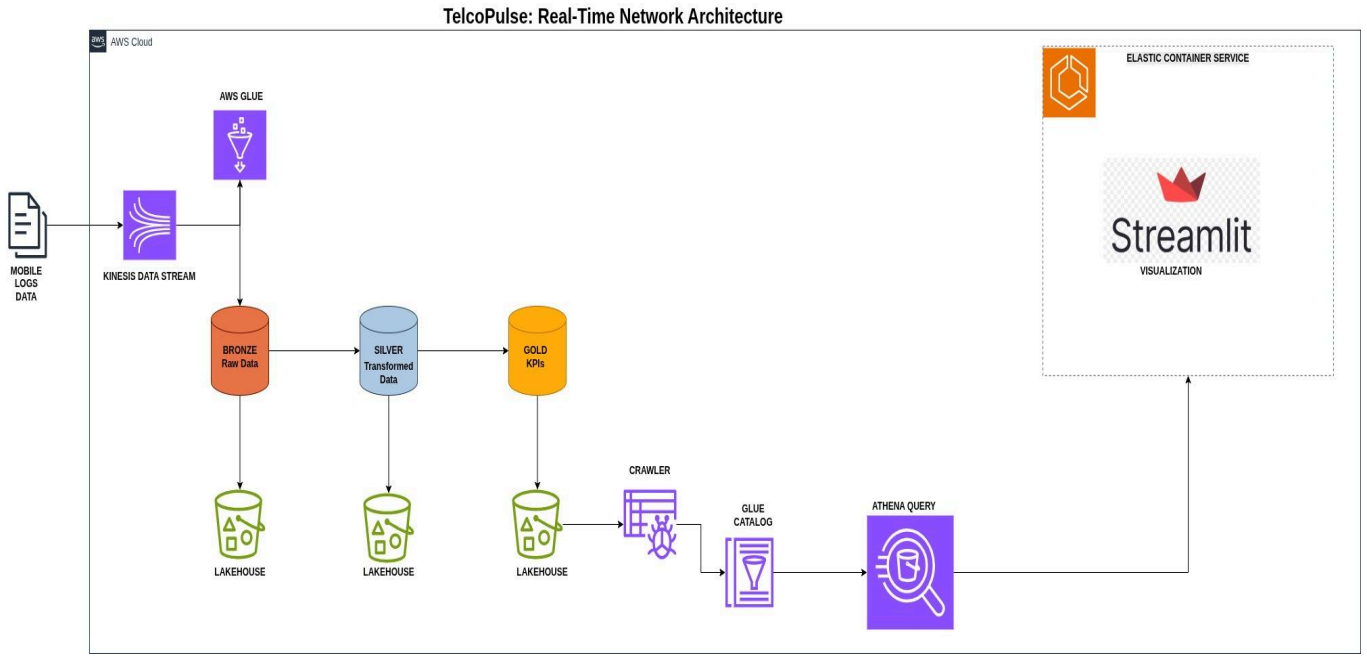
- Average Signal Strength per Operator
- Average GPS Precision per Operator
- Count of Network Statuses per Postal Code

DATA SCHEMA

COLUMN	DATA TYPE	Description
hour	timestamp	The date and time when the data point was recorded.
lat	double	Latitude coordinate representing the geographic location.
long	double	Longitude coordinate representing the geographic location.
signal	integer	Signal strength of the mobile or GPS device, typically measured in dBm.
network	string	Type of network connection
operator	string	Name of the network

		service provider or operator.
status	integer	Numerical status indicator (e.g., 0 = inactive, 1 = active).
description	string	Textual description of the current network or location context.
speed	double	Speed of the device in meters per second or kilometers per hour.
satellites	double	Number of satellites connected to the GPS receiver.
precission	double	Precision or accuracy of the location fix, typically in meters.
provider	string	Source of the location data (e.g., GPS, Network, Fused).
activity	string	Type of user activity detected (e.g., walking, driving, still).
postal_code	double	Postal code of the detected location. May be null or inaccurate in some cases.

ARCHITECTURE



Data Ingestion

- Mobile Logs Data: Source data from mobile devices containing information such as signal strength, location, speed, etc.
- Kinesis Data Stream (AWS): A real-time streaming service that ingests mobile log data as it is generated.

Data Lakehouse Tiers

The architecture follows a medallion architecture (Bronze → Silver → Gold), commonly used in modern data lakehouses:

Bronze (Raw Data)

- Stores unprocessed data from Kinesis.
- Persisted in an S3-based Lakehouse.
- Serves as the historical and audit layer.

Silver (Transformed Data)

- Processes and cleans the Bronze data using AWS Glue (ETL service).
- Adds structure, filters out noise, and handles missing values.
- Also stored in the Lakehouse**.**

Gold (KPIs)

- Final aggregated, enriched data layer.
- Stores Key Performance Indicators (KPIs) like network quality, activity patterns, etc.
- Used for querying and reporting.
- Also persisted in the Lakehouse

Metadata & Querying

- Glue Crawler: Scans the Gold data in S3 and populates the schema metadata into the Glue Data Catalog.
- Glue Catalog: Central metadata repository.
- Athena Query: Serverless query engine used to analyze Gold-layer data using standard SQL

Visualization

- Streamlit App (Hosted on ECS - Elastic Container Service)
- Streamlit reads the KPIs via Athena.
- Visualizes insights in dashboards for network analysts or business stakeholders.
- Deployed on ECS for scalability and containerized deployment.

Data Flow Summary

- Mobile Logs → Kinesis Stream
- Kinesis → Bronze (Raw) S3
- AWS Glue → Transform to Silver → Then Gold
- Gold → Cataloged by Glue Crawler
- Athena queries Gold KPIs
- Streamlit (on ECS) visualizes the results

Key Benefits

- Real-time data ingestion
- Scalable and serverless processing
- Structured medallion data layers
- Metadata management
- Interactive dashboards