Tribhuvan University

Institute of Engineering, Pulchowk Campus

Lalitpur, Nepal

15th March, 2022

A Lab Report on

Elastic Search

Report No. 4

Submitted By:

Name: Madhav Aryal

Roll No. : PUL074BCT520

Group: A

Submitted To:

Department of Electronics &

Computer Engineering

Table Of Contents

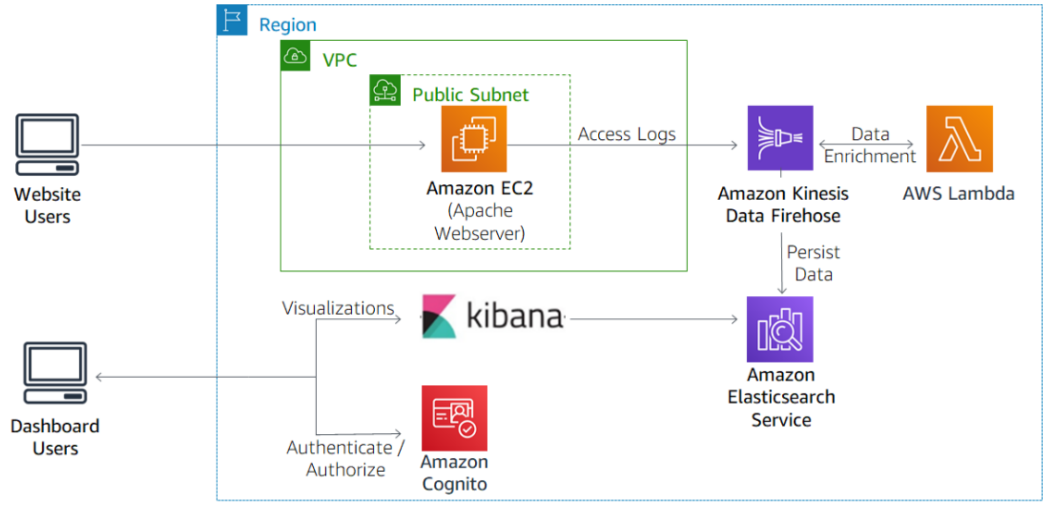
Introduction

Big data problems often require solutions in real time. This is the velocity part of the five Vs of big data (Volume, Variety, Velocity, Veracity, and Value). Some of the more common data sources for these scenarios include video streams, application logs, and infrastructure devices. Data in these velocity scenarios is called streaming data. Amazon Kinesis is a suite of services that we can use to analyze streaming data. Here, We have used Amazon Kinesis Data Firehose in this lab.

Objectives

- Access Amazon Kinesis Data Firehose and Amazon Elasticsearch Service (Amazon ES) in the AWS Management Console
- Create a Kinesis Data Firehose delivery stream
- Integrate a Kinesis Data Firehose delivery stream with Amazon ES
- Build visualizations with Kibana

Architecture



Task 1: Review the infrastructure

Streaming data is data that is generated continuously by thousands of data sources, which typically send the data records simultaneously. Generally, the data arrives as small data items. The data is often unstructured. Example sources of streaming data include:

- Web or mobile applications
- Medical devices or Internet of Things (IOT) devices
- Networking devices

The infrastructure we set up to analyze streaming data consists of the following five components:

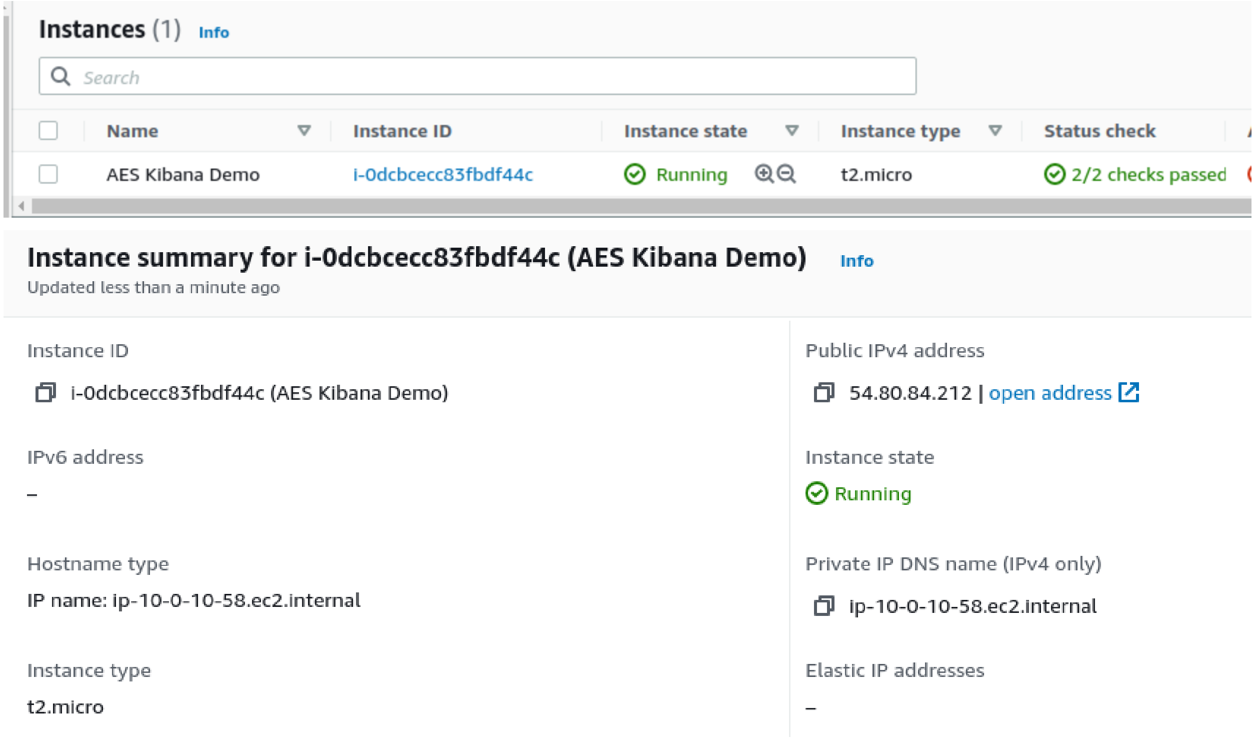
- An Amazon Elastic Compute Cloud (Amazon EC2) instance with a public subnet. The EC2 instance runs a web server.
- A Kinesis Data Firehose delivery stream that captures streaming data from the web server logs.
- An AWS Lambda function to transform the data.
- An Amazon ES cluster to store the data.

- A Kibana instance for building data visualizations.

Task 1.1: Review the Amazon EC2 instance

Amazon EC2 is a web service that provides elastic compute capacity for building and hosting applications and resources. The web server that runs in the EC2 instance includes a simple website that is composed of the following six pages:

- main.php – The home page for the website
- search.php – A page where a user can search for a product
- recommendation.php – A page that recommends a particular product based on the user search
- echo.php, kindle.php, firetvstick.php – Pages for the three products that are used in the PoC environment



□	Name	Instance ID	Instance state	Instance type	Status check
□	AES Kibana Demo	i-0dcbcecc83fbdf44c	✓ Running	t2.micro	✓ 2/2 checks passed

Instance summary for i-0dcbcecc83fbdf44c (AES Kibana Demo) Info

Updated less than a minute ago

Instance ID
i-0dcbcecc83fbdf44c (AES Kibana Demo)

Public IPv4 address
54.80.84.212 | open address

IPv6 address
–

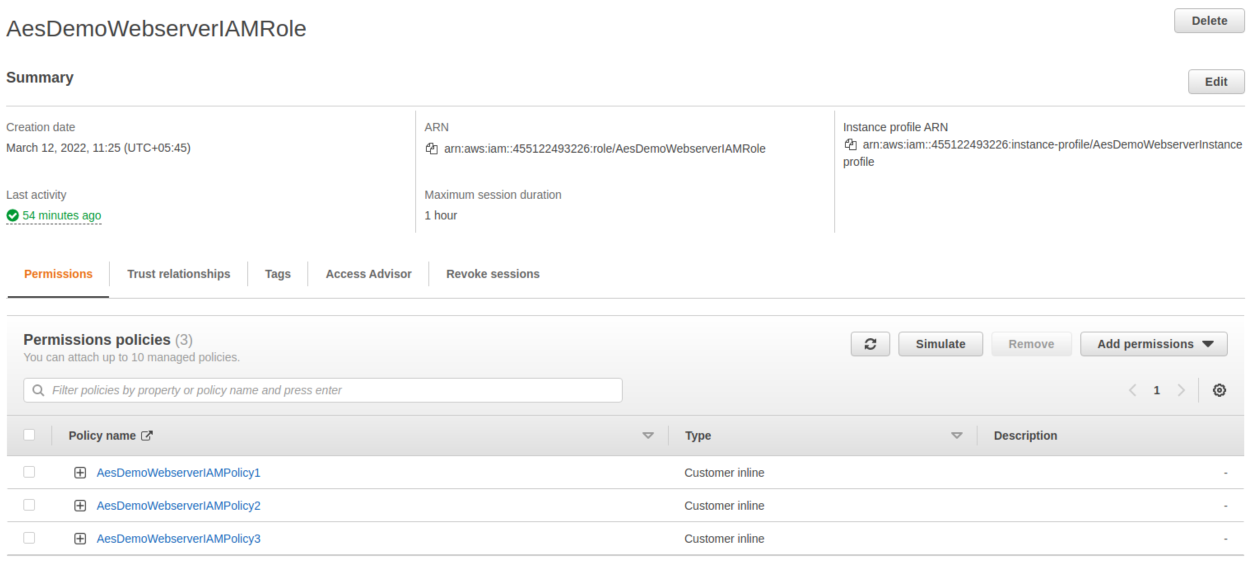
Instance state
✓ Running

Hostname type
IP name: ip-10-0-10-58.ec2.internal

Private IP DNS name (IPv4 only)
ip-10-0-10-58.ec2.internal

Instance type
t2.micro

Elastic IP addresses
–



AesDemoWebserverIAMRole

Summary

Creation date
March 12, 2022, 11:25 (UTC+05:45)

Last activity
54 minutes ago

ARN
arn:aws:iam::455122493226:role/AesDemoWebserverIAMRole

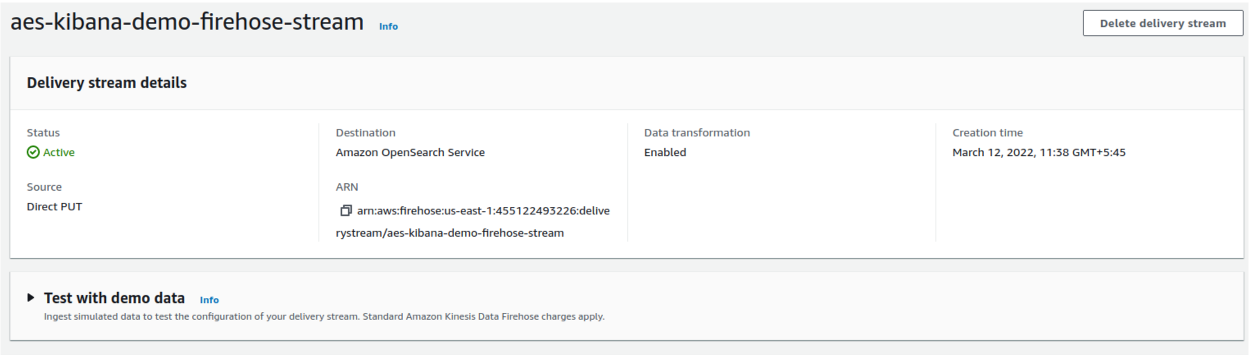
Maximum session duration
1 hour

Instance profile ARN
arn:aws:iam::455122493226:instance-profile/AesDemoWebserverInstance profile

Permissions policies (3)
You can attach up to 10 managed policies.

Policy name	Type	Description
AesDemoWebserverIAMPolicy1	Customer inline	-
AesDemoWebserverIAMPolicy2	Customer inline	-
AesDemoWebserverIAMPolicy3	Customer inline	-

Task 1.2: Review the Kinesis Data Firehose delivery stream



aes-kibana-demo-firehose-stream

Delivery stream details

Status
Active

Source
Direct PUT

Destination
Amazon OpenSearch Service

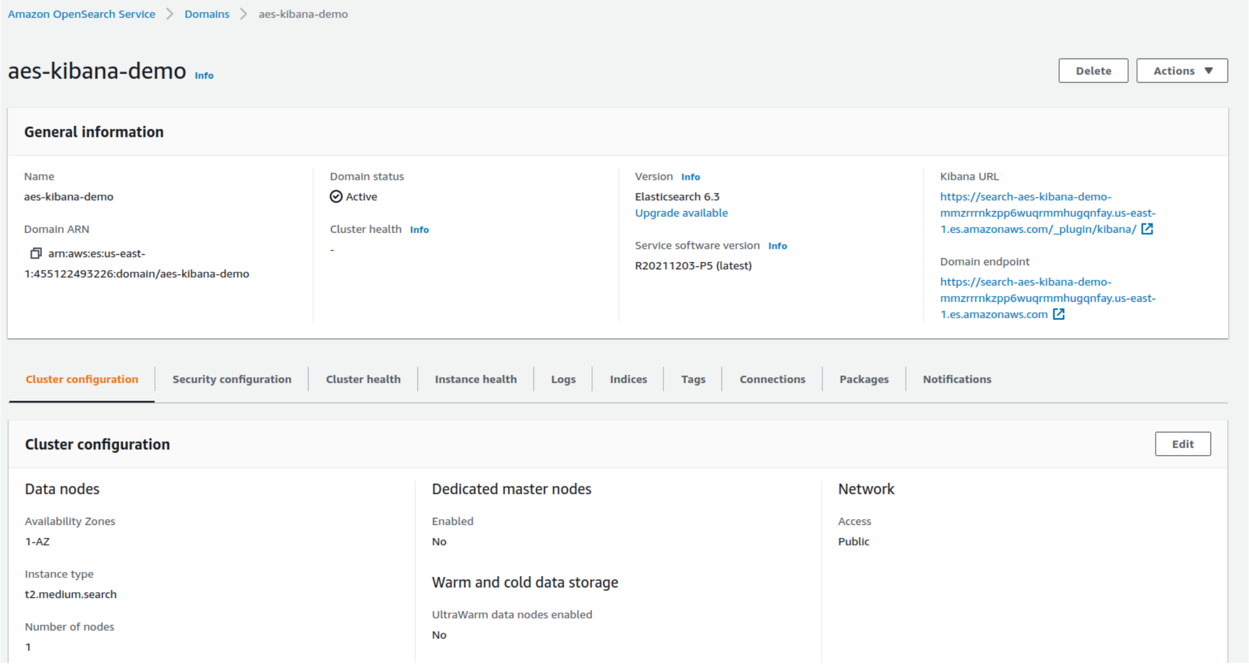
ARN
arn:aws:firehose:us-east-1:455122493226:deliverystream/aes-kibana-demo-firehose-stream

Data transformation
Enabled

Creation time
March 12, 2022, 11:38 GMT+5:45

Test with demo data
Ingest simulated data to test the configuration of your delivery stream. Standard Amazon Kinesis Data Firehose charges apply.

Task 1.3: Review the Amazon ES cluster



aes-kibana-demo Info

General information

Name
aes-kibana-demo

Domain ARN
arn:aws:es:us-east-1:455122493226:domain/aes-kibana-demo

Domain status
Active

Cluster health Info
-

Version Info
Elasticsearch 6.3
Upgrade available

Service software version Info
R20211203-P5 (latest)

Kibana URL
https://search-aes-kibana-demo-mmzrrrnkzpp6wuqrmmhugqnfay.us-east-1.es.amazonaws.com/_plugin/kibana/

Domain endpoint
https://search-aes-kibana-demo-mmzrrrnkzpp6wuqrmmhugqnfay.us-east-1.es.amazonaws.com

Cluster configuration

Data nodes

Availability Zones
1-AZ

Instance type
t2.medium.search

Number of nodes
1

Dedicated master nodes

Enabled
No

Warm and cold data storage

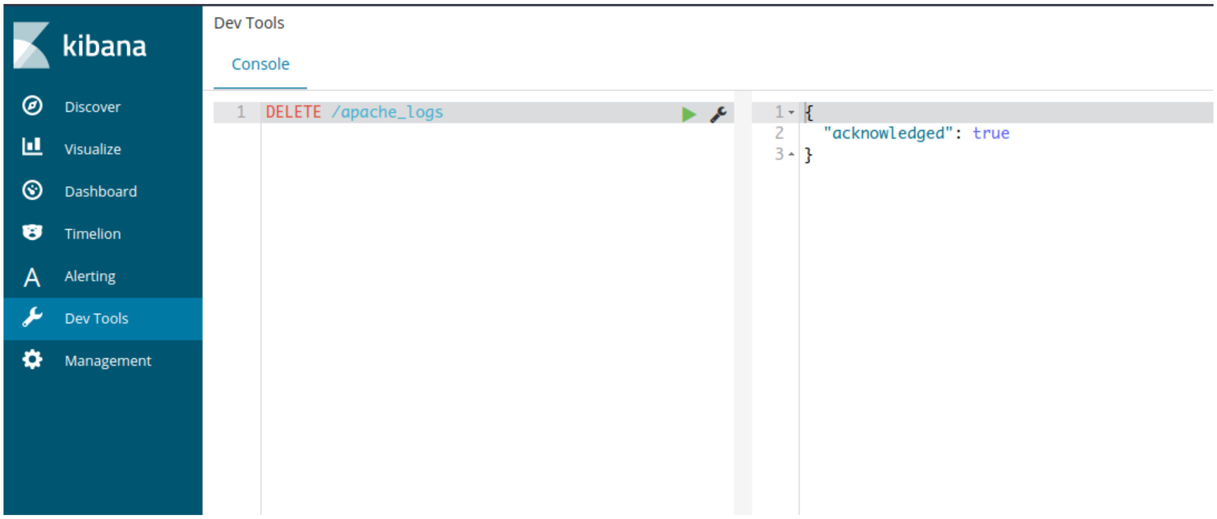
UltraWarm data nodes enabled
No

Network

Access
Public

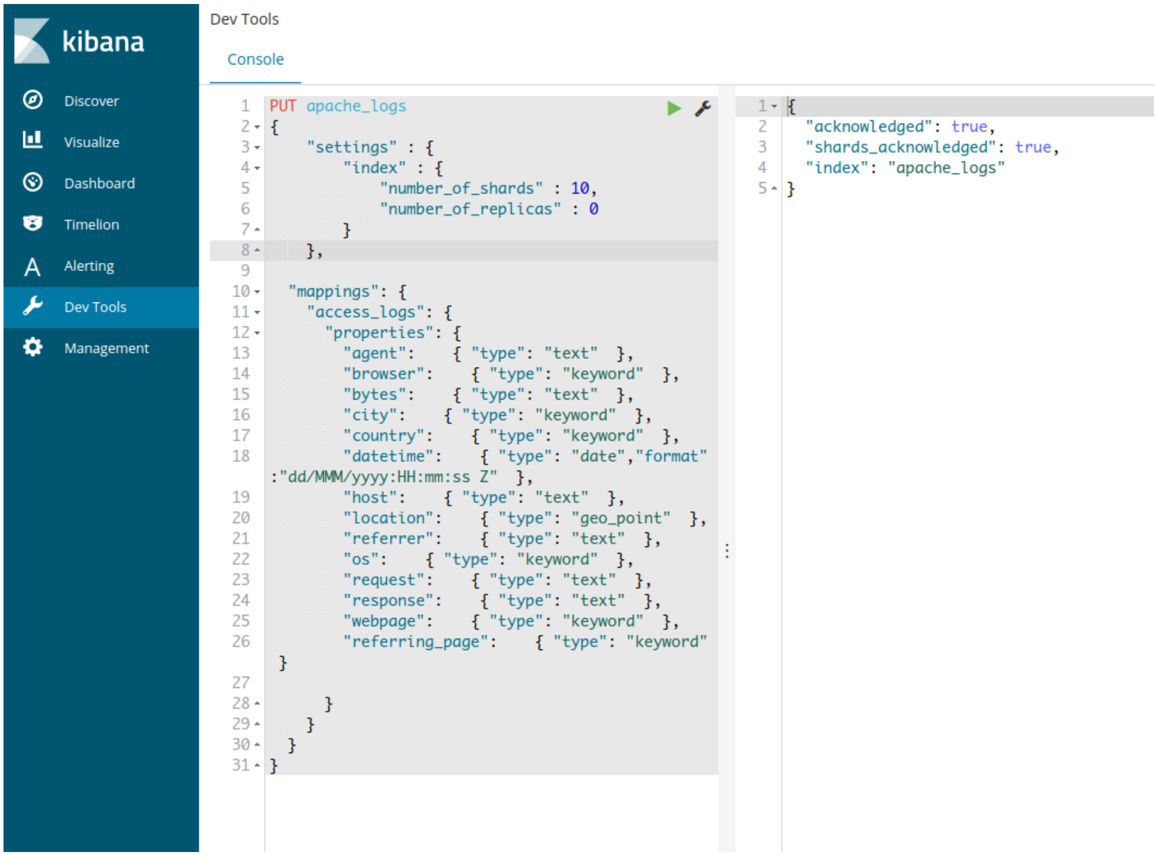
Task 2: Configure Kibana

Kibana is an open source data visualization tool for analyzing data in an Amazon ES cluster.



```
DELETE /apache_logs
```

```
{
  "acknowledged": true
}
```

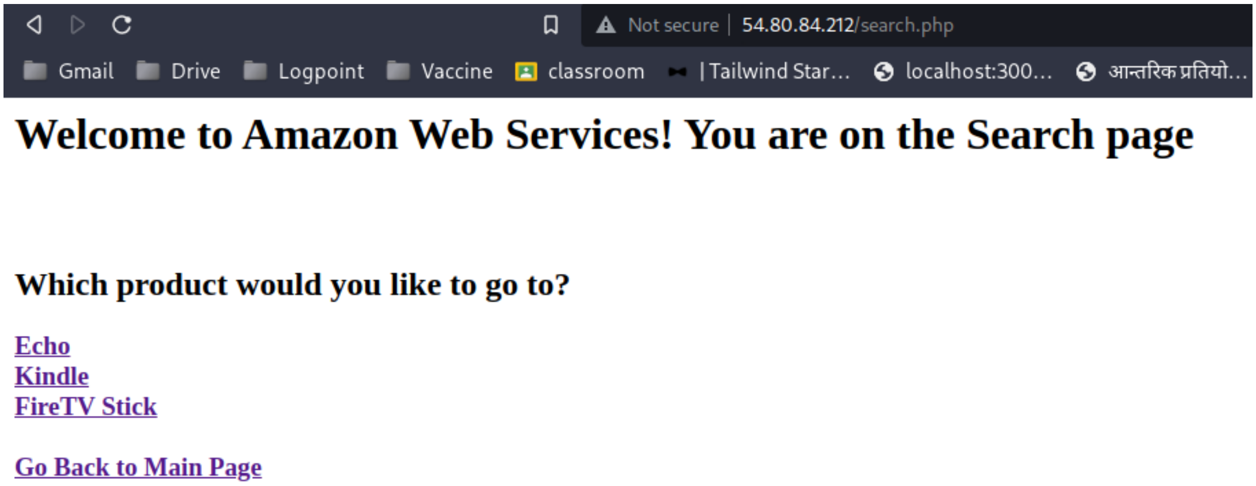


```
PUT apache_logs
{
    "settings" : {
        "index" : {
            "number_of_shards" : 10,
            "number_of_replicas" : 0
        }
    },

  "mappings": {
    "access_logs": {
      "properties": {
        "agent":    { "type": "text"  },
        "browser":    { "type": "keyword"  },
        "bytes":    { "type": "text"  },
        "city":    { "type": "keyword"  },
        "country":    { "type": "keyword"  },
        "datetime":    { "type": "date","format":"dd/MMM/yyyy:HH:mm:ss Z"  },
        "host":    { "type": "text"  },
        "location":    { "type": "geo_point"  },
        "referrer":    { "type": "text"  },
        "os":    { "type": "keyword"  },
        "request":    { "type": "text"  },
        "response":    { "type": "text"  },
        "webpage":    { "type": "keyword"  },
        "referring_page":    { "type": "keyword" }

      }
    }
  }
}
```

```
{
  "acknowledged": true,
  "shards_acknowledged": true,
  "index": "apache_logs"
}
```

Task 3: Populate the web server log with data

Several pages were opened by navigating through the website. Repeated the process with another web browser.



Welcome to Amazon Web Services! You are on the Search page

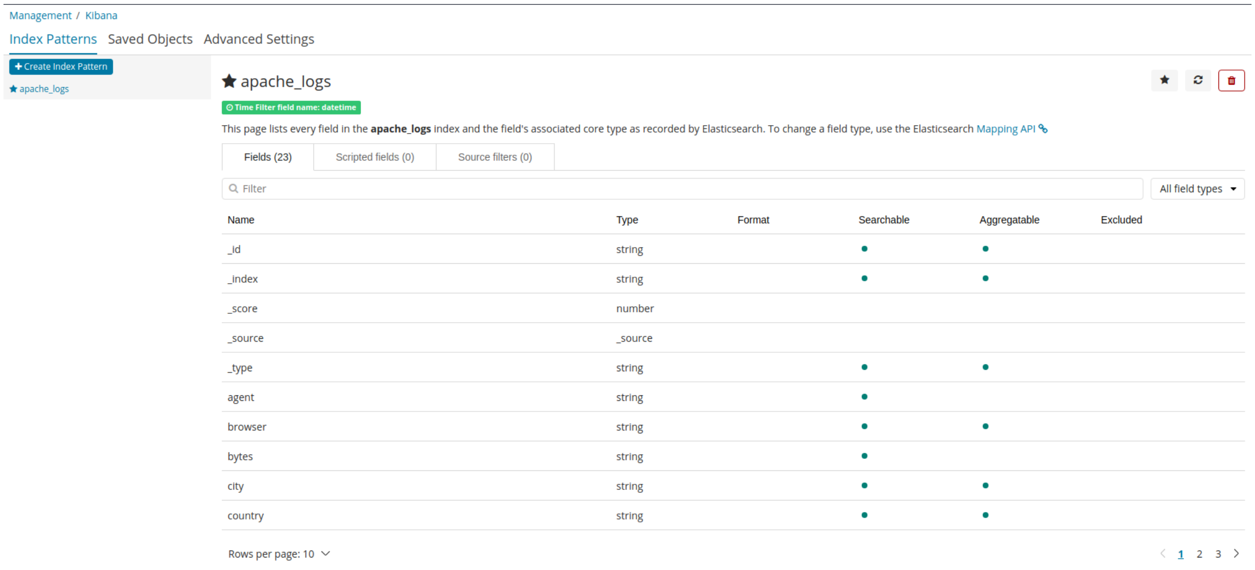
Which product would you like to go to?

Echo
Kindle
FireTV Stick

Go Back to Main Page

Task 4: Create the Kibana index

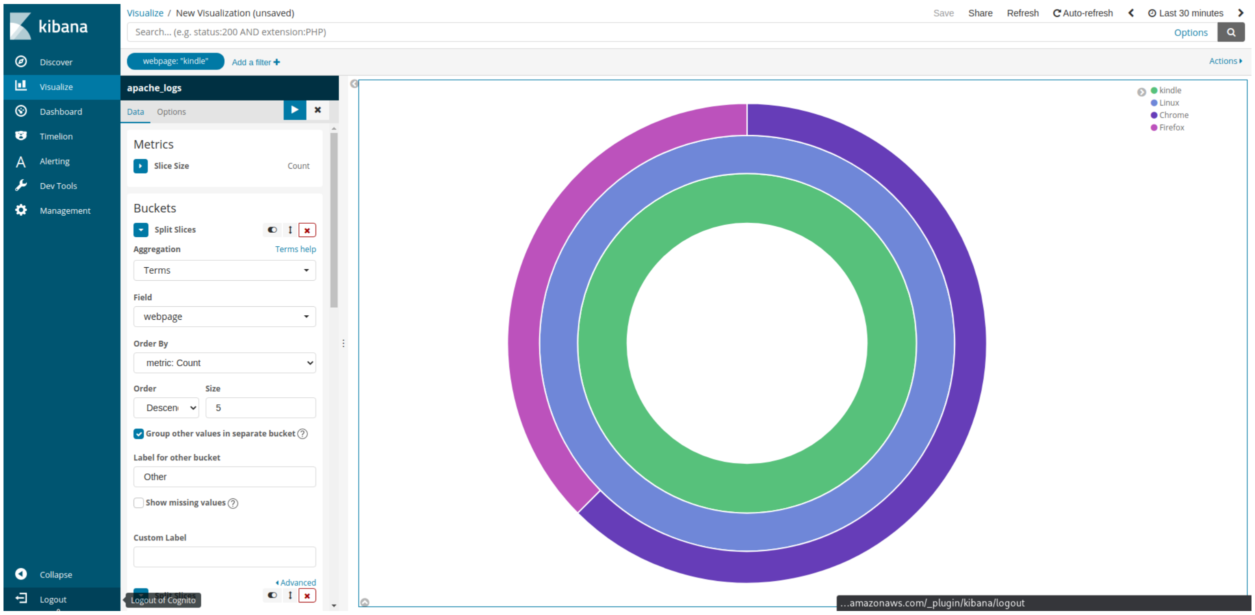
Kibana uses index patterns to identify the Elasticsearch indices that we want to use for creating our visualizations.



This page lists every field in the **apache_logs** index and the field's associated core type as recorded by Elasticsearch. To change a field type, use the Elasticsearch Mapping API

Name	Type	Format	Searchable	Aggregatable	Excluded
_id	string		●	●	
_index	string		●	●	
_score	number				
_source	_source				
_type	string		●	●	
agent	string		●		
browser	string		●	●	
bytes	string		●		
city	string		●	●	
country	string		●	●	

Task 5: Create the pie chart for PoC



webpage: "kindle"

Metrics

Slice Size: Count

Buckets

Split Slices

Aggregation: Terms

Field: webpage

Order By: metric: Count

Order: Descen

Size: 5

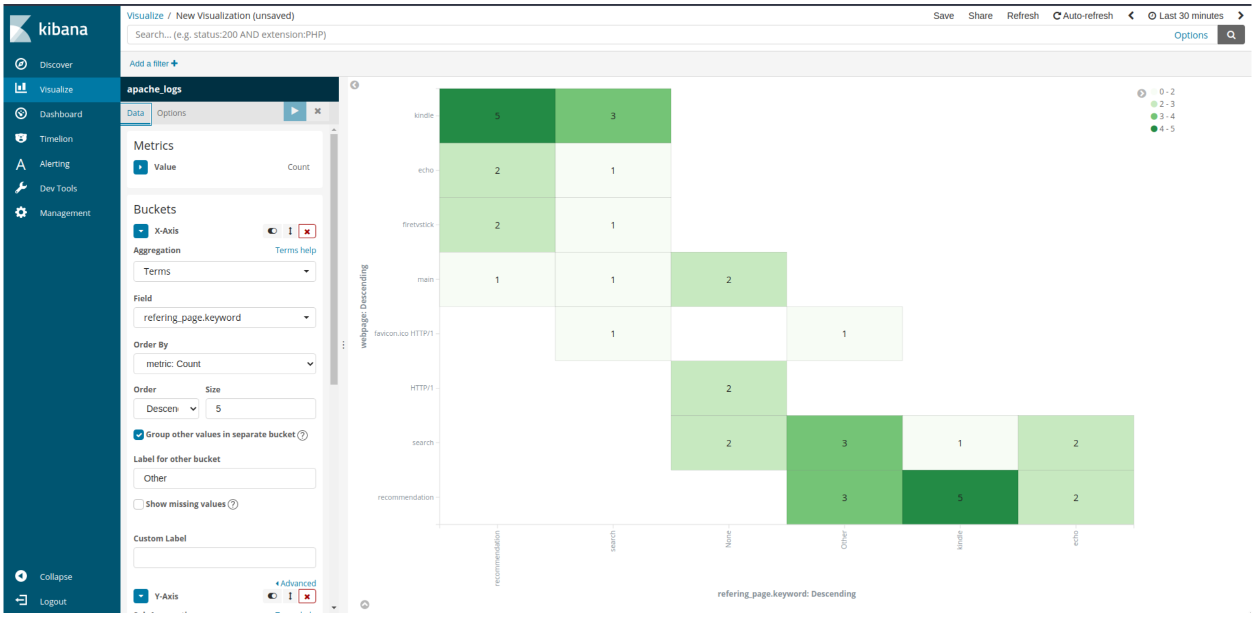
☑ Group other values in separate bucket

Label for other bucket: Other

☐ Show missing values

Custom Label:

Task 6: Create a heat map for PoC



Visualize / New Visualization (unsaved)

Search... (e.g. status:200 AND extension:PHP)

apache_logs

Metrics

Value: Count

Buckets

X-Axis

Aggregation: Terms

Field: refering_page.keyword

Order By: metric: Count

Order: Descen Size: 5

☑ Group other values in separate bucket

Label for other bucket: Other

□ Show missing values

Custom Label:

Y-Axis

	recommendation	search	None	Other	kindle	echo
kindle	5	3				
echo	2	1				
firetvstick	2	1				
main	1	1	2			
favicon.ico HTTP/1		1		1		
HTTP/1			2			
search			2	3	1	2
recommendation				3	5	2

Discussion and Conclusion

In this lab, we learnt about basic concepts of Elastic Search and its implementation in Amazon Elasticsearch Service(Amazon ES), Amazon Kinesis and Kibana. From this we understood that we can construct a pipeline to monitor real-time activity using Amazon Kinesis (for data collection), Amazon ES (for data indexing and searching), and Kibana (for real time visualization).