

MR. TRUMP AND MR. BIDEN TWITTER SENTIMENT ANALYSIS.

QUEENSLAND UNIVERSITY OF TECHNOLOGY



Roberto Carlos da Silva Junior

Student ID: n10374647

Master of Information technology

CONTENTS

1.	Introduction.....	1
2.	Technical description	2
2.1	Modules	2
2.2	APIs and Technologies Utilised	2
3.	User cases.....	3
4.	Technical breakdown	4
4.1	Architecture	4
4.2	Aws scaling info	4
4.2.1	Launch configuration.....	4
4.2.2	Auto scaling group.....	5
4.2.3	Load balancer	5
4.3	Scaling & Performance.....	6
5.	Test Plan & Results.....	6
6.	References	6

1. INTRODUCTION

The software developed as part of this assessment provides a resource for users to follow a comparative sentiment analysis from the American presidential candidates Mr. Trump and Mr. Biden on Tweets coming in from all around the world in real-time. The service automatically identifies and categorizes opinions expressed in a tweet to determine

whether the opinion expressed about Mr. Trump or Mr. Biden is positive, negative, or neutral. And then, demonstrate as percentage of positive, negative, and neutral comments in a Real-time pie chart, which automatically updates every three minutes, which causes scaling on the AWS instances.

2. TECHNICAL DESCRIPTION

2.1 MODULES

The application developed contains four modules. That is Aws, Sentiment, Stream, and Twitter. The module Aws is a bridge between the application and AWS DynamoDB, and it is the model responsible for uploading and downloading data in the database and verify whether a tweet exists as data in the database. The module Twitter connects to the Twitter endpoint returning the Twitter API about the American presidential candidates. The module Stream converts the Twitts API in JSON in save those in array to be understood by the module Sentiment, which performs the sentiment analysis of both candidates and calculates the percentual of good, bad, or neutral comments of both to be pop up on the screen as two pie charts representing each candidate.

2.2 APIs AND TECHNOLOGIES UTILISED

NodeJS [1]

It is an open-source and cross-platform JavaScript runtime environment. It is used to handle serving files as well as send requests and receive responses from APIs and Databases used in the Tweet Stream, Sentiment Analysis Processing, and Application Client section.

Express JS [2]

A web framework that sets the routes and views, allowing the client and server-side to be integrated, which allows for rapid development. It is used in the Sentiment Analysis processing and Application Client.

Twit [3]

Twitter's streaming API is utilized to gather data for sentiment analysis on tweets. It connects to Twitter's endpoint that provides tweets in real-time about Mr. Trump and Mr. Biden throughout HTTP requests.

Sentiment [4]

A Node.js module that performs sentiment analysis for the application once the twits are already collected, using AFINN-165-word list and Emoji Sentiment Ranking to analyze the sentiment ranking for arbitrary text blocks input.

AWS-SDK

A JavaScript library that enables developers to use AWS services in node.js. It is used as a bridge between the application and DynamoDB, permitting the upload and download of tweets.

Google Charts

Google Charts is a JavaScript API that allows structured data to be displayed in a visual diagram. It is used to display the sentiment analyzes of both presidential candidates in a pie chart.

AWS Elastic Load Balance

An AWS Elastic Load Balancer automatically distribute incoming application traffic across multiple targets. Tweetstorm utilises a load balancer in front of the Auto Scaling Group, this handles the large amounts of traffic coming in from the Twitter Stream to the Sentiment Analysis Servers and balances the traffic.

AWS Auto-Scaling Group

An AWS Auto-Scaling Group allows EC2 instances to be created dynamically based on defined conditions, this was utilised to balance the required number of instances in order to provide an application that is consistent despite the volume of task being completed.

JavaScript

This client-side of Javascript enables interactive features and dynamic functions. It is used to count the emotions of tweets and displays the overall sentiment.

Bootstrap

Most popular HTML, CSS and JavaScript framework used to develop simple and consistent UI for web page. It contains the HTML and CSS based design template for typography, forms, buttons, navigation and other interface components. The web pages in this services are styled using Bootstrap components.

Docker

Docker was utilised to encapsulate the application into a container for easy download and access. The figure below depicts the Dockerfile utilised for the containerisation of Tweet Your SentimentThe build is derived from version 10 of node. The source code is then copied into the base directory of the image along with setting the work directory for that source code. The *RUN* command is then performed to install the required Node packages for the Docker image. Finally, port 3000 is exposed and the necessary commands are utilised to run the application.

AWS DynamoDB

A no SQL database service that stores and retrieves any amount of data, and serve any level of request traffic.

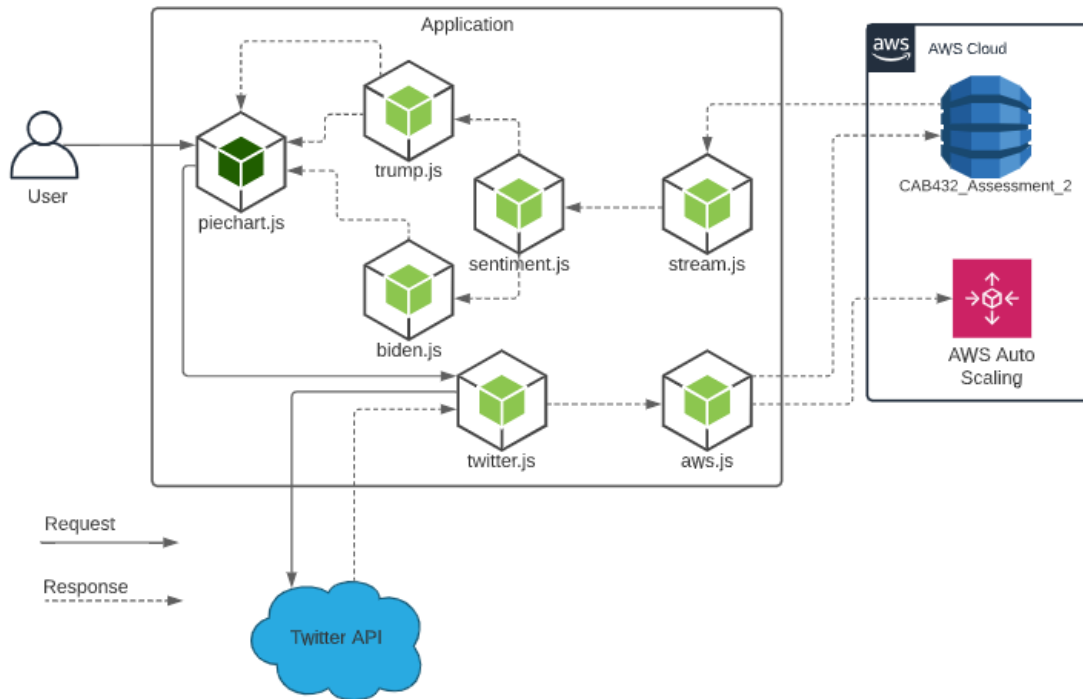
3. USER CASES

1. As a user, I want to view two graphs, one showing the sentiment analyse of Mr. Trump and the other showing the sentiment analyse of Mr. Biden from twitter. So that I can compare both American candidate's sentiment analyse.

- As a user, I want that each of the graphs updates by itself everyone three minutes. So that I can have a sentiment analyse view in Live time.

4. TECHNICAL BREAKDOWN

4.1 ARCHITECTURE



dyn

4.2 AWS SCALING INFO

4.2.1 Launch configuration

Launch configuration: Launch10374647

Details		
AMI ID ami-030bf1210a5023ee9	Instance type t2.micro	IAM instance profile -
Kernel ID -	Key name -	Monitoring false
EBS optimized false	Security groups sg-032bd1f8cf77dbb9	Spot price -
Create time Fri Oct 30 2020 12:23:01 GMT+1000 (Australian Eastern Standard Time)	RAM disk ID -	IP address type Default
Metadata accessible -	Token hop limit -	Metadata version -
User data -		

Copy launch configuration

4.2.2 Auto scaling group

EC2 > Auto Scaling groups > ASG10374647

Details

Activity

Automatic scaling

Instance management

Monitoring

Instance refresh

Group details

Desired capacity

1

Minimum capacity

1

Maximum capacity

3

Auto Scaling group name

ASG10374647

Date created

Fri Oct 30 2020 12:37:45 GMT+1000 (Australian Eastern Standard Time)

Amazon Resource Name (ARN)

arn:aws:autoscaling:ap-southeast-2:901444280953:autoScalingGroup:c8685263-0aee-45a9-8d75-60fb540e8093:autoScalingGroupName/ASG10374647

Launch configuration

Launch configuration

Launch10374647

Instance type

t2.micro

Storage (volumes)

/dev/sda1

View details in the launch configuration console

AMI ID

ami-030bf1210a5023ee9

Key pair name

-

Security groups

sg-032bd1ff8cf77dbb9

Create time

Fri Oct 30 2020 12:23:01 GMT+1000 (Australian Eastern Standard Time)

4.2.3 Load balancer

aws

Services

1. Configure Load Balancer

2. Configure Security Settings

3. Configure Security Groups

4. Configure Routing

5. Register Targets

6. Review

Step 6: Review

Please review the load balancer details before continuing

Load balancer

Name

LB10374647

Scheme

internet-facing

Listeners

Port:80 - Protocol:HTTP

IP address type

ipv4

VPC

vpc-007bab53289655834 (aws-controltower-VPC)

Subnets

subnet-05a3b8177138c8b14 (aws-controltower-PublicSubnet1), subnet-075811427d5564cf9 (aws-controltower-PublicSubnet2), subnet-04ca053dcbe5f49cc (aws-controltower-PublicSubnet3)

Tags

qut-username:n10374647

Security groups

Security groups

sg-078997505ad1c6bbc, sg-032bd1ff8cf77dbb9

Routing

Target group

Existing target group

Target group name

TG10374647

Port

3000

Target type

instance

Protocol

HTTP

Protocol version

HTTP1

Health check protocol

HTTP

Path

/

Health check port

traffic port

Healthy threshold

5

Unhealthy threshold

2

Timeout

5

Interval

30

Success codes

200

Targets

4.3 SCALING & PERFORMANCE

The scaling is possible through the load generated via processing the tweets and storing the tweets in the DynamoDB. A retrieving function to get the exact tweet that was just stored is also implemented in the Sentiment Analysis server to generate more load to be safe. An example of the metrics is shown by the image bellow.

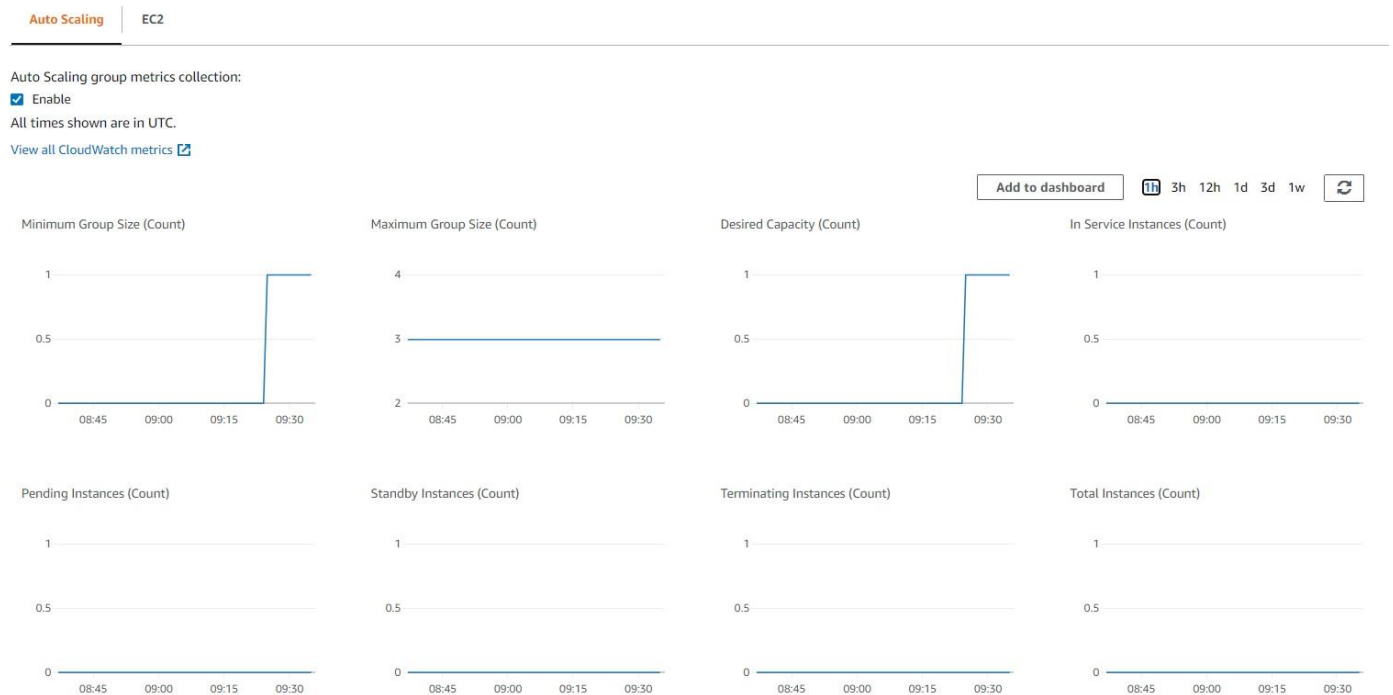


Figure 1 - Auto scaling metrics.

5. TEST PLAN & RESULTS

Purpose	Expected Outcome	Result
Access port 3000 of local network.	Application loads the home page.	PASS
Clear Real-Time chart	Real-Time line chart is cleared	PASS
Load Balancer		PASS

6. REFERENCES

- [1] N. Foundation, "Node.js", Node.js, 2017.
[Online]. Available: <https://nodejs.org/en/>. [Accessed: 10- Nov- 2020].
- [2] Express, "Express - Node.js web application framework", Expressjs.com.
[Online]. Available: <https://expressjs.com>. [Accessed: 10- Nov- 2020].
- [3] T. Tezel, "twit", npm, 2017.
[Online]. Available: <https://www.npmjs.com/package/twit>. [Accessed: 10- Nov- 2020].
- [4] A. Sliwinski, "sentiment", npm, 2017.
[Online]. Available: <https://www.npmjs.com/package/sentiment>. [Accessed: 10- Nov- 2020]