1. List some logging and visualization tools available in the market with the preferred scenario to use one over other.

Some of the tools for logging and visualization are as follows

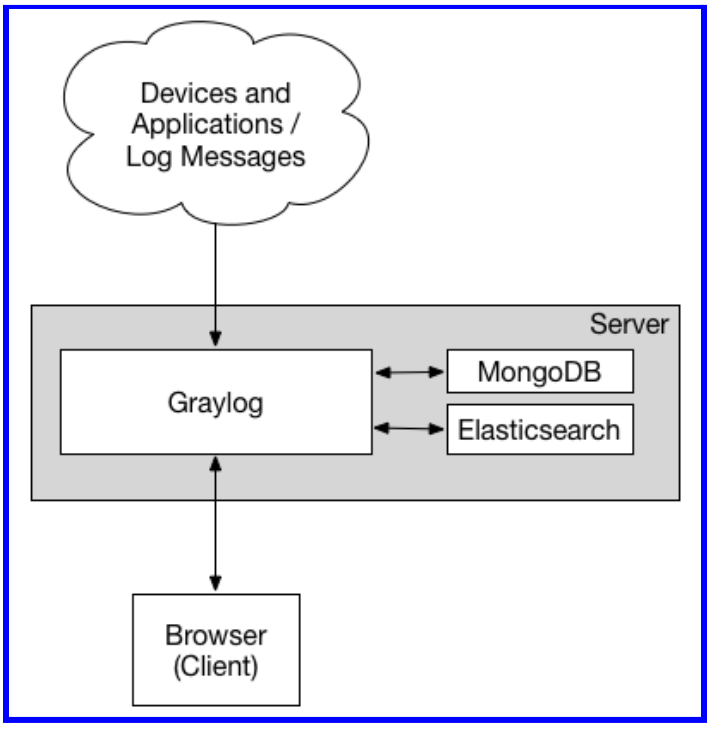
- **Graylog**
- **ELK (**Elasticsearch, Logstash and Kibana**)**
- **Splunk**
- **Datadog**

The feature of each log management tool for which it has been choose in the market is as follows

Graylog

The Graylog software centrally captures, stores, and enables real-time search and log analysis against terabytes of machine data from any component in the IT infrastructure and applications.

The software uses a three-tier architecture and scalable storage based on Elasticsearch and MongoDB.



This is a minimum to be used for smaller, non-critical, or test setups. None of the components are redundant, and they are easy and quick to set up.

Features of Graylog

- Ksystemlog can ingest any structured data, including log messages and network traffic.
- Provides a fully customizable dashboard with numbers of a widget.
- Use standard Boolean search terms for selecting fields and data types.
- Send real-time alert notifications to admin in various ways like email, text, and Slack.
- Graylog usually contains sensitive and regulated data so that the system itself remains accessible, secure, and speedy.
- Has predefined templates to display data.

Benefits over other tools

- **Real-time Answers and fast.**
- **Empower non-tech users**
- **Lower operations costs**
- **Explore your data**

ELK (Elasticsearch, Logstash and Kibana)

Elastic Stack, commonly abbreviated as ELK, is a popular three-in-one log centralization, parsing, and visualization tool that centralizes large sets of data and logs from multiple servers into one server.
ELK stack comprises 3 different products:

Logstash

Logstash is a free and open-source data pipeline that collects logs and events data and even processes and transforms the data to the desired output.

Elasticsearch

Built on Apache Lucene, Elasticsearch is an open-source and distributed search and analytics engine for nearly all types of data – both structured and unstructured.

Kibana

Data is finally passed on to Kibana, which is a WebUI visualization platform that runs alongside Elasticsearch. Kibana allows you to explore and visualize time-series data and logs from elasticsearch.

Features of Elastic Stack

- Clustering and high availability
- Automatic node recovery
- Index lifecycle management
- Secure settings and many more …..

Benefits over other tools

- **It is interoperable**

It can interoperate with other tools to get the job done.

- **It is open-source**

The Elastic Stack is a collection of open-source projects. If you want to deploy it yourself to test it out, you can.

- **It is managed**

If you have the money or don't want to spend any time maintaining your ELK Stack, you can subscribe to a managed service and reduce your time to value.

- **Parting thoughts**

There are many use cases with ELK, and you can break your deployment up into small parts if needed.

Splunk is a software platform widely used for monitoring, searching, analyzing and visualizing the machine-generated data in real time.

It performs capturing, indexing, and correlating the real time data in a searchable container and produces graphs, alerts, dashboards and visualizations.

Features of Splunk

- **Dashboards and Visualizations**

  Customized dashboards and data visualizations give voice to your data.

- **Monitoring and Alerting**

  Continuous monitoring of events, conditions, and critical KPIs helps keep your operations running smoothly.

- **Reporting**

  Reports can be created in real time, scheduled to run at any interval and used in your dashboards.

- **Machine Learning Toolkit (MLTK)**

  Use pre-built Splunk machine learning analytics for identifying use cases or create your own custom machine learning models to tackle impactful issues or opportunities in your company.

Benefits over other tools

- **Accelerate Your Digitization**

  Whether you're just starting to digitize, or you were born in the cloud, innovate with confidence with purpose-built solutions driven by AI and machine learning.

  Splunk solutions provide everything you need to ensure your digital initiatives succeed.

- **Ensure Business Resilience**

  Empower your people to predict, identify and solve problems in real time.

  Answer questions across business, IT, DevOps and security functions with world-class investigative capabilities, intuitive visualizations and seamless collaboration.

- **Meet the Data Opportunities of Today and Tomorrow**

  It’s flexible platform and purpose-built solutions scale with you as your data and organization evolve

Datadog Log Management, also referred to as Datadog logs or logging, provides decoupling log ingestion from indexing.

This enables you to cost-effectively collect, process, archive, explore, and monitor all of your logs without limitations, also known as Logging without Limits*.

Features of Datadog

- **See across systems, apps, and services**

With turn-key integrations, Datadog seamlessly aggregates metrics and events across the full devops stack.

- **Analyze and explore log data in context**

Quickly search, filter, and analyze your logs for troubleshooting and open-ended exploration of your data.

- **Visualize traffic flow in cloud-native environments**

Understand performance using meaningful, human-readable tags.

- **Build real-time interactive dashboards**

More than summary dashboards, Datadog offers all high-resolution metrics and events for manipulation and graphing. And many more …...

Benefits over other tools

- **Customizable Dashboard**

The dashboard used in the Datadog can be easily customizable as per the requirements.

- **Easy installation**

The configuration and installation of the Datadog tool are easy as it uses the SaaS service.

- **Cheaper than others**

The implementation of the tool is cheap.

There are also many more log management and visualization tools like, **LOGalyze, Glogg, GoAccess, Frontail, Multitail, Logwatch, Nagios, etc** which are open-source and efficient to use.

2. Mention 10 best practises when logging. Why is log formatting necessary?

The 10 important thing to be considered during logging are as follows:

- **Log only what's needed. Don't log everything**

  i.e Don't log too much or Too little

  Logging too much can cause storage volume sortage and it will really become hard to get any value from it.

  Too little log will risk not being able to troubleshoot problems.

- **Don't log sensitive information**

  i.e we should be aware of what not to log like

  If anyone gets the sensitive data from the log, then our system could be vulnerable and can help others to breach the security system.

  Some information should not be logged like:

  - Personally Identifiable Information (PII)
  - Business Names and Contact Informations
  - Financial Data (Back Accounts, Card Details, etc)
  - Passwords, security keys, secrets, etc

- **Format logs so they are easily parsable and readable**

  Use standard date and time format. Most probably UTC.

  Add more context to the logs Logging Practises.

- **Separation of concerns for logs**

  The log level is used to denote the severity of each event in the system.

  Using different log levels like **FATAL < ERROR < WARN < INFO < DEBUG < TRACE** can help to distinguish the troubleshooting priority too.

  E.g.

  2012-05-23T15:02:27Z | ERROR | access to /home/test/index.html is forbidden. Client: 202.123.22.23
  2018-03-22T11:34:12Z | WARN | The plugin xyz is deprecated and will be removed from future releases.

  Include developer-friendly log messages (short and sweet)

- **Centralize logs**

  It is mandatory to keep these logs preserved both locally and in central storage to make it hard for intruders and cybercriminals to access both locations and delete evidence of their activity at the same time.

  It helps us prevent breaches from going unnoticed.

  Logs can generate from multiple servers

  Accessing each server to retrieve the logs is painful

- **Log rotation**

  Eg. with logrotate utility, you can configure to rotate your logs with configuration as such

```
/var/log/nginx/access.log {
  Daily
  Rotate 7
  Size 20M
  Compress
}
```

  Log rotate can help us to take logs of the applications Daily/Weekly, keeping the require amount of logs 7 logs weekly, determine the size of the log, helps to compress the log, etc

- **Using Recommended Tools**

  Using recommended tools can help to get log in more efficient way like

  - Logging Frameworks

    We can use tools like JavaScript, TypeScript, Java, Golang, etc

  - Log inspection/aggregation/monitoring tools

    We can use CLI tools, Cloud tools, SEIM tools and others like ELK stack, Loggly, etc

- **Using the English Language**

  Some tools and terminal consoles do not support printing and storing log messages with certain Unicode characters. Hence localization and other advanced features may be challenging at the logging level.

  Therefore, make sure you stick to the English language and always use a widely accepted character set for writing log messages.

- **Make each log message unique across the system**

  One mistake that most developers make is to copy-paste the same log message in multiple files while letting the final log aggregate fill with similar log lines coming from multiple different parts of the system.

  Doing that,, it is not easy to trace the exact location that triggered the event in the code.

At least mentioning the log source with the log message to make the final log lines look different from each other.

- **Set up alerts and push notifications when critical incidents occur**
  Almost all the log monitoring tools include features to define custom thresholds at certain levels.
  When the system hits those levels, the monitoring tool will proactively detect them with the help of log data and notify SysAdmins via alarms, push notifications APIs (e.g. Slack Audit Logs API), emails, etc.
  Also they can be preconfigured to trigger automated processes like dynamic scaling, system backup, changeovers, etc.

Why is log formatting necessary ?

Generally when logs are generated,
The problem with log files is they are unstructured text data which makes it difficult to generate meaningful information by just looking at the logs.
Log formatting is necessary

- Process log files for analytics or business intelligence
- Searching log files
- To generate more meaningful information.
- To set severity levels such as "**FATAL, ERROR, CRITICAL**",etc so that we can focus on specific logs to solve.
- To solve problems like filtering all logs by a certain customer or transaction and allow additional analytics

3. Create a file in your system. Whenever someone performs some action(read, write, execute) on that file, the event should be logged somewhere.

To watch the file, we should need tools like

- **iwatch**
- **auditd, etc**

Installing auditd tool

sudo apt install auditd

sudo service start auditd

To see if auditd is active

sudo systemctl status auditd.service

```
bibek@bibek-LfTech:~$ sudo systemctl status auditd.service
● auditd.service - Security Auditing Service
     Loaded: loaded (/lib/systemd/system/auditd.service; enabled; vendor preset: e
     Active: active (running) since Sat 2021-11-27 20:47:45 +0545; 34s ago
       Docs: man:auditd(8)
             https://github.com/linux-audit/audit-documentation
   Main PID: 7063 (auditd)
      Tasks: 2 (limit: 2946)
     Memory: 384.0K
     CGroup: /system.slice/auditd.service
             └─7063 /sbin/auditd

नव  स्रबर  27  20:47:45     bibek-LfTech augenrules[7077]: backlog_wait_time 15000
नव  स्रबर  27  20:47:45     bibek-LfTech augenrules[7077]: enabled 1
नव  स्रबर  27  20:47:45     bibek-LfTech augenrules[7077]: failure 1
नव  स्रबर  27  20:47:45     bibek-LfTech augenrules[7077]: pid 7063
नव  स्रबर  27  20:47:45     bibek-LfTech augenrules[7077]: rate_limit 0
नव  स्रबर  27  20:47:45     bibek-LfTech augenrules[7077]: backlog_limit 8192
नव  स्रबर  27  20:47:45     bibek-LfTech augenrules[7077]: lost 0
नव  स्रबर  27  20:47:45     bibek-LfTech augenrules[7077]: backlog 4
नव  स्रबर  27  20:47:45     bibek-LfTech augenrules[7077]: backlog_wait_time 0
नव  स्रबर  27  20:47:45     bibek-LfTech systemd[1]: Started Security Auditing Service.
```

To show the rules

sudo auditctl -l

```
bibek@bibek-LfTech:~$ sudo auditctl -l
No rules
```

At first no rules are written

Creating a bash script file

vi echo.sh

```
#!/bin/bash
echo "Log of leapfrog"
```

Creating echo.log to store logs

touch echo.log

To write rule using auditd

sudo auditctl -w /home/bibek/assignment/auditd/echo.sh -p rwx -k

/home/bibek/assignment/auditd/echo.log

Here during writing log,

- **-w represent path to the file**
- **-p represent permission such as r(read), w(write), x(execute),a(change in the file's attribute)**
- **-k represent the file for storing logs**

Now doing cat action on echo.sh file

cat echo.sh

```
bibek@bibek-LfTech:~/assignment/auditd$ cat echo.sh
#!/bin/bash
echo "Log of leapfrog"
bibek@bibek-LfTech:~/assignment/auditd$
```

Checking the log file

sudo ausearch -k echo.log

```
time->Sat Nov 27 21:13:05 2021
type=PROCTITLE msg=audit(1638026885.704:172): proctitle=636174006563686F2E7368
type=PATH msg=audit(1638026885.704:172): item=0 name="echo.sh" inode=532249 dev=08:
05 mode=0100775 ouid=1000 ogid=1000 rdev=00:00 nametype=NORMAL cap_fp=0 cap_fi=0 ca
p_fe=0 cap_fver=0 cap_frootid=0
type=CWD msg=audit(1638026885.704:172): cwd="/home/bibek/assignment/auditd"
type=SYSCALL msg=audit(1638026885.704:172): arch=c000003e syscall=257 success=yes e
xit=3 a0=ffffff9c a1=7fff41fa66ac a2=0 a3=0 items=1 ppid=5399 pid=8008 auid=1000 ui
d=1000 gid=1000 euid=1000 suid=1000 fsuid=1000 egid=1000 sgid=1000 fsgid=1000 tty=p
ts1 ses=9 comm="cat" exe="/usr/bin/cat" subj=unconfined key="/home/bibek/assignment
/auditd/echo.log"
bibek@bibek-LfTech:~/assignment/auditd$
```

Now running script echo.sh

./echo.sh

```
bibek@bibek-LfTech:~/assignment/auditd$ ./echo.sh
Log of leapfrog
bibek@bibek-LfTech:~/assignment/auditd$
```

Again viewing the log

sudo ausearch -k echo.log

```
ts1 ses=9 comm="cat" exe="/usr/bin/cat" subj=unconfined key="/home/bibek/assignment
/auditd/echo.log"
----
time->Sat Nov 27 21:16:00 2021
type=PROCTITLE msg=audit(1638027060.803:179): proctitle=2F62696E2F62617368002E2F656
3686F2E7368
type=PATH msg=audit(1638027060.803:179): item=2 name="/lib64/ld-linux-x86-64.so.2"
inode=661064 dev=08:05 mode=0100755 ouid=0 ogid=0 rdev=00:00 nametype=NORMAL cap_fp
=0 cap_fi=0 cap_fe=0 cap_fver=0 cap_frootid=0
type=PATH msg=audit(1638027060.803:179): item=1 name="/bin/bash" inode=655452 dev=0
8:05 mode=0100755 ouid=0 ogid=0 rdev=00:00 nametype=NORMAL cap_fp=0 cap_fi=0 cap_fe
=0 cap_fver=0 cap_frootid=0
type=PATH msg=audit(1638027060.803:179): item=0 name="./echo.sh" inode=532249 dev=0
8:05 mode=0100775 ouid=1000 ogid=1000 rdev=00:00 nametype=NORMAL cap_fp=0 cap_fi=0
cap_fe=0 cap_fver=0 cap_frootid=0
type=CWD msg=audit(1638027060.803:179): cwd="/home/bibek/assignment/auditd"
type=EXECVE msg=audit(1638027060.803:179): argc=2 a0="/bin/bash" a1="./echo.sh"
type=SYSCALL msg=audit(1638027060.803:179): arch=c000003e syscall=59 success=yes ex
it=0 a0=5605df6fe720 a1=5605df70e730 a2=5605df6fb7b0 a3=8 items=3 ppid=5399 pid=803
0 auid=1000 uid=1000 gid=1000 euid=1000 suid=1000 fsuid=1000 egid=1000 sgid=1000 fs
gid=1000 tty=pts1 ses=9 comm="echo.sh" exe="/usr/bin/bash" subj=unconfined key="/ho
me/bibek/assignment/auditd/echo.log"
```

With iwatch

Installing iwatch

sudo apt install iwatch

Creating leapfrog.sh file and for storing log creating leapfrog.log

vi leapfrog.sh

```
#!/bin/bash
echo "IWATCH logs
```

touch leapfrog.log

Watching log of leapfrog file

iwatch -e access,close_nowrite,close_write,open leapfrog.sh

- **-e represents events access, close_nowrite(opened in read-only mode), close_write(opened in writable mode),open,etc**

```
bibek@bibek-LfTech:~/assignment/iwatch$ iwatch -e access,close_nowrite,close_write,
open leapfrog.sh
Wide character in print at /usr/bin/iwatch line 359.
[27/  नव ह्बर /2021    21:32:11]      IN_OPEN leapfrog.sh
Wide character in print at /usr/bin/iwatch line 359.
[27/  नव ह्बर /2021    21:32:11]      IN_CLOSE_NOWRITE leapfrog.sh
Wide character in print at /usr/bin/iwatch line 359.
[27/  नव ह्बर /2021    21:32:11]      IN_OPEN leapfrog.sh
Wide character in print at /usr/bin/iwatch line 359.
[27/  नव ह्बर /2021    21:32:11]      IN_ACCESS leapfrog.sh
Wide character in print at /usr/bin/iwatch line 359.
[27/  नव ह्बर /2021    21:32:11]      IN_CLOSE_NOWRITE leapfrog.sh
```

While appending log to leapfrog.log

iwatch -e access,close_nowrite,close_write,open leapfrog.sh &>> leapfrog.log

```
bibek@bibek-LfTech:~/assignment/iwatch$ cat leapfrog.log
Wide character in print at /usr/bin/iwatch line 359.
Wide character in print at /usr/bin/iwatch line 359.
Wide character in print at /usr/bin/iwatch line 359.
Wide character in print at /usr/bin/iwatch line 359.
Wide character in print at /usr/bin/iwatch line 359.
Wide character in print at /usr/bin/iwatch line 359.
Wide character in print at /usr/bin/iwatch line 359.
Wide character in print at /usr/bin/iwatch line 359.
bibek@bibek-LfTech:~/assignment/iwatch$
```

Only this message comes, the actual message was not appended

I personally found auditd tool best over iwatch

4. Install logstash in your system. download a sample nginx log from https://github.com/elastic/examples/blob/master/Common%20Data%20Formats/nginx_logs/nginx_logs , parse the logs using logstash. The parsed output must contain the geographical information like country, state etc. that the request is originating from. save the parsed output to a file in your system.

Installing Logstash

Download and install the Public Signing Key:

wget -qO - https://artifacts.elastic.co/GPG-KEY-elasticsearch | sudo apt-key add -

```
bibek@bibek-LfTech:~$ wget -qO - https://artifacts.elastic.co/GPG-KEY-elasticsea
rch | sudo apt-key add -
[sudo] password for bibek:
OK
bibek@bibek-LfTech:~$
```

Install the apt-transport-https package on Debian before proceeding:

sudo apt-get install apt-transport-https

Save the repository definition to /etc/apt/sources.list.d/elastic-7.x.list:

echo "deb https://artifacts.elastic.co/packages/7.x/apt stable main" | sudo tee -a /etc/apt/sources.list.d/elastic-7.x.list

Run sudo apt-get update and the repository is ready for use. You can install it with:

sudo apt-get update && sudo apt-get install logstash

```
Fetched 374 MB in 3min 54s (1,600 kB/s)
Selecting previously unselected package logstash.
(Reading database ... 173377 files and directories currently installed.)
Preparing to unpack .../logstash_1%3a7.15.2-1_amd64.deb ...
Unpacking logstash (1:7.15.2-1) ...
Setting up logstash (1:7.15.2-1) ...
Using bundled JDK: /usr/share/logstash/jdk
Using provided startup.options file: /etc/logstash/startup.options
OpenJDK 64-Bit Server VM warning: Option UseConcMarkSweepGC was deprecated in ve
rsion 9.0 and will likely be removed in a future release.
/usr/share/logstash/vendor/bundle/jruby/2.5.0/gems/pleaserun-0.0.32/lib/pleaseru
n/platform/base.rb:112: warning: constant ::Fixnum is deprecated
Successfully created system startup script for Logstash
bibek@bibek-LfTech:~$
```

Configuring logstash

sudo vi /etc/logstash

And make these changes

```
#
path.data: /var/lib/logstash
#
```

```
#
path.config: /etc/logstash/conf.d
#
```

```
# log.level: info
path.logs: /var/log/logstash
#
```

The nginx log file is downloaded and I renamed it with .log extension

```
bibek@bibek-LfTech:~/Downloads$ ls
nginx_logs.log
bibek@bibek-LfTech:~/Downloads$
```

Writing conf file for logstash

sudo vi /etc/logstash/conf.d/logstash-nginx.conf

```
input {
 file {
 type => "nginx-log"
 path => "/home/bibek/Downloads/nginx_logs.log"
 sincedb_path => "/dev/null"
 start_position => "beginning"
}
}

filter {
  grok {
    match => { "message" => %{ IPV4 } }
  }
  geoip {
    source => "IPV4"
  }
}

output {
 file {
 path => "/var/log/logstash/logstashnginx.log"
 }
}
```

Matching **IPV4** because grok filters IP with the name IPv4 and giving same **IPV4** as source to geoip.

Starting the logstash

sudo systemctl start logstash

sudo systemctl status logstash

```
bibek@bibek-LfTech:/etc/logstash/conf.d$ sudo systemctl status logstash.service

● logstash.service - logstash
     Loaded: loaded (/etc/systemd/system/logstash.service; enabled; vendor pre>
     Active: active (running) since Sat 2021-11-27 23:59:46 +0545; 5s ago
   Main PID: 15461 (java)
      Tasks: 14 (limit: 2946)
     Memory: 210.9M
     CGroup: /system.slice/logstash.service
             └─15461 /usr/share/logstash/jdk/bin/java -Xms1g -Xmx1g -XX:+UseCo>

नव ह्बर  27  23:59:46      bibek-LfTech systemd[1]: Started logstash.
नव ह्बर  27  23:59:46      bibek-LfTech logstash[15461]: Using bundled JDK: /usr/share/>
नव ह्बर  27  23:59:46      bibek-LfTech logstash[15461]: OpenJDK 64-Bit Server VM warni>
```

Checking for the log format in the nginx_log in the /var/log/logstash directory

tail -f /var/log/logstash/logstashnginx.log


```
{"@version":"1","type":"nginx-log","clientip":"80.91.33.133","message":"80.91.3
3.133 - - [04/Jun/2015:07:06:16 +0000] \"GET /downloads/product_1 HTTP/1.1\" 30
4 0 \"-\" \"Debian APT-HTTP/1.3 (0.8.16~exp12ubuntu10.16)\"","@timestamp":"2021
-11-27T18:51:16.517Z","host":"bibek-LfTech","geoip":{"latitude":59.955,"region_
name":"Oslo County","region_code":"03","country_code2":"NO","continent_code":"E
U","timezone":"Europe/Oslo","ip":"80.91.33.133","longitude":10.859,"location":{
"lon":10.859,"lat":59.955},"city_name":"Oslo","country_code3":"NO","postal_code
":"0114","country_name":"Norway"},"path":"/home/bibek/Downloads/nginx_logs.log"
}
{"@version":"1","type":"nginx-log","clientip":"144.76.151.58","message":"144.76
.151.58 - - [04/Jun/2015:07:06:05 +0000] \"GET /downloads/product_2 HTTP/1.1\"
304 0 \"-\" \"Debian APT-HTTP/1.3 (0.9.7.9)\"","@timestamp":"2021-11-27T18:51:1
6.518Z","host":"bibek-LfTech","geoip":{"latitude":51.1811,"region_name":"North
Rhine-Westphalia","region_code":"NW","country_code2":"DE","continent_code":"EU"
,"timezone":"Europe/Berlin","ip":"144.76.151.58","longitude":7.2171,"location":
{"lon":7.2171,"lat":51.1811},"city_name":"Remscheid","country_code3":"DE","post
al_code":"42855","country_name":"Germany"},"path":"/home/bibek/Downloads/nginx_
logs.log"}
{"@version":"1","type":"nginx-log","clientip":"79.136.114.202","message":"79.13
6.114.202 - - [04/Jun/2015:07:06:35 +0000] \"GET /downloads/product_1 HTTP/1.1\
" 404 334 \"-\" \"Debian APT-HTTP/1.3 (0.8.16~exp12ubuntu10.22)\"","@timestamp"
:"2021-11-27T18:51:16.518Z","host":"bibek-LfTech","geoip":{"latitude":59.3274,"
region_name":"Stockholm County","region_code":"AB","country_code2":"SE","contin
ent_code":"EU","timezone":"Europe/Stockholm","ip":"79.136.114.202","longitude":
18.0653,"location":{"lon":18.0653,"lat":59.3274},"city_name":"Stockholm","count
ry_code3":"SE","postal_code":"113 85","country_name":"Sweden"},"path":"/home/bi
bek/Downloads/nginx_logs.log"}
```


We can clearly see the country name, city name and other information