Offsite Test for DevOps

1. This is a technical test to understand our candidate better. Please try at your best efforts.
2. Please prepare document for the code / test case and detail setup instruction as a Git repository. Please use a random/meaningless project name for the git repository (e.g. 5848cf8a7dd2/wahaha)
3. The code needs to work after we clone it.
4. This test may contain bugs. Any bug fixing is appreciated.
5. Happy hacking!

Q1. Access Log analytics

You are given with an access log file (access.log.gz) containing some HTTP requests.

Part of the access log:

```
66.249.75.138 - - [19/May/2019:06:25:48 +0800] "GET
/doc/jumbo-tender-20170928.pdf HTTP/1.1" 200 2875782 "-"
"Mozilla/5.0 (Linux; Android 6.0.1; Nexus 5X Build/MMB29P)
AppleWebKit/537.36 (KHTML, like Gecko) Chrome/41.0.2272.96
Mobile Safari/537.36 (compatible; Googlebot/2.1;
+http://www.google.com/bot.html)" "66.249.75.138"
216.244.66.196 - - [19/May/2019:06:26:58 +0800] "GET
/debian-vim-syntax/ HTTP/1.1" 200 4985 "-" "Mozilla/5.0
(compatible; DotBot/1.1;
http://www.opensiteexplorer.org/dotbot, help@moz.com)"
"216.244.66.196"
216.244.66.196 - - [19/May/2019:06:27:00 +0800] "GET
/nian-nian-you-yu-yan/ HTTP/1.1" 200 6288 "-" "Mozilla/5.0
(compatible; DotBot/1.1;
http://www.opensiteexplorer.org/dotbot, help@moz.com)"
"216.244.66.196"
220.181.108.184 - - [19/May/2019:06:27:44 +0800] "GET
/2010/08/22/ HTTP/1.1" 404 2910 "-" "Mozilla/5.0 (compatible;
Baiduspider/2.0; +http://www.baidu.com/search/spider.html)"
"220.181.108.184"
```

Please prepare script(s) (any shell/scripting language executable on Linux in AWS environment) to:

- Count the total number of HTTP requests recorded by this access logfile
- Find the top-10 (host) hosts makes most requests from 2019-06-10 00:00:00 to 2019-06-19 23:59:59, inclusively
- Find out the country with most requests originating from (according to the source IP)

Q2. AWS API programming

Please help to prepare a script to query AWS API and look up the public IP of the instance with the specific EC2 Name tag. Then the script should execute the ssh command
ssh ec2-user@EC2_PUBLIC_IP .

Example interaction of the script (say the executable script is awssh):

```
$ awssh api-server-002
ssh ec2-user@120.45.256.57
[ec2-user@ip-120-45-256-57 ~]$
$ awssh not-existing-host
Host not found
```

You may test against your code with real AWS account OR AWS EC2 mock API moto (https://github.com/spulec/moto)

Q3. System design and Implementation

Please implement a URL shortener service (e.g. https://bitly.com/)

Implement a simple bit.ly like service, name the following 2 RESTful endpoints

1. RESTful endpoint for url submission

POST /newurl
- Request: { "url": "https://www.google.com" }
- Response: { "url": "https://www.google.com", "shortenUrl": "https://shortenurl.org/g20hi3k9"}

2. Shorten redirect URL

GET /[a-zA-Z0-9]{9} (regex, eg. g20hi3k9)
- HTTP 304 to saved link (eg. https://www.google.com according previous example)

- No update on the shorten link once created

System design concern

- High availability: Please make it highly available and no single point of failure.
- Scalability : Please make it scalable.
- Scaling target: 1000+ req/s, after scaling-up/out without major code change

Tech stack

- You could choose any technology / programming language / database / cache / AWS services, as long as the application code is executable on generic Linux EC2 node and the whole setup fits in AWS environment.

Deliverable & Documentation

- The application deliverable should be self-contained, preferably an automated deployment package / container image, such that we can deploy it easily

- The system/infrastructure should be also documented OR automated (via e.g. shell script)
- Both the system design and the actual system infrastructure configuration are the key measurement of this question.
- Please briefly explain your system and say why you are going to implement like that.
- Please state any assumption and limitation of the system implemented.