Twitter Realtime Sentiment Analysis

5th June, 2018

Patrick Schneider

Introduction and Use Cases

Twitter

Faster communication channel than traditional news outlets

Twitter Sentiment Analysis

Realtime insight about peoples mood

Location based info

Breaking down insight on location

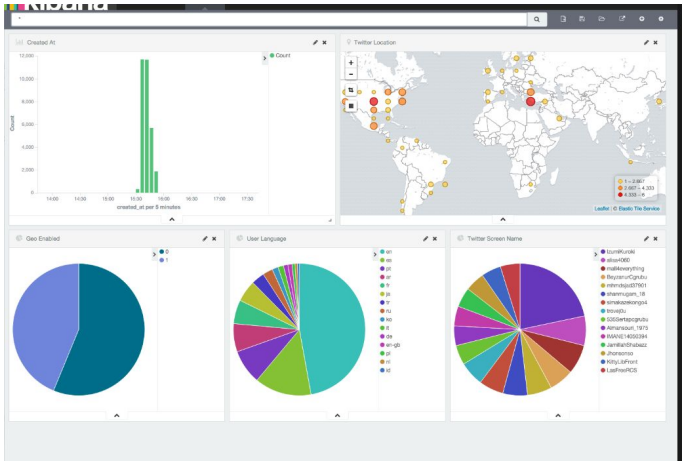
Introduction and Use Cases

Use cases

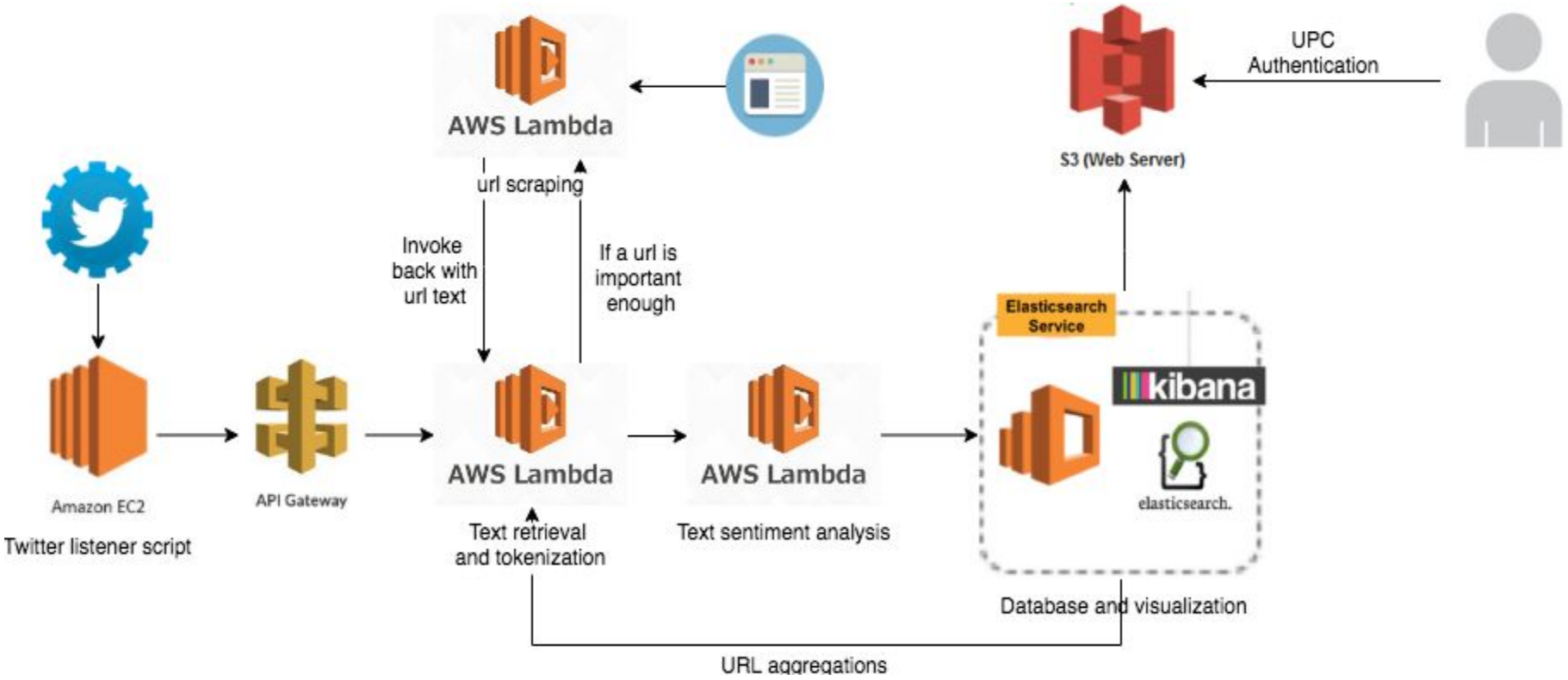
- Customer Service
    - Satisfaction Level
    - Identification of unexpected issues
    - Fast service support

- Marketing Campaigns
    - Activity measuring
    - Feedback

- Stockmarket trends
    - Insight on buy/sell incentive

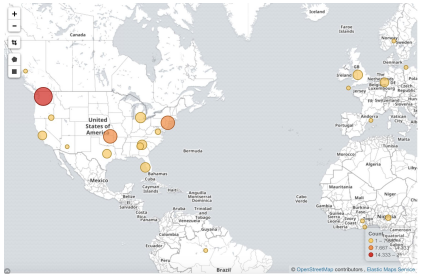


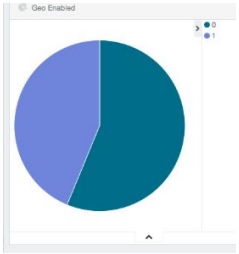
Architectural Design



Resulting PoC

User can display tweets origin on a spatial map based on tweet coordinates

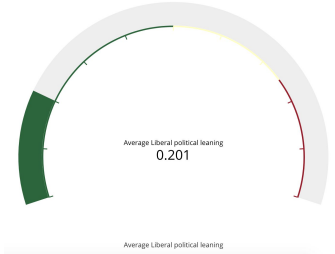




View tweet sentiment for a given topic

Over time development of sentiment and activitiy

...Political affiliation, detailed mood, tag cloud etc...



Initial plan vs current state

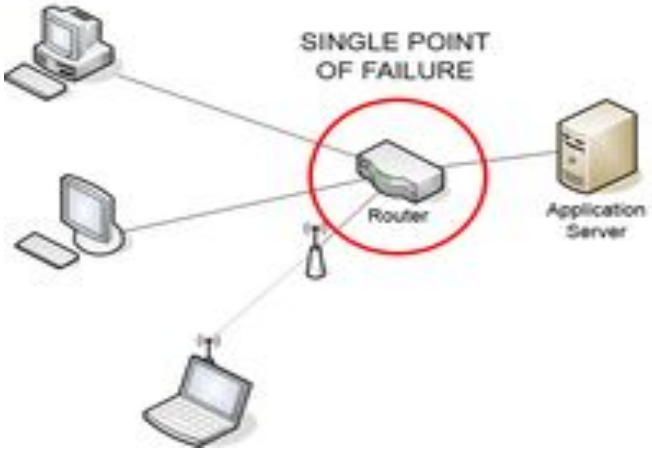
- S3 Bucket and kinesis firehose removed and replaced with Elasticsearch
- The Sentiment API calls were outsourced in a lambda function
- Google Maps API call for coordinates lookup
- The application data flow was changed to the one shown before, with three lambdas used
- The architecture was simplified as we learnt the full potential of the tools we were using.

Existing Dependencies

- TwitterListener (python 3.6)
  - tweepy, json, boto3, requests
- Sentiment Analysis Lambda function (python 2.7)
  - indico.io, boto3, json
- Tokenizer lambda function (python 3.6)
  - NLTK, requests_aws4auth, processTweet, requests, json, re
- Scraper lambda function (python 3.6)
  - requests_aws4auth, requests, json, re, BeautifulSoup

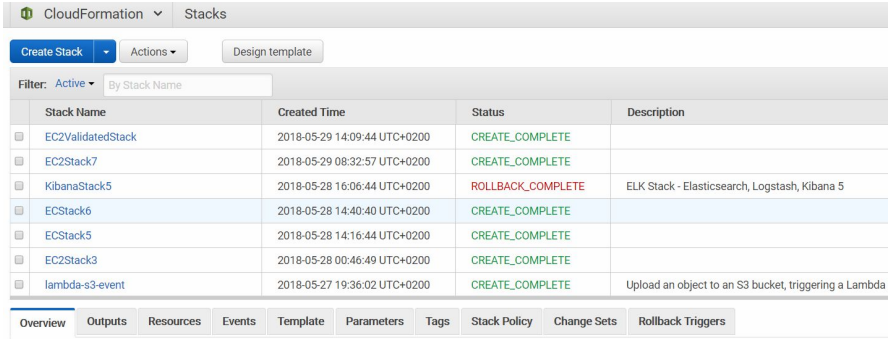
Any single point of failure

Since the architecture components are distributed, no single point of failure exist.



Automated Cloud Infrastructure

- CloudFormation
  - Code as Infrastructure
  - Backup and migration template
- Infrastructure configured as templates
  - EC2 instance
  - Lambda functions
  - Kibana & Elastic Search



The main challenges and solutions

Listener

- Retrieving more coordinates -> Google Maps API

Lambdas

- URL scraping -> Extract title and description from meta

Elasticsearch & Kibana

- Coordinates in Kibana -> Manually set geo-point mapping
- Creation of visualization -> Research and experimenting

Others

- CloudFormation -> Connection of the AWS Components