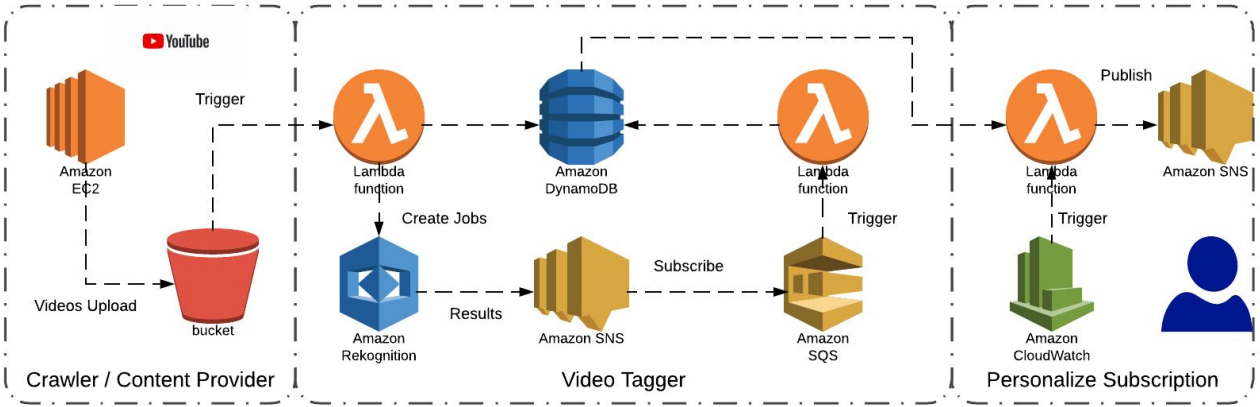
Video Classification & Delivery System

ELEN 6770 Next Generation Network
Haoran Guo hg2461 Hao Fang hf2345

Overview and Structure



Target: Crawl videos from Youtube, classify them using AWS, and deliver them to particular users.

1. Crawler System: Crawl videos from Youtube, videos delivered to S3
2. Video Tagger: Use AWS Rekognition to classify the videos
3. Video Delivery: Users receive e-mail or messages about the newly classified videos.

Video Crawler

:method	GET
:authority	www.youtube.com
:scheme	https
:path	/get_video_info?html5=1&video_id=aOF7_CNeSm4&cpn=RTb2akg2vbqqzLYu&eurl&
x-youtube-page-label	youtube.ytfe.desktop_20181205_5_RC2
user-agent	Mozilla/5.0 (Macintosh; Intel Mac OS X 10_14_1) AppleWebKit/537.36 (KHTML, like G
x-youtube-variants-checksum	fbda1bfc8c31c4e2d3d2e0c89188bb24
x-youtube-page-cl	224297709
x-youtube-utc-offset	-300
x-youtube-client-name	1
x-youtube-client-version	2.20181206
x-youtube-identity-token	[redacted]
accept	*/*
x-client-data	[redacted]
referer	https://www.youtube.com/watch?v=[redacted]
accept-encoding	gzip, deflate, br
accept-language	en-US,en;q=0.9,zh-CN;q=0.8,zh-TW;q=0.7,zh;q=0.6

Step 1. Packet analyzing

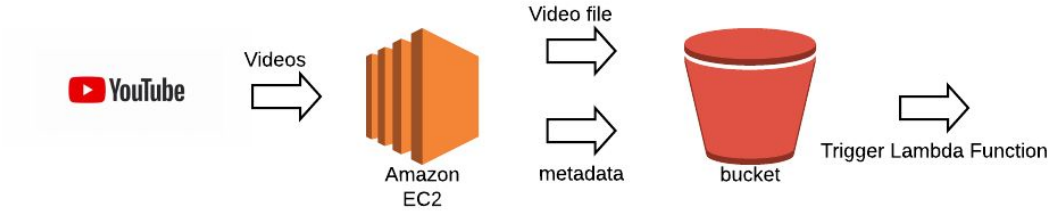
Tools: Charles Proxy & Postman

Steps: Analyze the packet browser sent to Youtube server

Filter out the essential parameters in headers and request body

Essential Parameters: **User-Agent** in headers, **video_id** in request body

Video Crawler



Step 2. Video Crawling

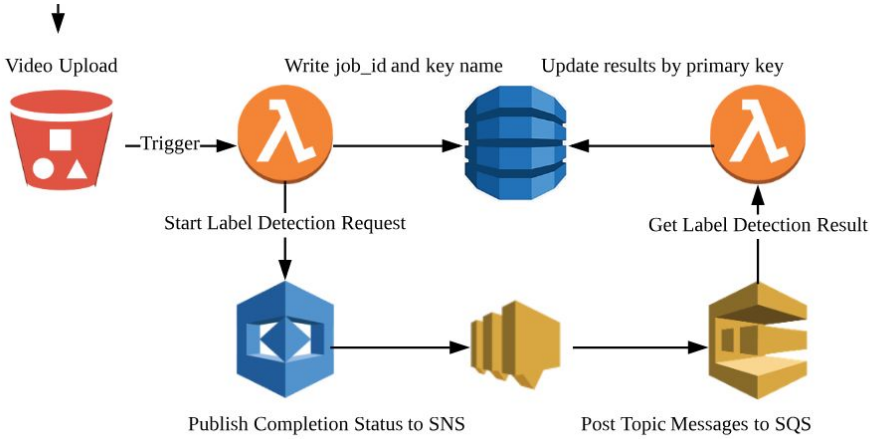
Parse the html and locate the video address, download it to EC2 instance

Step 3. Video Delivered to S3 buckets

Videos chunked into pieces and uploaded to S3 bucket in order.

Write metadata of the corresponding video object to S3

Video Processor



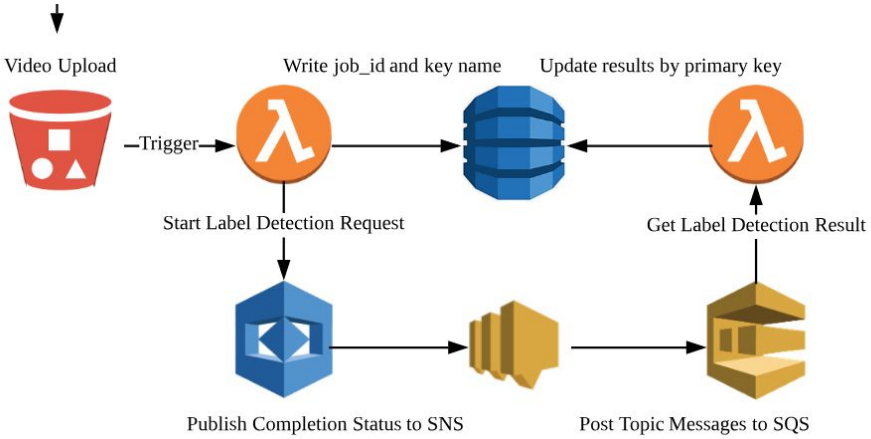
Step 1. Start Label Detection

Extract bucket and key name from S3 event trigger, call AWS Rekognition Video API, store response job id into dynamodb

Step 2. Monitor SNS & SQS

Rekognition API published the completion status of video analysis to SNS topic. SQS subscribe to SNS for receiving messages.

Video Processor



Step 3. Get Analysis Result

Lambda gets completion status and job id from SQS message queue. If succeed, pass job id to Rekognition API to get final analysis results,

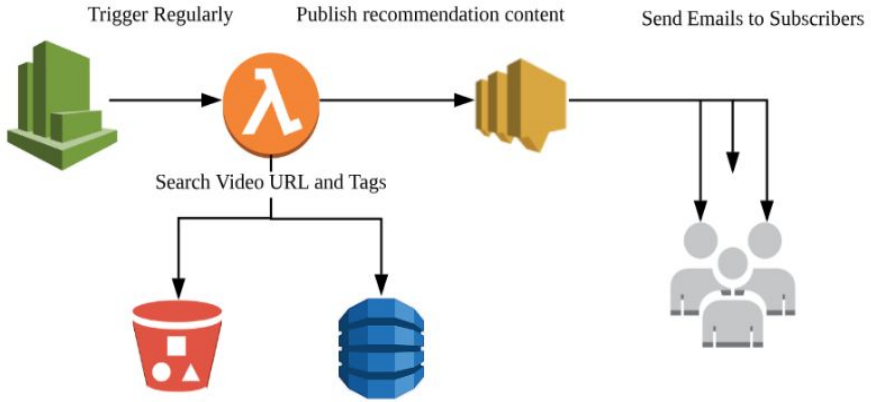
Step 4. Update dynamodb item

Use unique job id to locate item in dynamodb, update result labels.

Regular Classification and Delivery

CloudWatch trigger Lambda function daily

Lambda function would scan DynamoDB table and filter out the video users may be interested in.



Publish recommendation results to SNS topic.

Users who subscribed the corresponding topic would receive message or e-mail from SNS.

Demo