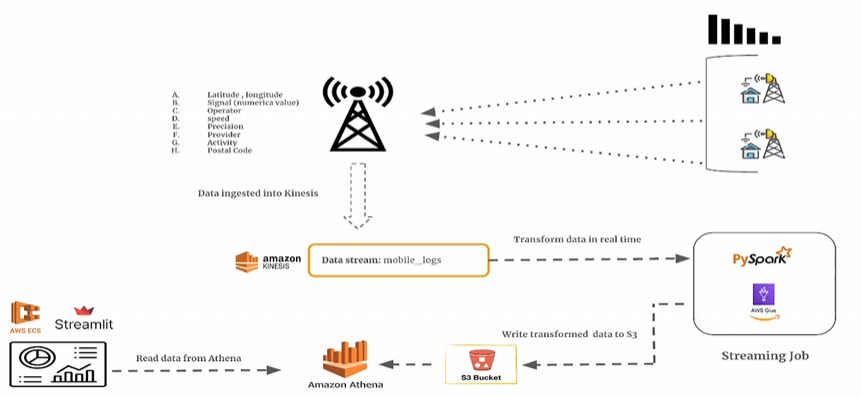
Real-Time Streaming Data Processing with PySpark and Kinesis

# Overview

This real-time streaming pipeline processes telecommunication network data, leveraging AWS Kinesis and PySpark. The system enables live tracking of signal strength, network provider analytics, and geographical insights.



# Key Features

✔️ \*\*Live Data Ingestion\*\*: Captures network logs in real time using AWS Kinesis.   
✔️ \*\*Real-Time Processing\*\*: Uses AWS Glue and PySpark Streaming for transformations.   
✔️ \*\*Efficient Storage\*\*: Stores processed data in an AWS S3 data lake.   
✔️ \*\*Instant Analytics\*\*: Queries data via Athena for insights.   
✔️ \*\*Interactive Dashboard\*\*: A web dashboard built with Streamlit and deployed on ECS.

# How It Works

🔹 \*\*Step 1\*\*: Data is ingested from mobile network logs into AWS Kinesis.   
🔹 \*\*Step 2\*\*: PySpark processes the data in real-time and writes output to S3.   
🔹 \*\*Step 3\*\*: AWS Glue crawlers catalog the data, making it queryable.   
🔹 \*\*Step 4\*\*: Athena enables SQL-based queries on transformed data.   
🔹 \*\*Step 5\*\*: A Streamlit dashboard visualizes metrics like average signal strength and GPS precision.