Created a Network Infrastructure on AWS

For Hosting Tweet Mapper Web Servers

Dr. Reza Eftekari
Cloud Computing, ENTS 649B
University of Maryland
College Park, Maryland

Sahil Khanna
University of Maryland
College Park, Maryland
114840134

Naveen Achyuta
University of Maryland
College Park, Maryland
114854452

Susmita Pradhan
University of Maryland
College Park, Maryland
114932860

Laveena Dulani
University of Maryland
College Park, Maryland
114924216

Apeksha Chauhan
University of Maryland
College Park, Maryland
115035027

ABSTRACT

The project concentrates on the significance of cloud computing in the IT industry. It mainly focuses on Amazon Web Services cloud to demonstrate cloud services. The main aim of the project is to explore various services of AWS to create a secure, scalable, and efficient cloud network to support various industry applications. Using this infrastructure, we have hosted web servers for Tweet Mapper application, to demonstrate the capabilities of the infrastructure.

OBJECTIVE

Create a network infrastructure on Amazon Web Service cloud using Amazon Elastic Compute Cloud, Auto Scaling, Elastic Load Balancing, Identity and Access Management, Amazon Virtual Private Cloud, Elastic IPs, Amazon Simple Storage Service, and Amazon Elastic Block Store.

INTRODUCTION

Tweet Mapper is a concept to map tweets on the google maps in real time to find location specific performance of the product. The application utilizes the big data generated by social network website like Twitter to derive consumer behavior towards the product. This sentiment analysis can be utilized in various applications such as target marketing, elections polling, disaster management and personalize recommendations. To host this application, we have created an infrastructure on AWS which involves private network, database, load balancer, auto scaling, user directory, VPN tunnel, storage repositories and queuing service.

The Tweet Mapper application performs sentiment analysis of the tweet using Natural Language Processing (NLP) techniques. In the application, twitter Api is integrated to the private network using Amazon's Simple Queue Service (SQS), which forms a reliable buffer of the data stream and pushed the data into the database. Amazon Elastic Search is incorporated as a database for its quick search capabilities. To map the results of the analysis on map in real time, we integrated Google Map Api which runs in a project, created on Google Compute Engine. The application uses all the services mentioned in the project as a system infrastructure to host web servers of the application on AWS.

The cycle of application begins by retrieving data from the twitter stream and queuing the data in simple queuing service of Amazon by executing python scripts on Windows EC2 instance. Once the data is queued in the SQS, NLP techniques is applied on the data and the result is stored in Elastic search. The web servers refer data from elastic search in real time to map the results on google maps. The web servers are hosted on two Linux EC2 instances, which are integrated into a private network using VPC service of Amazon. The http traffic is distributed among both the web servers using Elastic Load Balancer. The infrastructure is scaled automatically using Amazon's Auto Scaling service for the seasonal demand. We are using OpenVPN to create a private tunnel to the instances, and using Amazon's Cloud Watch and S3 for monitoring of EC2 instance performance and storing logs.
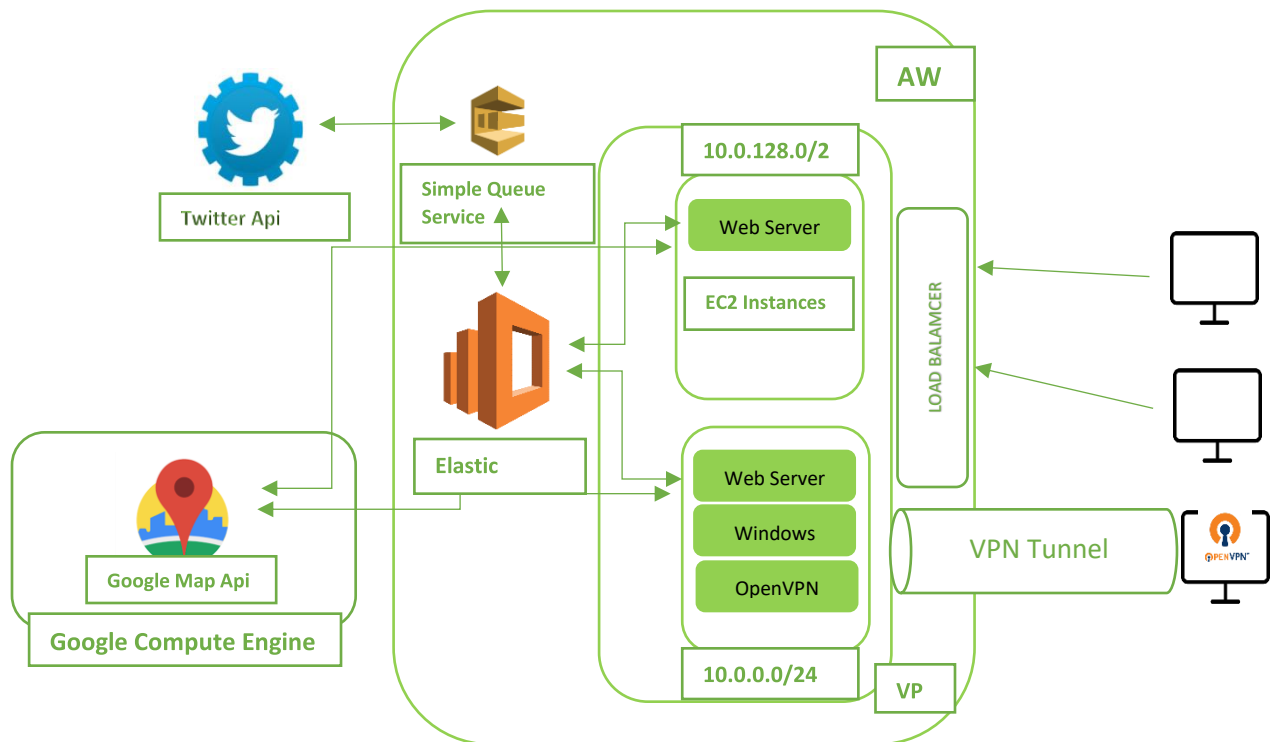
Figure 1 Network Architecture for Tweet Mapper

APPLICATION ARCHITECTURE'S SERVICES DESCRIPTION

AWS Identity and Access Management (IAM)

AWS Identity and Access Management (IAM) enables you to securely control access to AWS services and resources for the users. IAM service can create and manage AWS users and groups, and use permissions to allow and deny their access to AWS resources.

Figure 2 Amazon IAM Dashboard to manage users

Advantages of IAM

- It offers interactive management of user directory which can create users, assign individual security credentials, and request temporary security credentials. It can even help to specify permissions based on services.

Challenges/Learnings

- It is very important to secure the main root key, it is advised not to make a root access key. If you are making an access key, hide it while making versions on public repository or cloning the code. During the process of application development, someone get an access to our root key and we were charged with $1355 bill overnight because of the hacker's malicious activities. Fortunately, amazon web service had notified us for this malicious attack and credited the bill.

Amazon Elastic Compute Cloud

Amazon Elastic Compute Cloud (Amazon EC2) provides scalable computing capacity in the Amazon Web Services (AWS) cloud. Using Amazon EC2 eliminates your need to invest in hardware up front, so you can develop and deploy applications faster. You can use Amazon EC2 to launch as many or as few virtual servers as you need, configure security and networking, and manage storage.
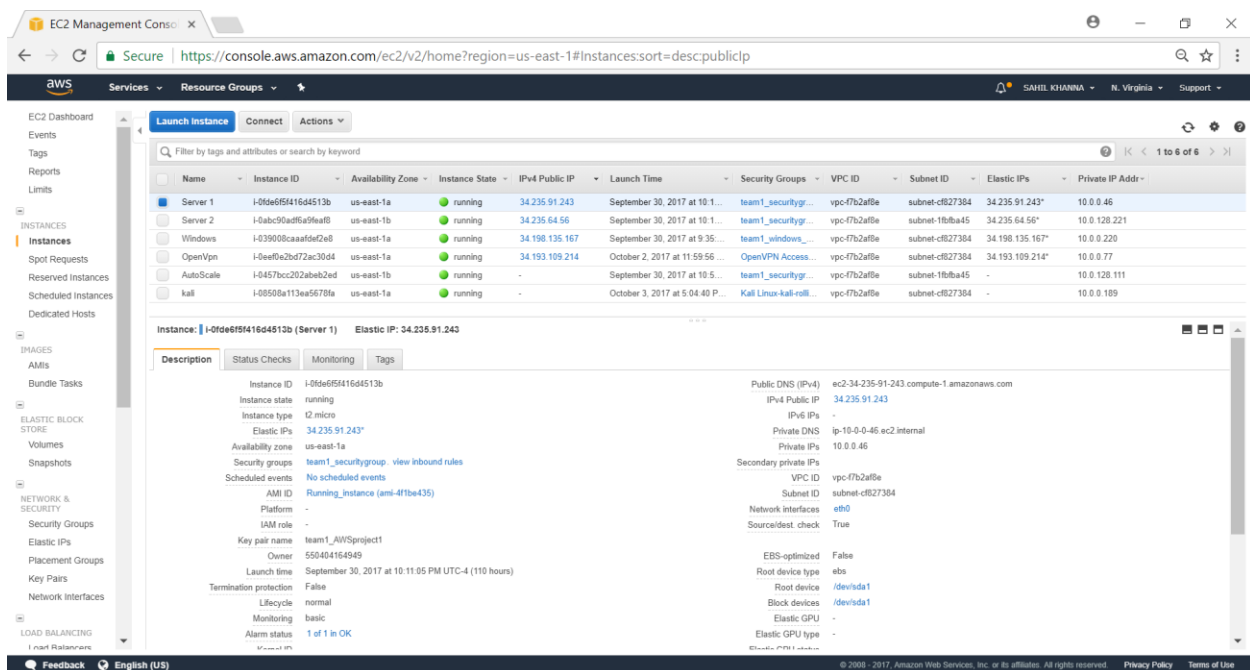

Figure 3 Amazon EC2 Dashboard

Advantages of using Amazon EC2

- Preconfigured templates for your instances, known as Amazon Machine Images (AMIs)
- Various configurations of CPU, memory, storage, and networking capacity for your instances
- Secure login information for your instances using key pairs

- Multiple physical locations for your resources categorizes as Regions and Zones
- A firewall that enables you to specify the protocols, ports, and source IP ranges
- Static IPv4 addresses for dynamic cloud computing, known as Elastic IP addresses

Challenges/Learnings
- Launched EC2 instance with wrong configuration.
- The configuration needed before launching an instance is very comprehensive.

Amazon Virtual Private Cloud

Amazon Virtual Private Cloud (Amazon VPC) enables you to launch Amazon Web Services (AWS) resources into a virtual network that you've defined. This virtual network closely resembles a traditional network that you'd operate in your own data center, with the benefits of using the scalable infrastructure of AWS.
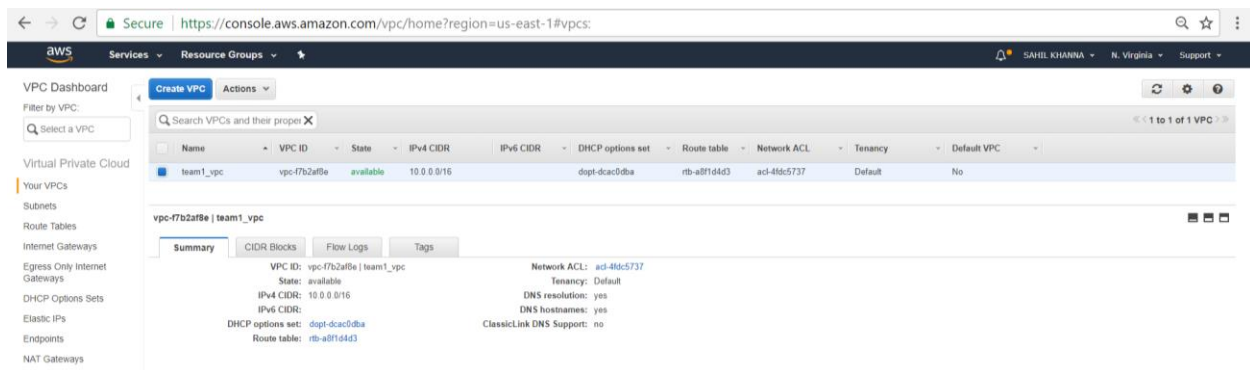
Figure 4 Amazon VPC Dashboard with one private network for the application

Advantages of using Amazon VPC
- Assign static private IPv4 and IPv6 addresses to your instances that persist across starts and stops
- Define network interfaces, and attach one or more network interfaces to your instances
- Add an additional layer of access control to your instances in the form of network access control lists
- Run your instances on single-tenant hardware

Challenges/Learnings
- Wrong subnet configuration initially because load balancer required two subnets
- Security configuration is comprehensive because we needed to provide access to multiple services

Amazon Simple Storage Service (S3)

Amazon S3 is object storage, built to store and retrieve any amount of data from anywhere. It is designed to deliver 99.99% durability. S3 provides comprehensive security and compliance capabilities that meet even the most stringent regulatory requirements. It allows user to manage their data for cost
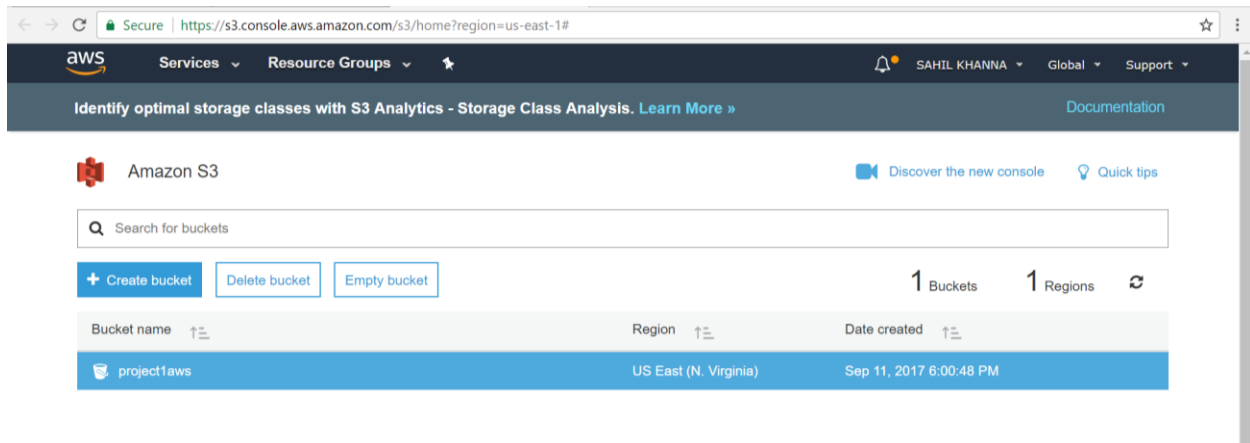
optimization, access control and compliance.

Figure 5 Amazon S3 Dashboard with one bucket created for the application

Advantages of using S3

- Durability, reliability and scalability.
- Security and compliance capabilities.
- It offers flexible management.
- It provides query in place feature.
- It is very cheap and provide easy data transfer.

Elastic Block Storage

EBS allows you to create storage volumes and attach them to Amazon EC2 instances. Once attached, it can be used to create file system on top on these volumes, run a database. EBS volumes are placed in a specific availability zone, where they are automatically replicated to uses from the failure of single component. It offers very low latency for read and write operations.

Elastic Volumes is a feature of Amazon EBS that allows users to dynamically increase capacity, tune performance, and change the type of live volumes with no downtime or performance impact. This allows users to easily right-size your deployment and adapt to performance changes.

Figure 6 Amazon EBS Dashboard displaying the integrated drives for the application servers

Advantages of using EBS

- Elastic Volumes allows users to dynamically increase capacity, tune performance, and change the type of any new or existing current generation volume with no downtime or performance impact.
- Amazon EBS volumes are designed to be highly available and reliable. Amazon EBS volume data is replicated across multiple servers in an Availability Zone to prevent the loss of data from the failure of any single component.
- Amazon EBS encryption offers seamless encryption of EBS data volumes, boot volumes and snapshots, eliminating the need to build and manage a secure key management infrastructure.

Challenges/Learnings

- It does not lock EC2 and EBS accessible data, thus we lost configurations during initial setups.

Classic Load Balancer

Classic load balancer provides load balancing across multiple EC2 instances and operates at both the request level and connection level. Classic load balancer is intended for applications that were built within the EC2-Classic network.
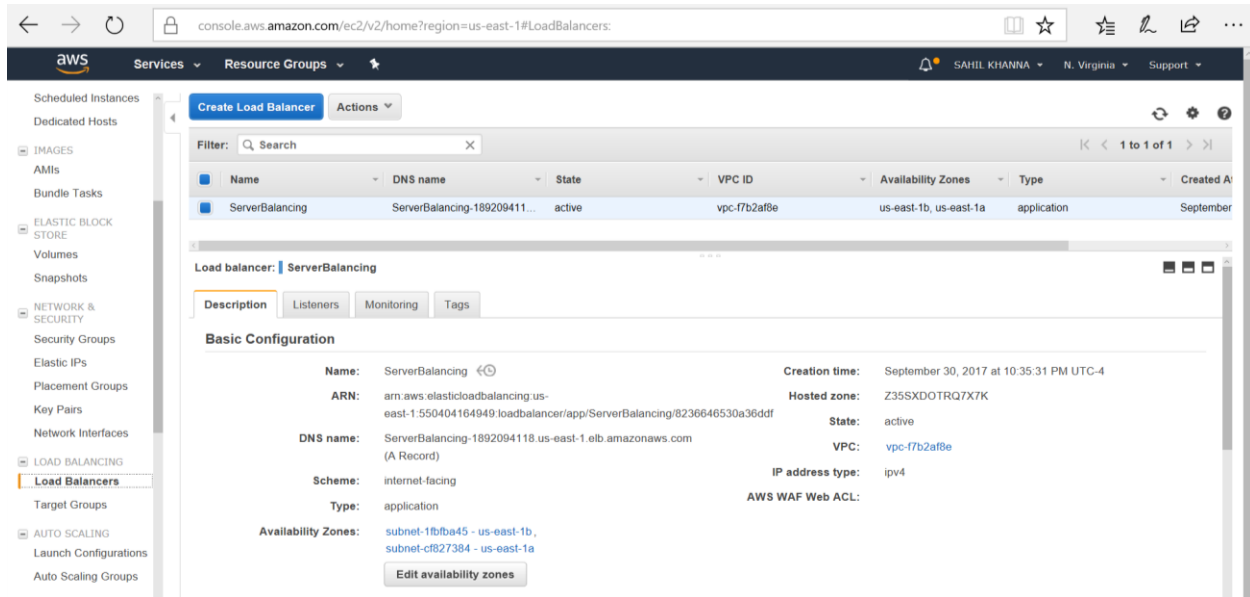
Figure 7 Amazon Load Balancer Dashboard displaying the load balancer instance created for the application

Advantages of Classic Load Balancer

- Incoming traffic across EC2 instances in a single or multiple availability zones can be distributed using load balancing. It automatically scales its request handling capacity in response to incoming traffic.
- Classic Load Balancer can detect the health of Amazon EC2 instances. When it detects unhealthy EC2 instances, it no longer routes traffic to those instances and spreads the load across the remaining healthy instances.
- Classic Load Balancer supports the ability to stick user sessions to specific Amazon EC2 instances using cookies. Traffic will be routed to the same instances as the user continues to access your application.

Challenges/Learnings

- ELB configuration was tough in initial stages because it requires subnets in different zones
- Network connectivity between web servers and elastic load balancer requires correct configuration
- CertificateNotFound: when there is a delay in propagating the certificate to all regions when it is created using the AWS Management Console. When this delay occurs, the error message is shown in the last step in the process of creating the load balancer.
- OutofService (A transient Error Occurred): There is a transient internal problem within the Elastic Load Balancing service or the underlying network. This temporary issue might also occur when Elastic Load Balancing queries the health of the load balancer and its registered instances.

Amazon Cloud Watch

Amazon Cloud Watch is a monitoring service that enables us to collect metrics and monitor them, analyze system behavior, create alarms, monitor bills and create custom metrics. We can create Dashboards to collectively monitor services in a single widget. Cloud Watch also provides information on

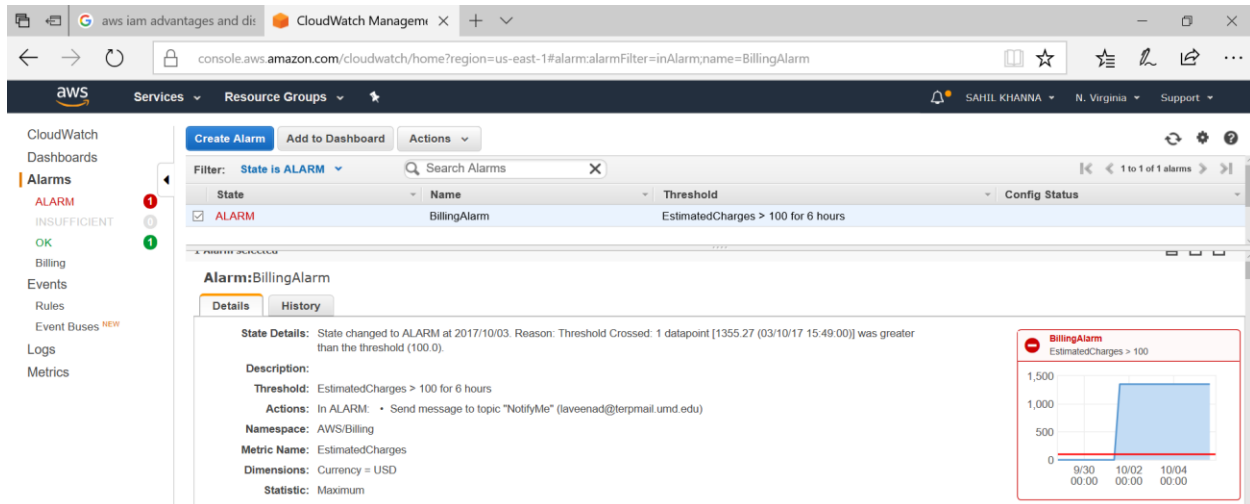the selected instance in the EC2 console.

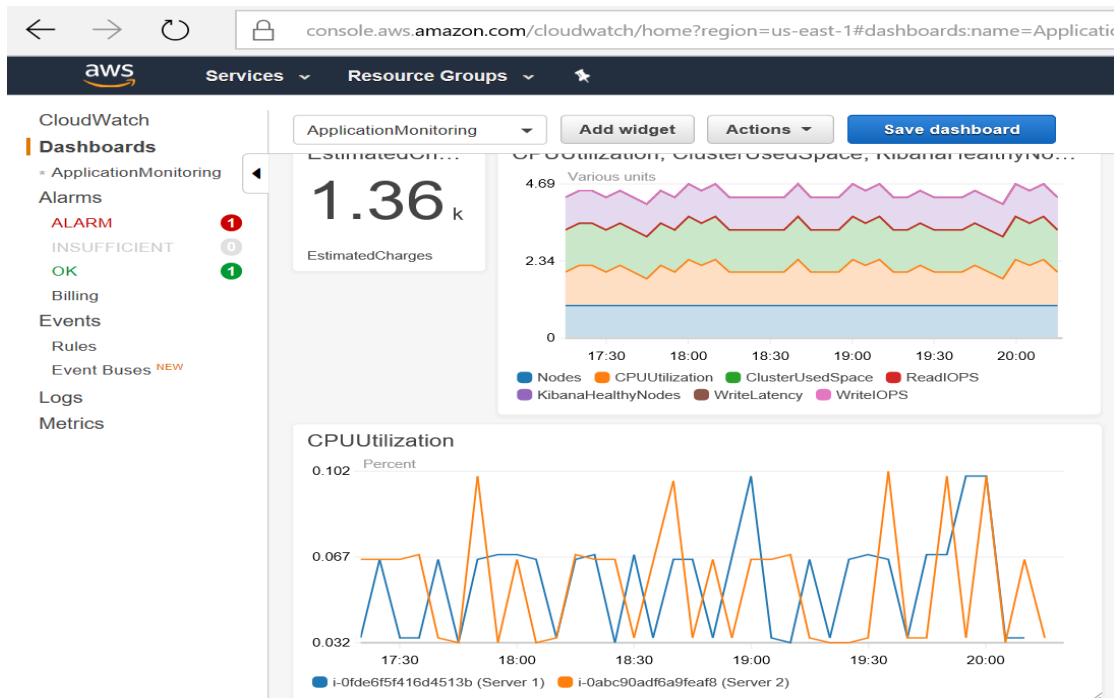Figure 8 Displaying the billing alarm notification of cloud watch

Figure 9 Dashboard created in cloud watch to monitor bill, elastic load balancer and web servers CPU utilizations

Advantages of using Cloud Watch

- It easily integrates with other services and provides with a set of utilities an administrator can use to monitor the system.
- It can set alarms for notifying the administrator when any of the instance CPU utilization shoots up significantly. It can set billing alerts to keep expenses in check, especially if with budget constraints.
- We can also set up dashboards to monitor a given set of services in a single widget.

Amazon Auto Scaling

Amazon Auto Scaling is a service through which one can always ensure that a correct number of instances are running to handle the load on the system. Auto Scaling can be a perfect fit to adhere to the needs of an environment with varying system needs (example: a web server). Auto Scaling follows a normal set up asking the user to fill in the maximum and minimum system needs, example: an EC2 instance, and it will ensure that at any given time, the system has those many instances available. In case an instance is terminated due to some error, or is crashed, Auto Scaling automatically will create an instance to ensure that the minimum system requirements are fulfilled.

Challenges/Learnings

- It requires to create an exact image of the web servers and needs to be updated every time you make some change is the server.

Test Drives

Test drives are pre-configured environments which offers tutorials to the consumers to run company's product for hands-on experience. For designers who are looking forward to using such tools, using cloud based services might be easier to follow these labs in comparison to having on site software.

High Performance Computing

AWS provides hands on lab sessions to test systems hosted on AWS. For high performance computing we have selected the Mentor Graphics' tool Mentor Graphics HyperLynx PI which lets the user test for power integrity of a PCB. The test drives come with manuals and pre-installed pcb files for users to experiment.

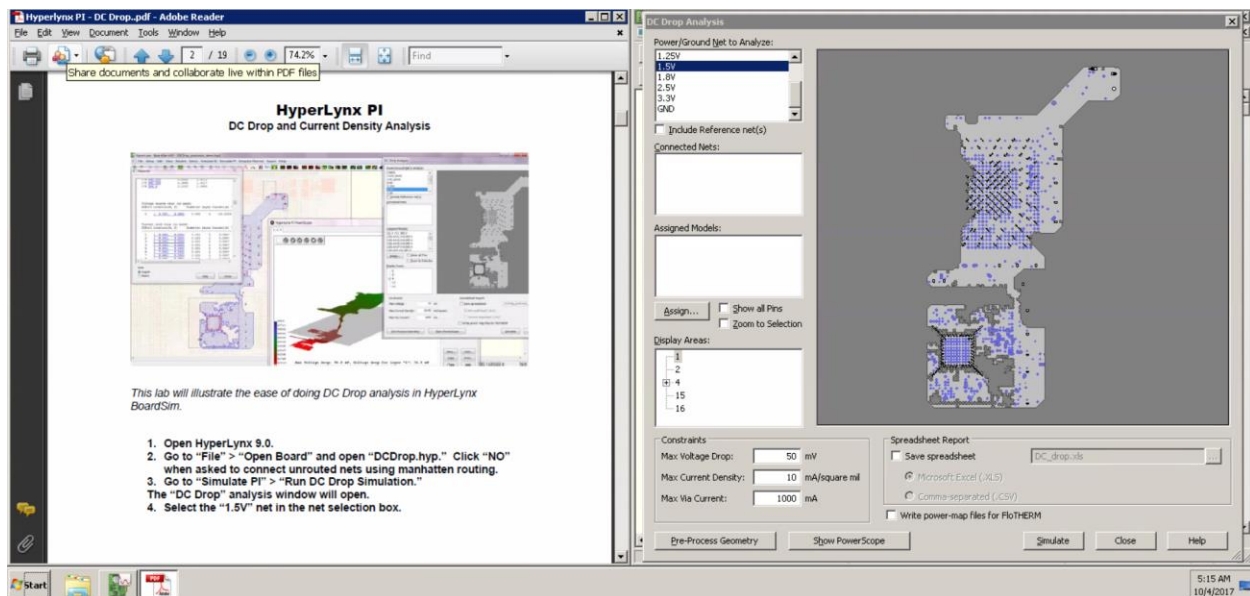*Figure 10 Lab initial interface*

Figure 11 Lab demo and results

SECURITY SPHERE

This test drive, creates an environment on AWS cloud infrastructure services to test SecuritySphere Web Application Firewall (WAF) features. It demonstrates how SecureSphere protects against advanced cyber threats such as SQL Injection and Zero-Day Attacks. In this test drive, we remotely connect to Windows server to run web application and perform cyber-attacks. The demo compares the performance of the firewall by changing and monitoring configuration parameters in the software.
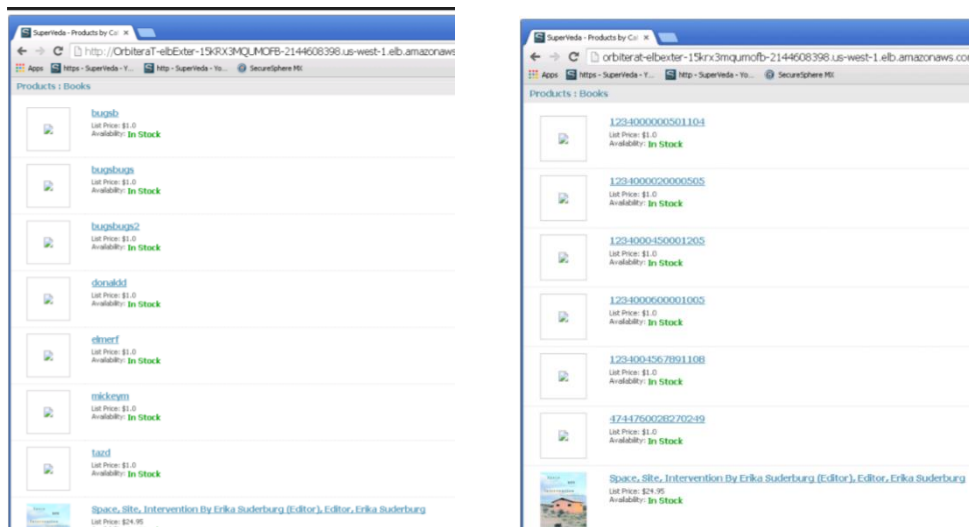

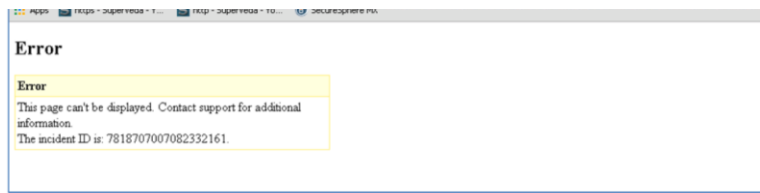*Figure 12 Web server responses after SQL injection*

Error

Error

This page can't be displayed. Contact support for additional
information.
The incident ID is: 7818707007082332161.

Figure 13 Webpage response after activating the firewall

MARKET PLACE TOOLS

Security: Kali Linux

Kali Linux is a Debian-derived Linux distribution designed for digital forensics and penetration testing. It is maintained and funded by Offensive Security Ltd. Mati Aharoni, Devon Kearns and Raphaël Hertzog are the core developers.

```
ec2-user@kali:~$ sudo nmap 10.0.0.46

Starting Nmap 7.60 ( https://nmap.org ) at 2017-10-03 21:13 UTC
Nmap scan report for ip-10-0-0-46.ec2.internal (10.0.0.46)
Host is up (0.010s latency).
Not shown: 999 closed ports
PORT    STATE SERVICE
22/tcp open  ssh
MAC Address: 0A:D1:08:C7:7C:C4 (Unknown)

Nmap done: 1 IP address (1 host up) scanned in 1.96 seconds
ec2-user@kali:~$
```

Advantages of using Kali
- It is an open source project.
- It contains many penetration tests and vulnerability assessment tools, which can used interactively using graphical user interface.

Networking: OpenVPN

OpenVPN is an open-source software application that implements virtual private network (VPN) techniques for creating secure point-to-point or site-to-site connections in routed or bridged configurations and remote access facilities. It uses a custom security protocol that utilizes SSL/TLS for key exchange. It is capable of traversing network address translators (NATs) and firewalls.

Figure 14 OpenVPN interface to see the server status

Figure 15 VPN configuration interface

Figure 16 Interface to setup security rules

Advantages of using OpenVPN

- It offers security features such as peer authentication using pre-shared keys, certificates and other usual forms of authentication, strong encryption standards using the OpenSSL Library, and HMAC packet authentication, OpenVPN is ideal for people who want to keep their networks safe and secure from prying eyes and hackers.
- When OpenVPN goes down, the network is brought to a pause to allow for repair or reconfiguration, thereby ensuring that no data loss or corruption or miscommunication happens.
- Being an open source solution, OpenVPN enjoys the support and community focus that comes with almost every software licensed under the GNU GPL.

Challenges/Learnings

- Setting up an OpenVPN network is a daunting task because it requires depth knowledge of network protocols and security concepts for configuration. It takes time to learn even the basics of OpenVPN. If it is not configured properly, it can be disastrous.

GOOGLE MAP API

Google Map Api provide various functions to run the application on google maps in real time. Map api instance runs on the project, created in Google Compute Engine. We used javascript script to integrate the Google Map api with our server applications. This api provides an interactive map which has features like zoom-in and zoom-out, and scrolling. These features allow consumers to see the result in detail and help them to design efficient models for their problems.

TWITTER API

Twitter Api produces a stream of tweets which are in json format. The twitter stream is integrated with the amazon simple queue service using python libraries. After the data is queued in SQS, we perform

sentiment analysis on this stream data using natural language processing libraries of python. The result of the sentiment analysis is further stored in the elastic search. The stored data is in json format which is manipulated using python libraries in the web servers.

CONCLUSION

The design and implementation of Tweet Mapper application covers the objective, to work on different Amazon cloud services to create an infrastructure for an industry application. With this application, we extended our research and built a tool which can be utilized to solve many real-life problems such as disaster management, election polling, health awareness and target marketing. The development of the project did not require any server costs and amazon usage charges. We can conclude that the cloud is the best option for the industry applications which solves major problems such as server maintenance and installation cost, seasonal load, human labor cost and low efficiency of resource utilization by providing on-demand usage functionality.

REFERENCES

1. Amazon Documentation, Available: https://aws.amazon.com/documentation/
2. S3 and its key features, Available: https://aws.amazon.com/s3/details/
3. EBS and key features, Available: https://aws.amazon.com/ebs/details/
4. Elastic Load Balancing: Classic ELB, https://aws.amazon.com/elasticloadbalancing/details/