ASSIGNMENT

Problem Statement

- Install Logstash. This can be on a PC/laptop or in a virtual machine.
- Create a Logstash configuration file. The Logstash configuration file should do the following:
  - Allow for ingestion of the **Security Log Data** below via a "Logstash Input Plugin", your choice.
  - Allow for output of data via a "Logstash Output Plugin", your choice.
  - Utilizing any number of "Logstash Filter Plugins", transform data from the **Security Log Data** from its original format to a normalized structure described below.
- Design your solution so that a different piece of **Security Log Data** structured the same, but containing different data in the log message, could be ingested and achieve a similar output based on the data contained within the log.
- Send the **Security Log Data** example to Logstash and capture the output from Logstash in some form (text, screenshot, etc). Include this in the Git repo.
- Save the Logstash configuration created above in a public Git repo.
- Write a short description of how data is ingested and output from Logstash based on your solution. Include this write-up in the Git repo.

Security Log Data:

<14>1 2016-12-25T09:03:52.754646-06:00 acmeinchost1 antivirus 2496 - - alertname="Virus Found" computername="acmeincpc42" computerip="10.58.194.142" severity="1"

Output as JSON formatted data containing (but not exclusively) the following fields:

description: "Virus Found"

hostname: " acmeincpc42"

source_ip: "10.58.194.142"

severity: "High"

Solution

1. Installed Elasticsearch, Kibana and Logstash and configured to connect through local server
2. Verified the working of ELK

3. I used the **stdin** and **stdout** as input and output plugin to do **console testing** wherein, I am ingesting data (a sample log) from the terminal and seeing the ouput on the terminal as well.
4. Initially, I created the config file by directly parsing the fields from the log in the **grok** pattern. Attached below is the config file and the output:

Config File.txt

Ouput

```
Windows PowerShell
[2025-04-05T23:29:58,936][WARN ][logstash.filters.grok    ][main] ECS v8 support is a preview of the unreleased ECS v8, and uses the
v1 patterns. When Version 8 of the Elastic Common Schema becomes available, this plugin will need to be updated
[2025-04-05T23:29:58,938][INFO ][logstash.outputs.elasticsearch][main] Using a default mapping template {:es_version=>8, :ecs_compati
bility=>:v8}
[2025-04-05T23:29:59,096][INFO ][logstash.javapipeline    ][main] Starting pipeline {:pipeline_id=>"main", "pipeline.workers"=>8, "pi
peline.batch.size"=>125, "pipeline.batch.delay"=>50, "pipeline.max_inflight"=>1000, "pipeline.sources"=>["D:/Elastic Stack/Logstash/l
ogstash-8.17.4/config/logstash.conf"], :thread=>"#<Thread:0x7e3f7da3 D:/Elastic Stack/Logstash/logstash-8.17.4/logstash-core/lib/logs
tash/java_pipeline.rb:138 run>"}
[2025-04-05T23:30:00,435][INFO ][logstash.javapipeline    ][main] Pipeline Java execution initialization time {"seconds"=>1.34}
[2025-04-05T23:30:00,471][INFO ][logstash.javapipeline    ][main] Pipeline started {"pipeline.id"=>"main"}
The stdin plugin is now waiting for input:
[2025-04-05T23:30:00,485][INFO ][logstash.agent           ] Pipelines running {:count=>1, :running_pipelines=>[:main], :non_running_p
ipelines=>[]}
1 2016-12-25T09:03:52.754646-06:00 acmeinchost1 antivirus 2496 - - alertname="Virus Found" computername="acmeincpc42" computerip="10.
58.194.142" severity="1"
{
     "event_type" => "antivirus",
    "description" => "Virus Found",
    "device_name" => "acmeincpc42",
       "severity" => "High",
       "hostname" => "acmeinchost1",
      "source_ip" => "10.58.194.142",
     "@timestamp" => 2016-12-25T15:03:52.754Z,
            "pid" => "2496"
}
```

5. Further **fine-tuning** the config file, I used the **kv** filter to segregate the remaining part of the message which had a key and a value. But the output had a backtracking due to **severity** value giving **string** values when it is configured to support **long-int** values

Config_File_With_KVP_Technique.txt

Output

```
Windows PowerShell
[2025-04-05T23:48:05,742][WARN ][logstash.filters.grok    ][main] ECS v8 support is a preview of the unreleased ECS v8, and uses the
v1 patterns. When Version 8 of the Elastic Common Schema becomes available, this plugin will need to be updated
[2025-04-05T23:48:05,748][INFO ][logstash.outputs.elasticsearch][main] Using a default mapping template {:es_version=>8, :ecs_compati
bility=>:v8}
[2025-04-05T23:48:06,079][INFO ][logstash.javapipeline    ][main] Starting pipeline {:pipeline_id=>"main", "pipeline.workers"=>8, "pi
peline.batch.size"=>125, "pipeline.batch.delay"=>50, "pipeline.max_inflight"=>1000, "pipeline.sources"=>["D:/Elastic Stack/Logstash/l
ogstash-8.17.4/config/logstash.conf"], :thread=>"#<Thread:0xde792bf D:/Elastic Stack/Logstash/logstash-8.17.4/logstash-core/lib/logst
ash/java_pipeline.rb:138 run>"}
[2025-04-05T23:48:08,372][INFO ][logstash.javapipeline    ][main] Pipeline Java execution initialization time {"seconds"=>2.29}
[2025-04-05T23:48:08,518][INFO ][logstash.javapipeline    ][main] Pipeline started {"pipeline.id"=>"main"}
The stdin plugin is now waiting for input:
[2025-04-05T23:48:08,559][INFO ][logstash.agent           ] Pipelines running {:count=>1, :running_pipelines=>[:main], :non_running_p
ipelines=>[]}
1 2016-12-25T09:03:52.754646-06:00 acmeinchost1 antivirus 2496 - - alertname="Virus Found" computername="acmeincpc42" computerip="10.
58.194.142" severity="1"
{
      "source_ip" => "10.58.194.142",
            "pid" => "2496",
     "event_type" => "antivirus",
    "description" => "Virus Found",
    "device_name" => "acmeincpc42",
       "hostname" => "acmeinchost1",
       "severity" => "High",
     "@timestamp" => 2016-12-25T15:03:52.754Z
}
```

In both the above configs, I used the **rubydebug** codec to parse the output data

6. Analyzing further, I see that using a different label for **severity** field would be an ideal approach as I observed that field mappings were blocked.
   Hence, I modified the config file to parse **severity** in an additional field **device_severity** and mapped the **string** value to this field.

   <u>Input Plugin</u>
   stdin

   <u>Filters Used</u>
   grok – To match the header part of the log
   kv – To split the key and value from the remaining part of the log
   mutate + **gsub** – To remove any double quotes in the severity value which has an integer
   translate – For converting integer severity values to strings with “Low”, “Medium” and “High” severity levels
   mutate + **rename** – Renamed the field names given in the **grok** pattern to match to match the problem statement
   mutate + **remove_fields** – Removed some of the fields from the **stdout** output JSON

   <u>Output Plugin</u>
   stdout

   <u>Elasticsearch Index and Host</u>
   hosts => ["http://localhost:9200"]
   index => "test.logstash"

   Attached below is the config file and the output:

   logstash.conf

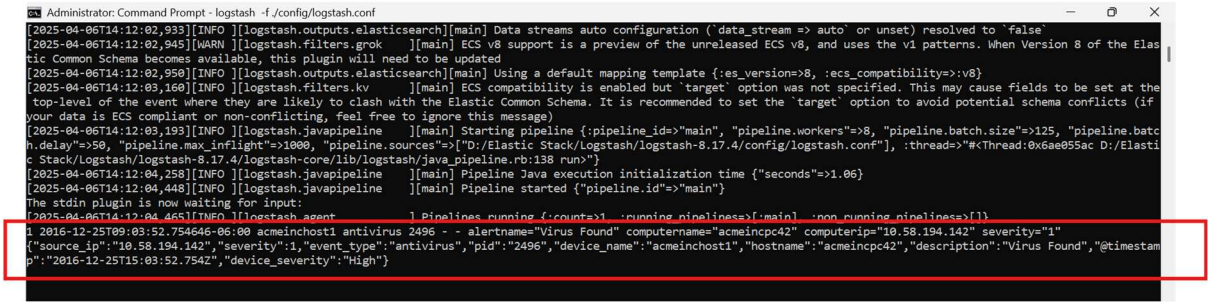
   <u>Input</u>

```
D:\Elastic Stack\Logstash\logstash-8.17.4\bin>logstash -f ./config/logstash.conf
"Using bundled JDK: D:\Elastic Stack\Logstash\logstash-8.17.4\jdk\bin\java.exe"
Sending Logstash logs to D:/Elastic Stack/Logstash/logstash-8.17.4/logs which is now configured via log4j2.properties
[2025-04-06T12:50:51,901][WARN ][logstash.runner          ] NOTICE: Running Logstash as a superuser is strongly discouraged as it poses a security risk. Set 'allow_superuse
r' to false for better security.
[2025-04-06T12:50:51,907][INFO ][logstash.runner          ] Log4j configuration path used is: D:\Elastic Stack\Logstash\logstash-8.17.4\config\log4j2.properties
[2025-04-06T12:50:51,907][INFO ][logstash.runner          ] Starting Logstash {"logstash.version"=>"8.17.4", "jruby.version"=>"jruby 9.4.9.0 (3.1.4) 2024-11-04 547c6b150e O
penJDK 64-Bit Server VM 21.0.6+7-LTS on 21.0.6+7-LTS +indy +jit [x86_64-mswin32]"}
[2025-04-06T12:50:51,907][INFO ][logstash.runner          ] JVM bootstrap flags: [-Xms1g, -Xmx1g, -Djava.awt.headless=true, -Dfile.encoding=UTF-8, -Djruby.compile.invokedyn
amic=true, -XX:+HeapDumpOnOutOfMemoryError, -Djava.security.egd=file:/dev/urandom, -Dlog4j2.isThreadContextMapInheritable=true, -Dlogstash.jackson.stream-read-constraints.m
ax-string-length=200000000, -Dlogstash.jackson.stream-read-constraints.max-number-length=10000, -Djruby.regexp.interruptible=true, -Djdk.io.File.enableADS=true, --add-expor
ts=jdk.compiler/com.sun.tools.javac.api=ALL-UNNAMED, --add-exports=jdk.compiler/com.sun.tools.javac.file=ALL-UNNAMED, --add-exports=jdk.compiler/com.sun.tools.javac.parser=
ALL-UNNAMED, --add-exports=jdk.compiler/com.sun.tools.javac.tree=ALL-UNNAMED, --add-exports=jdk.compiler/com.sun.tools.javac.util=ALL-UNNAMED, --add-opens=java.base/java.se
curity=ALL-UNNAMED, --add-opens=java.base/java.io=ALL-UNNAMED, --add-opens=java.base/java.nio.channels=ALL-UNNAMED, --add-opens=java.base/sun.nio.ch=ALL-UNNAMED, --add-open
s=java.management/sun.management=ALL-UNNAMED, -Dio.netty.allocator.maxOrder=11]
[2025-04-06T12:50:51,965][INFO ][org.logstash.jackson.StreamReadConstraintsUtil] Jackson default value override `logstash.jackson.stream-read-constraints.max-string-length`
 configured to `200000000`
[2025-04-06T12:50:51,965][INFO ][org.logstash.jackson.StreamReadConstraintsUtil] Jackson default value override `logstash.jackson.stream-read-constraints.max-number-length`
 configured to `10000`
[2025-04-06T12:50:52,015][WARN ][logstash.config.source.multilocal] Ignoring the 'pipelines.yml' file because modules or command line options are specified
[2025-04-06T12:50:52,941][INFO ][logstash.agent           ] Successfully started Logstash API endpoint {:port=>9600, :ssl_enabled=>false}
[2025-04-06T12:50:53,501][INFO ][org.reflections.Reflections] Reflections took 96 ms to scan 1 urls, producing 152 keys and 530 values
[2025-04-06T12:50:55,531][WARN ][logstash.filters.translate] You are using a deprecated config setting "destination" set in translate. Deprecated settings will continue to
```

Output


```
{
 "source_ip": "10.58.194.142",
 "severity": 1,
 "event_type": "antivirus",
 "pid": "2496",
 "device_name": "acmeinchost1",
 "hostname": "acmeincpc42",
 "description": "Virus Found",
 "@timestamp": "2016-12-25T15:03:52.754Z",
 "device_severity": "High"
}
```


Final Output from Terminal



In the above output, I mapped **string** values of **severity** to **device_severity** and used **JSON** as the **codec.**

The remaining attributes are mapped as per what is expected in the output.

Hence, the final version of the config file that I came up with, is attached below:

logstash.conf