

TelcoPulse: Real-Time Network Metrics Dashboard

Objective: Design and implement a real-time data pipeline to monitor and analyze network performance metrics using AWS-native services and display the results in a dynamic dashboard.

Project Context

You've joined the data engineering team at **TelcoPulse**, a fictional telecommunications analytics company that aggregates real-time data from multiple mobile network providers across the country.

Your job is to build 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 dashboard.

This project mirrors a **real-world streaming data use case** and is designed to help you build hands-on skills in **data ingestion**, **real-time transformation**, **data lake architecture**, and **dashboards**—all using **AWS-native tools**.

Your Mission

Build an end-to-end real-time analytics system that:

- 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 & Services You'll Use

- **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
 - **AWS Athena** – for running SQL queries on the S3 data
 - **Amazon ECS** – to deploy a real-time Streamlit dashboard
-

KPI Requirements (to be displayed on the dashboard)

You will compute and visualize the following metrics **in near real-time**:

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

Your dashboard should **auto-refresh at configurable intervals** (e.g., every 5 minutes) and allow users to view the latest available metrics.

Deliverables

- Spark Streaming job code on AWS Glue
- Glue Crawler + Athena setup
- Streamlit dashboard deployed on ECS
- README with architecture diagram and deployment instructions