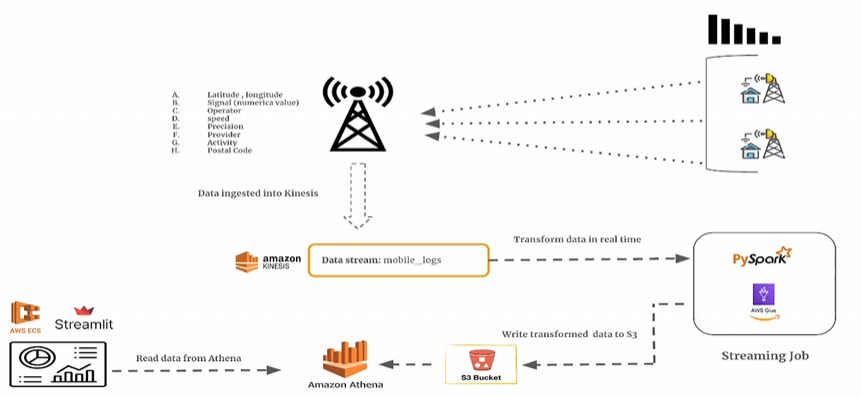
Professional Report: Real-Time Streaming Data Processing with PySpark and Kinesis

# Introduction

This report details the implementation of a real-time data streaming pipeline using AWS Kinesis and PySpark. The system is designed to process telecommunication network data streams and provide analytical insights using Athena and Streamlit dashboards.

# Architecture Overview

The architecture comprises multiple AWS services:   
- \*\*AWS Kinesis\*\* for real-time data streaming.   
- \*\*AWS Glue\*\* for PySpark-based streaming transformations.   
- \*\*S3\*\* as the storage layer for transformed data.   
- \*\*Athena\*\* for querying processed data.   
- \*\*Streamlit\*\* for real-time dashboards deployed on ECS.



# PySpark Streaming Job Implementation

The PySpark streaming job ingests data from AWS Kinesis, applies transformations, and writes the results to an S3 bucket for further analysis. The job includes:   
- Schema definition for the streaming data.   
- Conversion of raw JSON data into structured DataFrames.   
- Aggregations such as average signal strength and network status counts.   
- Watermarking to handle late-arriving data efficiently.

# Deployment Steps

1. \*\*Set up AWS Kinesis Data Stream\*\*: Create a Kinesis stream for ingesting network logs.   
2. \*\*Deploy PySpark Streaming Job on AWS Glue\*\*: Submit the PySpark script to AWS Glue as a Spark Streaming Job.   
3. \*\*Store Processed Data in S3\*\*: The output data is stored in partitioned Parquet format.   
4. \*\*Query Data with Athena\*\*: AWS Athena allows users to query and analyze the processed data.   
5. \*\*Deploy Streamlit Dashboard on ECS\*\*: A containerized Streamlit application is deployed on AWS ECS for real-time visualization.

# Conclusion

This implementation provides a scalable solution for real-time data transformation and analytics using AWS-managed services. The combination of Kinesis, PySpark, and Athena allows efficient data processing and visualization, making it suitable for high-frequency data ingestion scenarios.