Aggregation, indexing and visualization using the ELK Stack

by

Bradlee Gene Beadle

A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Major: Information Assurance and Computer Engineering

Program of Study Committee:
Doug Jacobson, Major Professor

Iowa State University

Ames, Iowa

2019

DEDICATION and ACKNOWLEDGEMENTS

The dedication of my creative component goes out to my family. My grandmother, Vondean, my parents, Chuck and Cheryl, my wife, Denise, and my children, Malevon and Audriena. I appreciate each and every one of you. I LOVE YOU!!!

I would like to express my sincere gratitude to the many individuals who have given their help, support, and belief in me during my academic journey. Chris Garrison, Cimone Wright-Hamor, Zach Beese, and Nik Kinkel have been amazing. Thank you!!

TABLE OF CONTENTS

CHAPTER 1. OVERVIEW

My idea of what this final creative component of my education should be is related to the next step in the journey of life. I'm not aspiring to break new ground at this point, I am aspiring to teach those who desire to learn. I want to help build a strong foundation of learning for each student or individual I have the privilege to engage with in whatever capacity that entails.

With that in mind, I decided to approach this last experience from a different perspective. This is not something that has never been done before. However, it is something that is not traditionally submitted as a creative component. Of course, I am going to submit the paper portion as required. What makes this more important to me is the additional video component.

The next intention in my life is to become an adjunct instructor at community college. Not only to fulfill a commitment, but to advance the abilities of those I come in contact with while fostering an excitement for education. I want to drive others to be better than the day before, to stand for what they believe in, and to offer support as well as encouragement to achieve their dreams.

This creative component is going to be a tutorial on how to do a simple install of the a log analysis platform in a network environment. This paper will provide screenshots and verbiage on each step I took to accomplish the endeavor. In addition, I will create video tutorials that will allow the viewer to see the installation procedure firsthand in hopes that viewing of such material will help those who enjoy learning with that medium.

1.1 Introduction

As a proponent of open source software, I chose the ELK Stack as the analytic engine for my creative component. This project is installed and maintained on a CyberCorps server.

1.1.1 Installation Environment

Environment specifications that the ELK Stack is performing in are contained below.

1.1.1.1 Hypervisor

VMWare ESXi 6.5

1.1.1.2 Virtual Machine Specifications

This VM was created with 4vCPUs, 4gb RAM and a 16gb hard drive. As the ELK Stack is updated and other services are added, the environment will be increased to allow for a smoother flow experience.

1.1.1.3 Operating System

Ubuntu 18.04

CHAPTER 2. ELASTICSEARCH

This chapter will cover the installation, tips and problems encountered while installing Elasticsearch.

NOTE: This is the only section that will include the public key and repository installation sequence as it only needs to be achieved once. Many of the steps may seem remedial to some, but are necessary in this order to ensure proper installation.

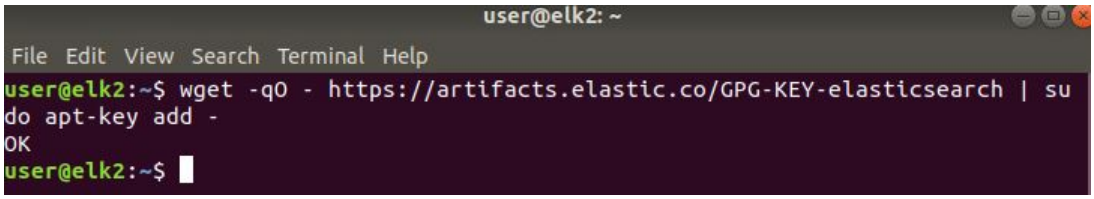
2.1 Installation

2.1.1 PGP Key

The Elk Stack package is signed with a PGP key. This public signing key must be downloaded and installed prior to any of the Elastic products.

2.1.1.1

Note that **'OK'** is the confirmation that the key has been added successfully. Enter the following command to download and install the PGP key:

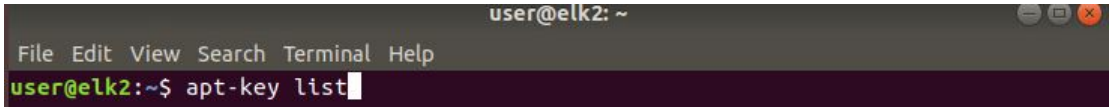


In this sequence, wget downloads the PGP key from the public key server. The key is piped to apt-get which adds it to the keyring. It is good practice to verify the key has been

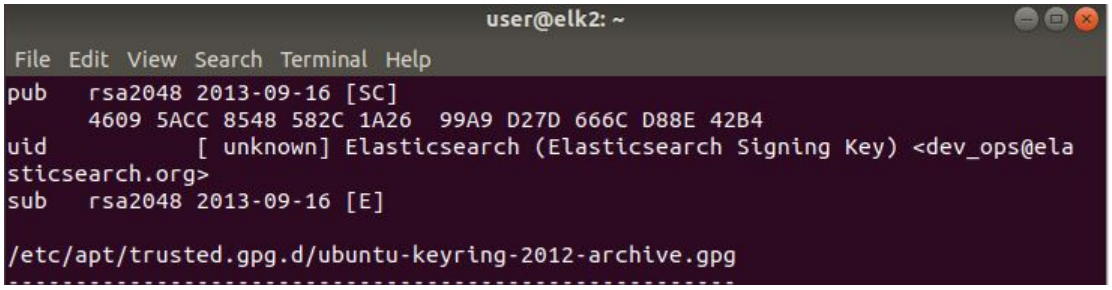
added to ensure trust in the repository.

2.1.1.2

Note that the cursor is shown prior to hitting enter. Enter the following command to ensure the Elasticsearch PGP key has been added to the trusted list:



Once the trusted list output is on the screen, you should see the Elasticsearch PGP key. Note that there will be multiple keys and you may have to scroll through the list to find the key. It should appear similar to the following:

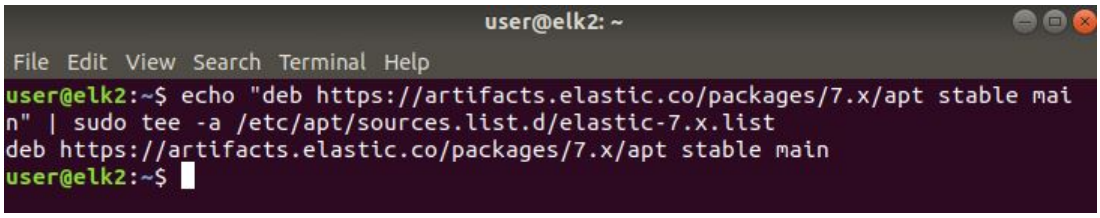


2.1.2 Repository

In order to download the ELK Stack, you must download the repository definition from the repository where the latest stable environment is located. As mentioned earlier, this is the only time you will have to perform this task.

2.1.2.1

Note that the third line is not included in your input as it is the output of the repository URL. Enter the following command to save the repository:



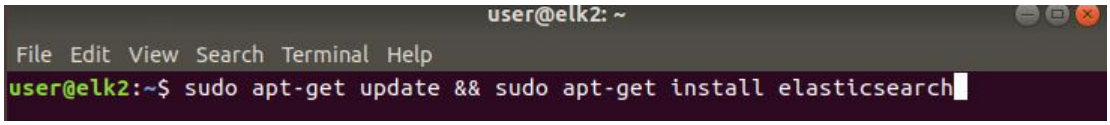
In this sequence, the web address is piped to the tee command. The URL is then echoed to the screen and is also saved as a repository definition in the ***/etc/apt/sources.list.d/*** folder.

2.1.3 Debian Package

This section will install the highly scalable Elasticsearch Debian package with the newest version of all required files.

2.1.3.1

Note that the cursor is shown prior to hitting enter. Enter the following command to install Elasticsearch:



At this point, patience is a virtue. There will be an enormous amount of streaming text while you wait for the installation to complete. That text will include reading package lists, building dependency tree, and fetched amount among other items.

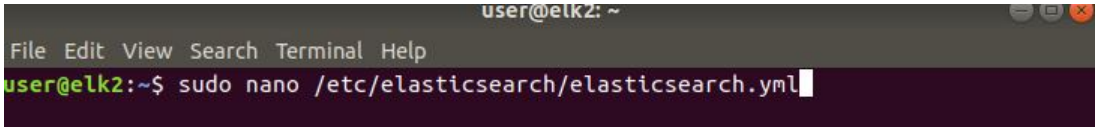
2.1.4 Configuration

It is important to understand where Elasticsearch loads its configuration from as well as how to edit it. The config file can be found at /etc/elasticsearch/elasticsearch.yml.

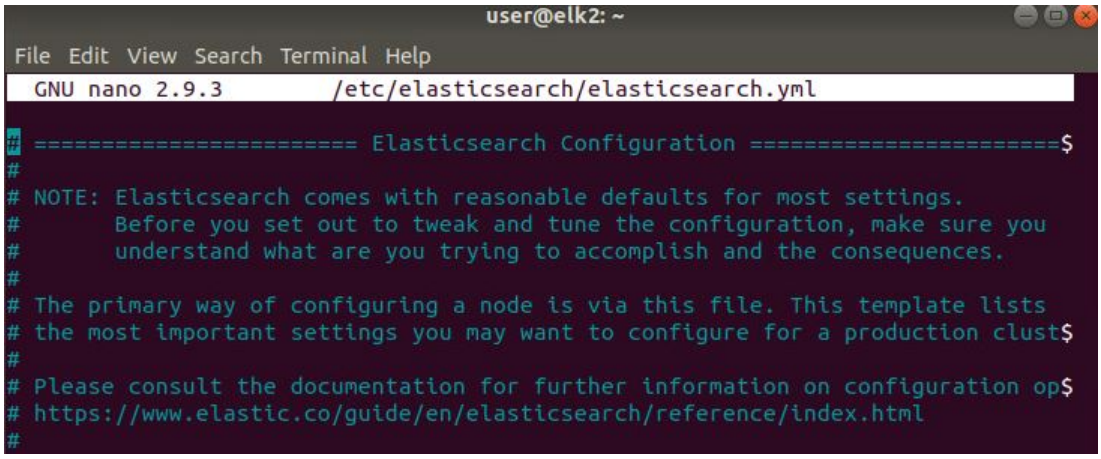
For this basic tutorial, we will only verify a couple of configuration entries using Nano, a text editor that is included with Ubuntu.

2.1.4.1

Note that the cursor is shown prior to hitting enter. Enter the following command to access the Elasticsearch configuration file using Nano:



In this sequence, Nano is used to open elasticsearch.yml for editing. Once Nano is open, the terminal window should look similar to this:

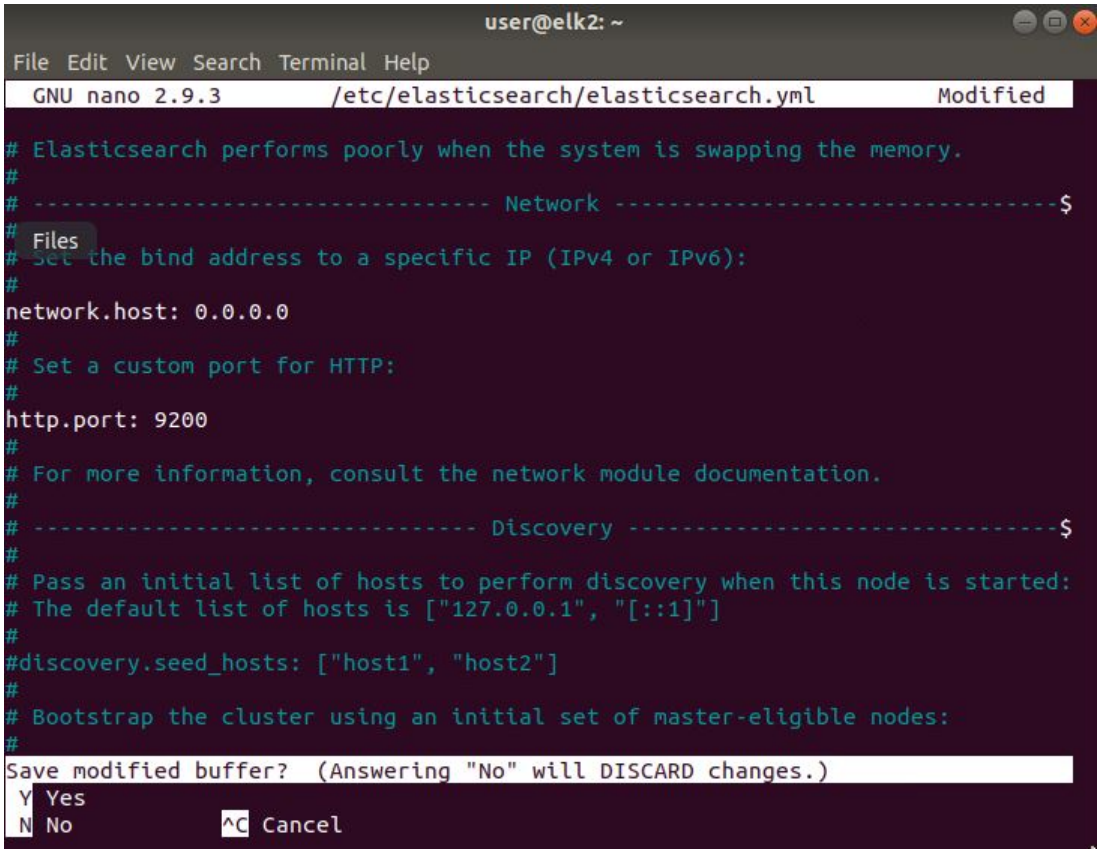


The view will have more information than the screenshot above as that is the first section of the file.

2.1.4.2

Using the arrow key, scroll down to the network section. There are only two changes that need to be made at this point. Delete the pound symbol before ***network.host*** and

before ***http.port***. Press ctrl-x to exit Nano. The terminal should now look similar to this:



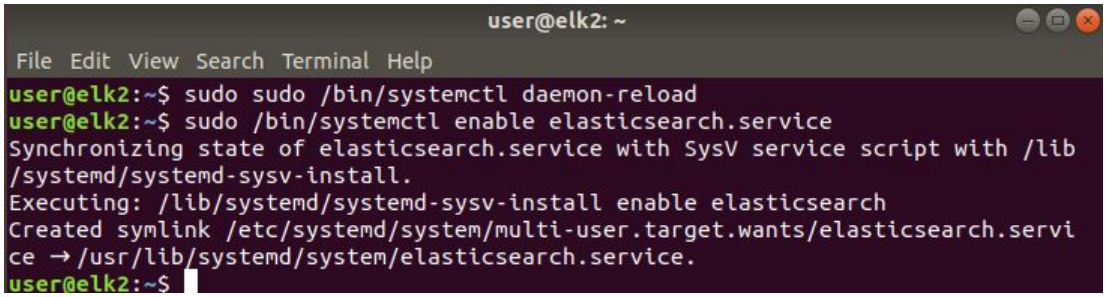
Press Y to save the file. At the file name to write prompt, press enter to confirm and exit the Nano text editor. Network host is now set to listen on all interfaces and the custom port is 9200.

2.1.5 Startup

This section will enable Elasticsearch to automatically start when the system boots up which is vital to ensure it is indexing and storing accordingly in the event of a power failure or reboot. I have also included the service commands to start and stop a program as you may have the desire or need to do so.

2.1.5.1

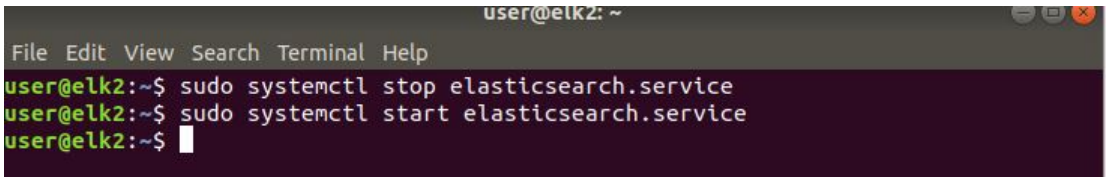
Note that user input is ONLY the first two lines in the image below. Enter the following commands to enable automatic start:



When an update is done to the configuration, you must execute daemon-reload. According to the man page, this reloads files and recreates the dependency tree while regenerating the configuration of the systemd manager. These commands must be entered in this sequence. (...)

2.1.5.2

The order of commands below is irrelevant at this time. This is just for informational purposes only. Enter the appropriate command to stop or start service:



Note that you will not receive confirmation that Elasticsearch has stopped or started.

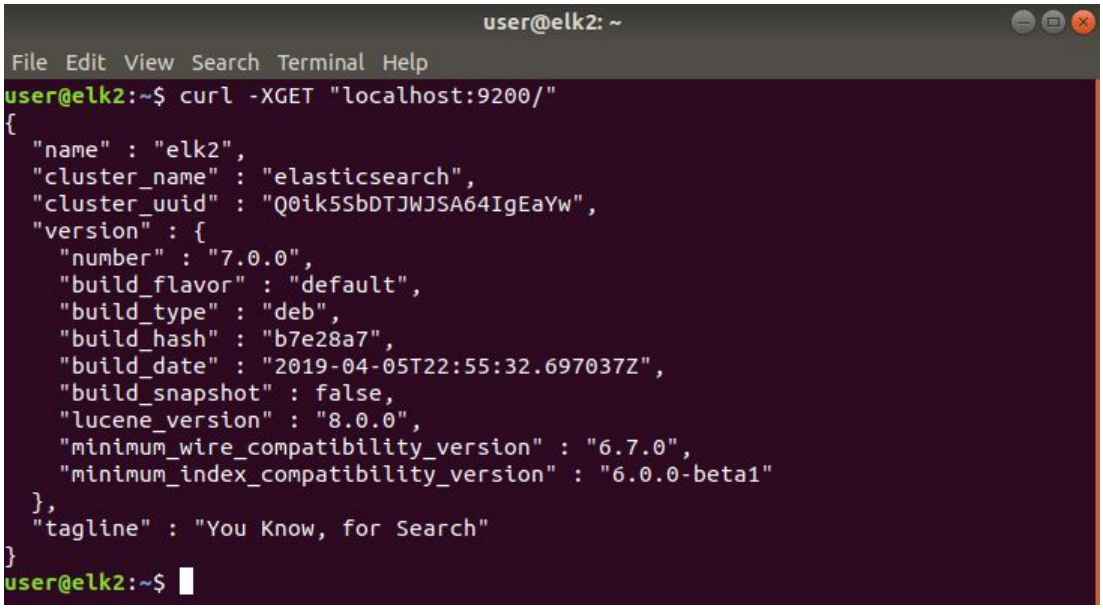
2.2 Is Elasticsearch running?

At this point in the installation, it is time to check if Elasticsearch is running. This section provides two options to verify if Elasticsearch is running. One via the command

line and the other via the browser. The command line is listed first as it is my preferred method.

2.2.1 Command Line

Note that the command is only the first line and the output follows. Enter the following command to verify Elasticsearch is running:

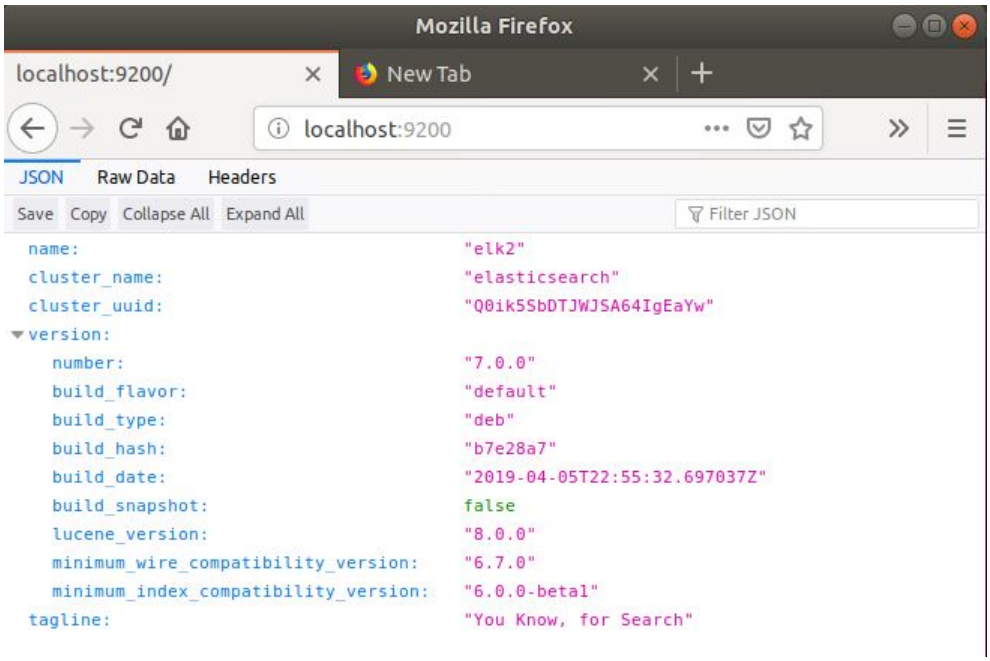


In this sequence, curl is the tool used to transfer data from the localhost. -XGET is a custom get request.

The response you receive should look similar to the above output if Elasticsearch is running. If it is not running, you will receive a ***connection refused error*** informing you of the failed connection. This error should be eliminated by starting Elasticsearch as detailed previously. Once started, input the curl command from this section again and the response should be active.

2.2.2 Browser

Open up the Firefox browser and put in localhost:9200 as the address. You should see output similar to the following if Elasticsearch is running:



If the Elasticsearch service is not running, you will be unable to connect. This error should be eliminated by starting the service as detailed previously. Once started, refresh the browser to see the connection. Note that it may take a few moments for Elasticsearch to start.

CHAPTER 3. KIBANA

.

With the verification Elasticsearch is running, the next step is to install Kibana. This service acts as the visualization tool that provides many features including histograms, graphs, charts, and maps among other user defined features.

3.1 Installation

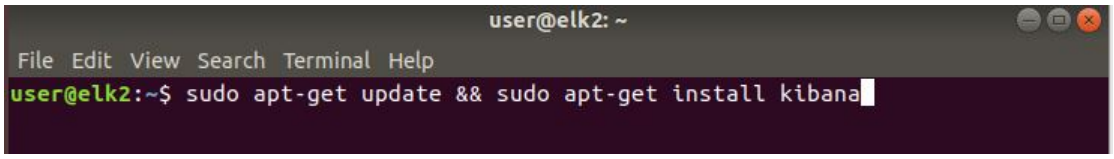
This chapter will cover the installation, tips and any problems encountered while installing Kibana. This will be the easiest part of the ELK stack installation process.

3.1.1 Debian Package

This section will install the Kibana Debian package with the newest version of all required files.

3.1.1.1

Note that the cursor is shown prior to hitting enter. Enter the following command to install Kibana:



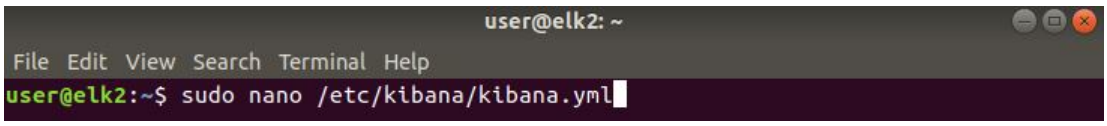
As this is a much smaller package than Elasticsearch, Kibana installation is much faster. Once again, screen text will include reading package lists, building dependency tree, and fetched amount among other items.

3.1.2 Configuration

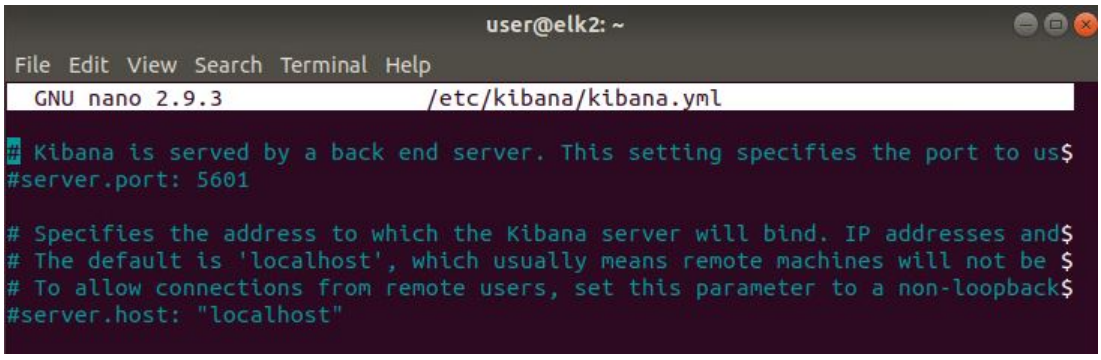
The Kibana configuration file can be found at /etc/kibana/kibanba.yml. Once again, Nano will be used to edit the file.

3.1.2.1

Note that the cursor is shown prior to hitting enter. Enter the following command to access the Kibana configuration file using Nano:



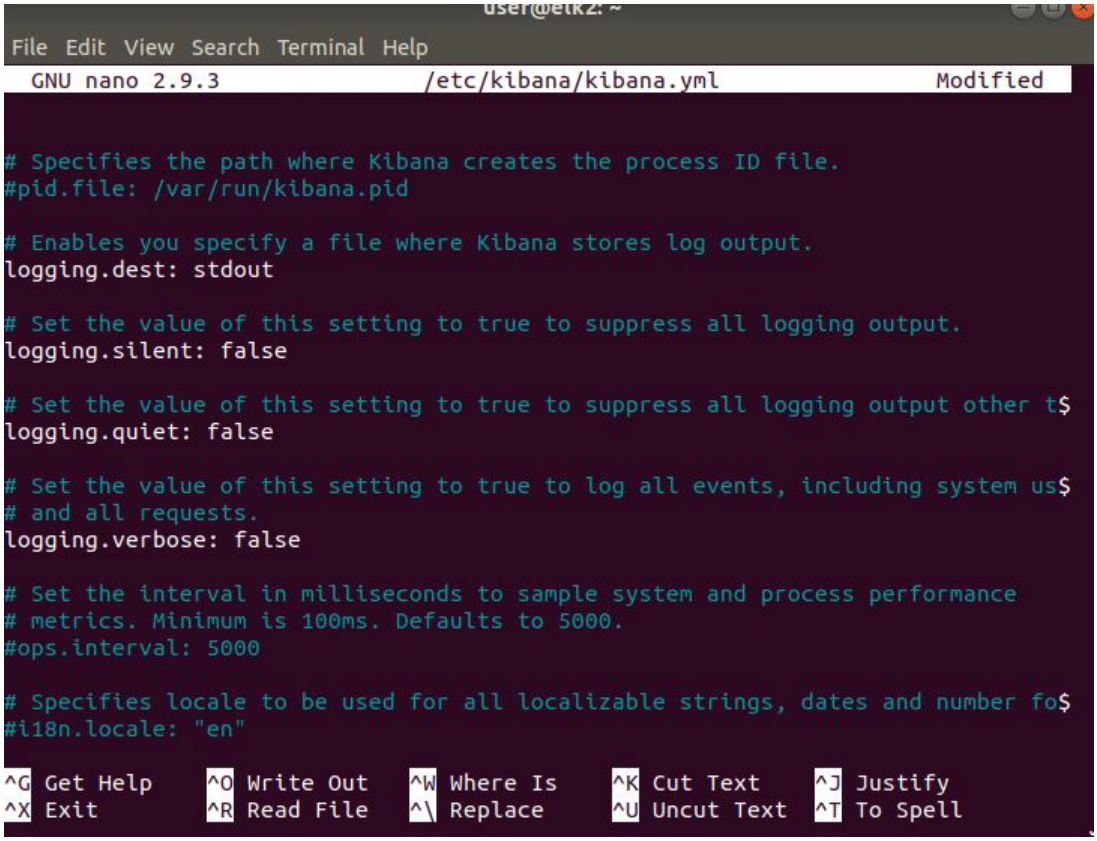
In this sequence, Nano is used to open kibana.yml for editing. Once Nano is open, the terminal window should look similar to this:



The view will have more information than the screenshot above as that is the first section of the file.

3.1.2.2

In this section, there are multiple changes that must be made. Use the arrow keys to navigate to the appropriate lines. Delete the pound symbol before the following line entries: ***server.port***, ***server.host*** changing its value to 0.0.0.0 instead of localhost, ***elasticsearch.hosts***, ***logging.dest***, ***logging.silent***, ***logging.quiet***, and ***logging.verbose*** changing its value to true. Press ctrl-x to exit Nano. The terminal should now look similar to this:



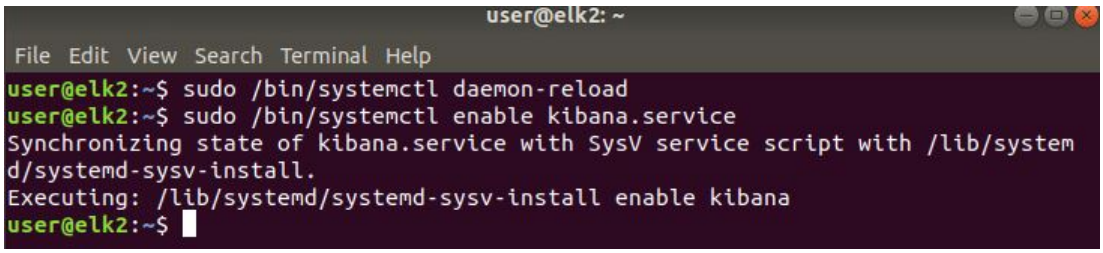
Press Y to save the file. At the file name to write prompt, press enter to confirm and exit the Nano text editor. Kibana is now set to listen on all interfaces, logging each event and run on localhost:5601.

3.1.2.3 Startup

This section will enable Kibana to automatically start when the system boots up which is vital for visualization and integration with Elasticsearch.

3.1.2.4

Note that user input is ONLY the first two lines in the image below. Enter the following commands to enable automatic start:



As we learned earlier, when an update is done to the configuration, you must execute daemon-reload and, once again, these commands must be entered in sequence.

CHAPTER 4. LOGSTASH

The next step is to install Logstash. As we have already verified the PGP key and the current version of Jave, this installation will be fairly simple.

4.1 Installation

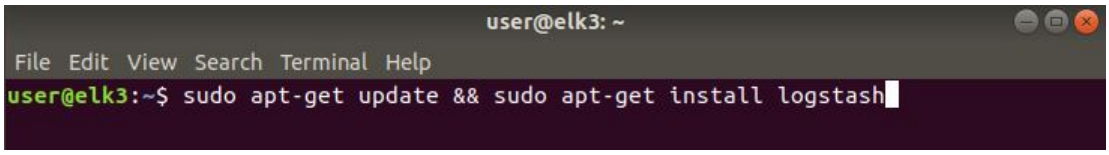
This chapter will cover the installation, tips and any problems encountered while installing Logstash.

4.1.1 Debian Package

This section will install the Logstash Debian package with the newest version of all required files.

4.1.1.1

Note that the cursor is shown prior to hitting enter. Enter the following command to install Logstash:



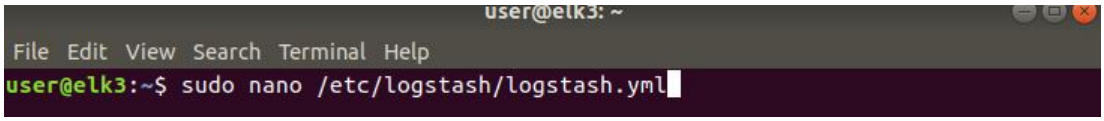
Once again, this is a quick download with screen text that will include reading package lists, building dependency tree, and fetched amount among other items.

4.1.2 Configuration

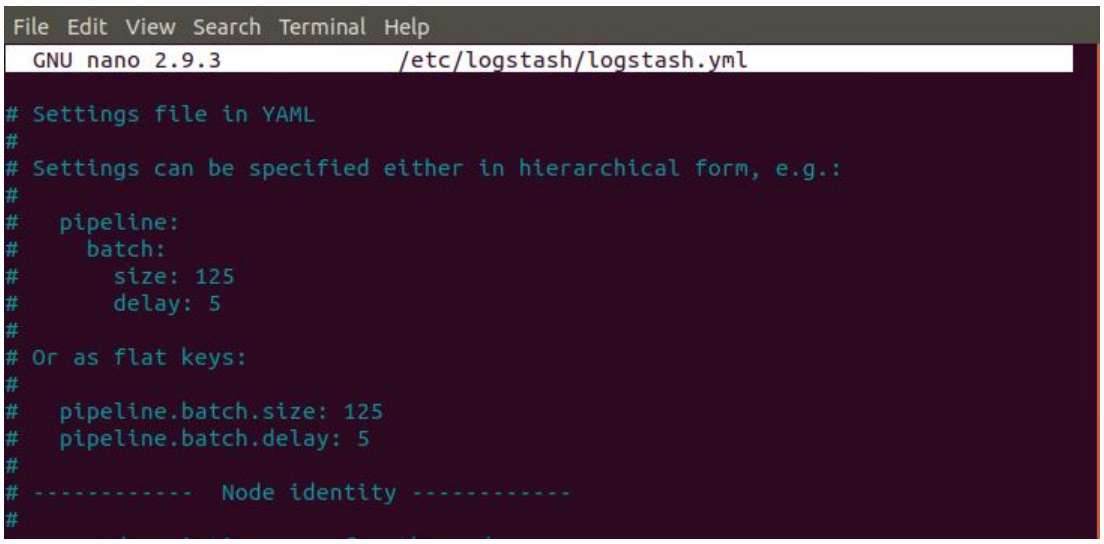
The Logstash configuration file can be found at /etc/logstash/logstash.yml. Once again, Nano will be used to edit the file.

4.1.2.1

Note that the cursor is shown prior to hitting enter. Enter the following command to access the Logstaash configuration file using Nano:



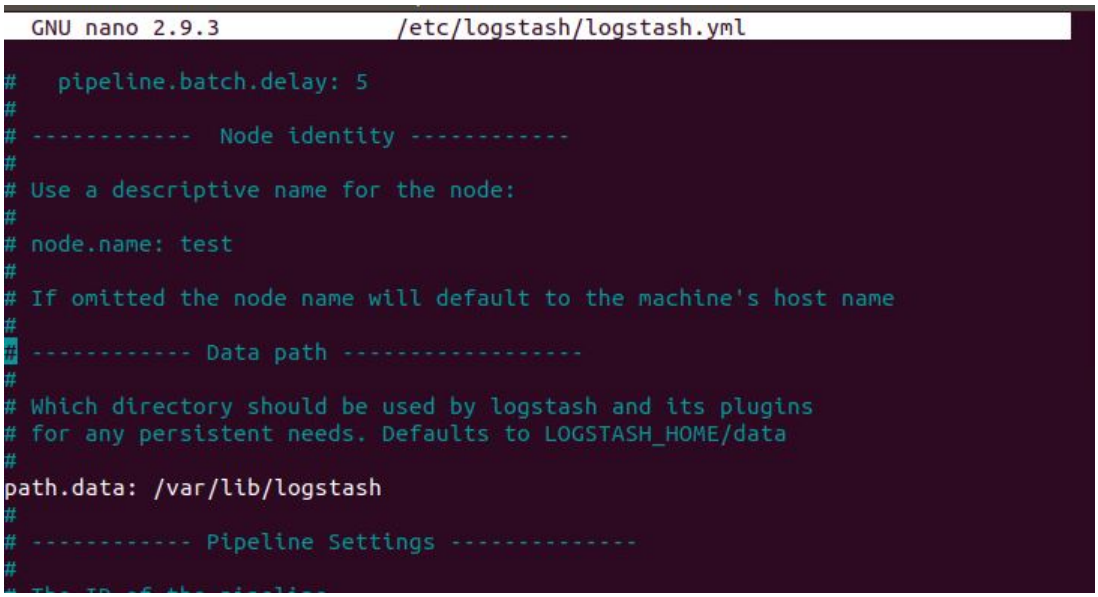
In this sequence, Nano is used to open logstash.yml for editing. Once Nano is open, the terminal window should look similar to this:

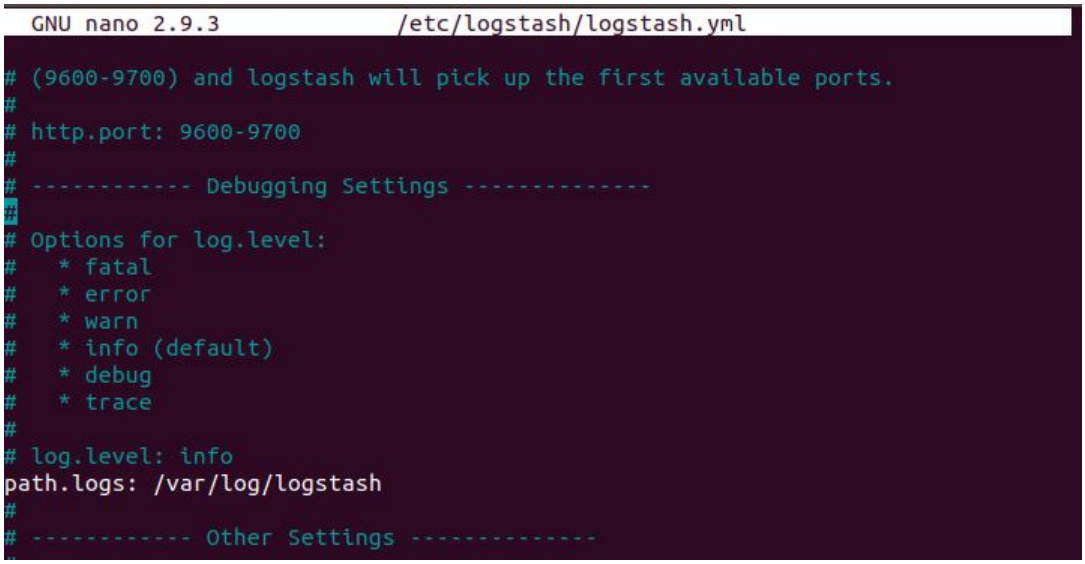


The view will have more information than the screenshot above as that is the first section of the file.

4.1.2.2

In this section, you need to ensure two lines are not commented out. Use the arrow keys to navigate to the appropriate lines. If needed, delete the pound symbol before the following line entries: ***path.data*** and ***path.logs***. The terminal should look similar to this for each:





Press ctrl-x to exit Nano and press Y to save the file. At the file name to write prompt, press enter to confirm and exit the Nano text editor.

4.1.3 Startup

This section will enable Logstash to become available.

4.1.3.1

Enter the following commands to start and enable automatic start:

```
:~$ systemctl start logstash
:~$ systemctl enable logstash
:~$
```

4.1.3.2

At this point, Logstash should be running and template file will need to be made based on the desired output needs for the environment. The following was selected as the template for this installation:

```
template(name="json-template"
  type="list") {The
    constant(value="{")
      constant(value="\"@timestamp\":\"")      property(name="timereported" dateFormat="rfc3339")
      constant(value="\",\"@version\":\"1")
      constant(value="\",\"message\":\"")      property(name="msg" format="json")
      constant(value="\",\"sysloghost\":\"")   property(name="hostname")
      constant(value="\",\"severity\":\"")     property(name="syslogseverity-text")
      constant(value="\",\"facility\":\"")     property(name="syslogfacility-text")
      constant(value="\",\"programname\":\"")  property(name="programname")
      constant(value="\",\"procid\":\"")       property(name="procid")
    constant(value="\"}\n")
}
```

This must be saved in the /etc/rsyslog.d folder as a ***.conf*** file. This is the system folder of rsyslog, which is an open source utility for message logging.

CHAPTER 5. CONCLUSION

This concludes the simple installation of the Elk Stack. Note that each installation will have to be configured according the expectations and restraints of that environment. With that in mind, research should be asserted that will further enhance all aspects of what is needed for any log analysis monitor. Care will need to be taken to ensure any expansion on this simple installation.

BIBLIOGRAPHY

Basic Concepts Elasticsearch Reference [7.0] Elastic. Elastic.Co, 2019, www.elastic.co/guide/en/elasticsearch/reference/current/getting-started-concepts.html, 7.0.

Berman, Daniel. The Complete Guide to the ELK Stack. Logz.Io, Logz.io, 23 May 2016, logz.io/learn/complete-guide-elk-stack/intro.

Elahi, Urfeena. Elastic Stack A Brief Introduction. Hacker Noon, Hacker Noon, 18 July 2018, hackernoon.com/elastic-stack-a-brief-introduction-794bc7ff7d4f.

Jain, Ayush. Introduction to Elasticsearch and the ELK Stack. Blogspot.Com, 2019, fullstackgeek.blogspot.com/2019/03/introduction-to-elasticsearch-and-elk-stack.html.

Ornbo, George. Linux and Unix Wget Command Tutorial with Examples. George Ornbo, 16 Sept. 2016, shapeshed.com/unix-wget/.