

# Real-Time Network Traffic Analysis for Telecommunications

## Problem Statement

A telecommunications company wants to monitor its network traffic in real-time to identify any anomalies or patterns that could indicate issues or opportunities for improvement. The company has a large volume of network traffic data generated every second and needs to process this data in real-time. They also want to be able to visualize the data to provide insights to the network operation team.

We will develop a real-time network traffic analysis system using Apache Kafka and Structured Spark Streaming. The system will ingest and process network traffic data in real-time and identify any anomalies or patterns. The data will be visualized using a web-based dashboard that provides real-time insights into the network traffic.

## Hints

- Set up a Kafka cluster on Confluent Cloud and configure Kafka topics for ingesting network traffic data.
- Use Structured Spark Streaming to ingest data from Kafka topics and perform real-time analytics on the data.
- Implement stateless transformations such as select, filter, and groupBy to analyze the data in real-time.
- Use sliding window operations and window-based aggregations to identify any patterns or anomalies in the data.
- Visualize the data using a web-based dashboard such as Streamlit or Grafana.

## Guidelines

- Set up a Kafka cluster on Confluent Cloud and create two Kafka topics named network-traffic and processed-data.
- Implement a Python script using the kafka-python package to generate network traffic data and publish it to the network-traffic Kafka topic.
- Use Structured Spark Streaming to ingest data from the network-traffic Kafka topic and perform real-time analytics on the data.
- Implement a sliding window operation to identify patterns in the data, such as sudden spikes or drops in traffic.

- Use a window-based aggregation to identify any anomalies in the data, such as unexpected patterns or traffic from unusual sources.
- Publish the processed data to the processed-data Kafka topic.
- Use Grafana to visualize the processed data in real-time. Create graphs that show traffic trends, identify any issues, and provide insights to the network operation team.

You can use the following [sample code](#) to generate network traffic data and publish it to the Kafka topic. However, you'll need to write code for ingesting and processing the data in real-time.