Streaming Data Processing in Telecom Data Engineering

This document outlines the step-by-step approach for real-time streaming of telecom data using Apache Kafka, Databricks, and Power BI. The goal is to stream real-time network metrics such as signal strength and call drop rates, process them in Databricks, and visualize the data in Power BI.

# Tech Stack

1. Apache Kafka: Used for real-time stream ingestion of network performance metrics.  
2. Databricks: Used for processing streaming data with Structured Streaming.  
3. Delta Lake: Storage layer for real-time data with ACID properties.  
4. Power BI: Visualize streaming data in real-time dashboards.  
5. Python: Used for streaming ETL scripts.

# Step-by-Step Solution

## 1. Data Ingestion Using Kafka

Set up Apache Kafka to stream network performance data (e.g., signal strength, call drop rates) from various network devices. Kafka will act as the real-time ingestion layer for data sources such as network routers and devices.

## 2. Real-Time Data Processing in Databricks

Use Databricks Structured Streaming to process data as it arrives from Kafka. Apply business rules such as detecting network anomalies (e.g., signal drops or high call drop rates) and calculating metrics like average signal strength per region.

## 3. Storage in Delta Lake

Store the processed streaming data in Delta Lake. Use Delta Lake to manage real-time data with ACID transactions. This ensures data consistency while handling high-velocity data streams.

## 4. Sending Data to Power BI Using Streaming API

After processing the data, use the Power BI Streaming API to push real-time data to Power BI. Visualize metrics such as signal strength trends and call drop rates in real-time. Configure dashboards to update dynamically as new data is streamed into Power BI.

## Use Case: Real-Time Network Performance Monitoring

This use case involves streaming network performance metrics like signal strength and call drop rates. The data is processed in real-time and stored in Delta Lake. Power BI dashboards visualize the real-time metrics, providing insights into network health and potential outages.