2023-09-28 Group Therapy
Casper Kristiansson
Nicole Wijkman

Intensive Data Processing for Highway Camera Logs

The United States of America has an active proposal called 511 Travel Information Telephone Services [1]. In this proposal, each state (around 90%) has some type of service to provide traffic information to its citizens. Most of these states therefore provide a web application that allows a user to view highway cameras [2]. While each state doesn't allow public access to all cameras each state provides access to usually 1,000-5,000 cameras.

PKTraffic LLC. [3] is a company that uses these cameras to capture and collect traffic data more specifically incident data. Their system is developed to efficiently and accurately collect vast amounts of data with a low-cost solution, enabling it to broaden our impact and provide a sought-after service for multiple entities.

The issue with highway cameras is that a large majority of them receive issues like 404 not found, 503 service unavailable, and a lot of other errors. Because the company collects currently footage from around 2,000-2,500 cameras the logs generated from all of them are quite large. In their current situation, PKTraffic doesn't have any good system in place for storing the logs in a manner so that they can easily be accessed and managed.

The goal of the logs generated is both to keep track of why a specific camera wasn’t recording at a specific given time but also to be able to display the current active cameras. This means that they on the fly want to be able to visualize why a group of cameras from an incident was recording or not and why they weren't recording at the given time.

The use of cloud products especially AWS is crucial in this project. Due to the specific use cases with the querying of the data, DynamoDB is a perfect product to use. While this project will be limited to mostly structure and collect the specific cameras it will be crucial to be able to perform analytics in the future like the usage of Kinises or DynamoDB streams. The logs generated by the programs come from ffmpeg which decodes m3u8 streams. While these do generate a lot of other information only the connection of the request will be tracked (start the stream, stop the stream, and reason why it stopped).

The most difficult of this process will be to build a system in the best way possible to structure the data and store it in DynamoDB. Because the querying of the data in DynamoDB is limited (querying is used on the pk and sk or Secondary index) it is crucial to structure the data in the correct way. Because there will most likely be multiple writes to the database every second building a solution to optimize the storing and writing is crucial to keep costs low.

The entire pipeline of the project technologies will look like this:

Data Collection

- Highway CCTV Cameras decoded using ffmpeg
- HTTP(s) requested performed towards the streaming server

Stream Processing & Data Ingestion

- Amazon Kinesis Data streams for processing the data
- Amazon Lambda to perform further analytics and save the data

Data Storage

- Amazon DynamoDB to store the data

Visualization

- React.js and other external libraries to integrate new visualization into an existing dashboard

References

[1] https://ops.fhwa.dot.gov/511/
[2] https://fl511.com/
[3] https://pktraffic.com/