Graylog Cookbook

authored during

Big Data Internship

at

Sifast

in order to

Solve the Documentation Problem

of Graylog

by

Maryem ALOULOU

Isslem ZOUARI

« Graylog »

First Edition

supervised by :

Mr.	**Ghazi ABIDA**	*Industrial supervisor*
Mr.	**Ahmed KALLEL**	*Industrial supervisor*
Mr.	**Mohamed JMAL**	*Industrial supervisor*

Servers Setup

Graylog Server Side

To establish our Graylog server, we need a virtual machine equiped with at least the minimum hardware requirements and software dependencies of Graylog.

Software Dependencies

- **Java:** is a widely-used programming language known for its platform independence and versatility. In this tutoriel we will be using Java 8
- **MongoDB:** is a popular NoSQL database that provides a flexible and scalable way to store and retrieve data. In this tutoriel we will be using MongoDB 3.2.
- **Elasticsearch:** is a distributed search engine built on top of the Lucene library and can be used as a database. In this tutoriel we will be using Elasticsearch 7.x

> **Info**
> We can use OpenSearch or Elasticsearch. However, Elasticsearch is the commonly used.

- **Graylog:** is a log management and analysis tool. It allows you to collect, index, and analyze log data from various sources. In this tutoriel we will be using Graylog 4.2.

Hardware Requirements

The recommended resources' size may vary depending on your environment and the amount of data you are processing. Therefore, you can monitor the CPU and RAM usage of your Graylog environment accordingly.
If you are running all components (Graylog, Elasticsearch, MongoDB) on a single machine, reasonable sizes may be:

- **RAM Capacity :** 8GB (4GB of RAM for Elasticsearch and 4GB of RAM for Graylog).
- **CPU :** 4 with 4 cores each.

> **Tip**
> There are few rules you should take into consideration when scaling resources for Graylog:
>
> - Graylog nodes should have a focus on CPU power especially if the nodes are also serving the user interface to the browser.
>   --> The CPU power of Graylog nodes is crucial because log processing and indexing operations can be computationally intensive.
>   --> As log data is ingested, it needs to be parsed, normalized, and indexed for efficient searching and analysis. These operations can put a significant load on the CPU, especially when dealing with high volumes of log data.

--> Serving the user interface to the browser requires CPU resources to handle user requests, process search queries, generate visualizations, and deliver the interface elements to the browser in a timely manner. The responsiveness of the user interface depends on the CPU capacity available to handle these tasks efficiently.
--> To ensure optimal performance, it is recommended to provision Graylog nodes with an adequate amount of CPU power. The specific CPU requirements will depend on factors such as the volume of log data being processed, the complexity of parsing and indexing operations, and the expected number of concurrent users accessing the interface.

- Elasticsearch/OpenSearch nodes are responsible for storing and searching log data. Thus, they should have as much RAM as possible and the fastest disks you can get. Everything depends on I/O speed here.
- Graylog uses MongoDB to store configuration data, not log data. This means that only metadata such as user information or stream configurations are stored in MongoDB. As a result, MongoDB does not have a big system impact and does not require many resources. MongoDB runs alongside the Graylog server processes and takes up minimal space

- **OS:** We can use an operating systems that is Linux-based such as Debian, Ubuntu, and CentOS. In this tutorial, we will be using Ubuntu 20.04.

> ⚠ **Caution**
> Running Graylog on Windows is not supported.

Installation & Configuration

1. Before any installation, make sure to execute the command:

```
sudo apt-get update -y
sudo apt-get upgrade -y
```

> **Hint**
>
> To avoid repeating the key word "sudo" in each command, you can execute at the beginning of your session:
>
> ```
> sudo su
> ```
>
> This will prompt you to enter your VM password, and then you can execute all commands as the root user. However, it is important to note that executing commands as the root user can be dangerous and should be done with caution.

2. **Java**

> **Info**
> Graylog is a Java-based application, and it requires a Java Runtime Environment (JRE) to run. Therefore, Java must be installed on the system before installing Graylog.

- To install java, you can follow these commands:

```
sudo apt-get install openjdk-8-jre -y
sudo apt-get install openjdk-8-jdk -y
```

> **Tip**
>
> After installing Java, it is recommended to set the environment variables to ensure that the system recognizes the Java installation correctly. There are 2 methods to set the environment variables:
> #Method1
> Through `/etc/environment` which is a system-wide configuration file that sets environment variables globally for all users on the system, regardless of their shell.
>
> - Use the following command to open the `/etc/environment` file in a text editor:

```
sudo nano /etc/environment
```

> - Within the text editor, add the following line at the end of the file:

```
JAVA_HOME="/usr/bin/java"
```

> - Press `Ctrl+s` to save the file, then press Enter to confirm the filename.
> - Press `Ctrl+x` to exit the text editor.
> - To apply the changes and update the environment variables, run the following command:

```
source /etc/environment
```

> #Method2
> Through `~/.bashrc` which is a user-specific configuration file that is executed whenever a new interactive Bash shell is opened for that particular user.
>
> - Use the following command to open the `~/.bashrc` file in a text editor:

```
sudo nano ~/.bashrc
```

.

- Within the text editor, add the following line at the end of the file:

```
export JAVA_HOME="/usr/bin/java"
```

- To apply the changes and update the environment variables, run the following command:

```
source ~/.bashrc
```

Hint

Verify if the environment variable is set correctly by running the following command:

```
echo $JAVA_HOME
```

You should see the output showing the path to your Java 8 installation.

Attention

If the installation directory of Java 8 is different on your system, you should adjust the path accordingly.

Success

You can verify if Java is installed correctly on your system using the following command:

```
java -version
```

When you run this command in the terminal, it will display the version information of the installed Java runtime.

3. **MongoDB**

Info
we should install mongodb first to store the configuration file of graylog

- Download the MongoDB public GPG key from a specific URL:

```
wget -qO - https://www.mongodb.org/static/pgp/server-4.4.asc
```

- Add the downloaded GPG key to the system's keyring, allowing the system to verify the authenticity of MongoDB packages.

```
sudo apt-key add -
```

- Add the MongoDB repository to the package manager (the list of package sources in `/etc/apt/sources.list.d/mongodb-org-4.4.list`):

```
echo "deb [ arch=amd64,arm64 ] https://repo.mongodb.org/apt/ubuntu
focal/mongodb-org/4.4 multiverse" | sudo tee
/etc/apt/sources.list.d/mongodb-org-4.4.list
```

- Update the package manager (update the package lists from all configured repositories, including the newly added MongoDB repository):

```
sudo apt-get update
```

- Install MongoDB. By using this command, the package manager will download and install the necessary MongoDB packages and dependencies:

```
sudo apt-get install -y mongodb-org
```

✓ Success

You can verify if MongoDB is installed correctly on your system using the following command:

```
mongod -version
```

When you run this command in the terminal, it will display the version information of the installed Java runtime.

- Reload the `systemd` configuration:

```
sudo systemctl daemon-reload
```

Tip

This command is generally used after making changes to service unit files to ensure that the changes take effect.

- Enable the `mongod.service` unit, which is responsible for managing the MongoDB service:

```
sudo systemctl enable mongod.service
```

Tip

Enabling the service means it will start automatically at boot time.

- Start the MongoDB service:

```
sudo systemctl start mongod.service
```

Success

To check if the MongoDB service is currently active, we can use 2 methods:

- Listing all active services on the system and filters the output to show only the services with `mongod` in their names.

```
systemctl --type=service --state=active | grep mongod
```

- Providing detailed information about the service, including whether it is running, any errors or warnings, and the most recent logs.

```
sudo systemctl status mongod
```

4. **Elasticsearch**

- Install `curl` which is a command-line tool used to transfer data over various protocols, including HTTP. we will use this command to download the Elasticsearch GPG key and for later use:

```
sudo apt install curl
```

- Import the Elasticsearch GPG key by retrieving it and adding it to your system's keyring, allowing you to verify the authenticity of the Elasticsearch packages during installation:

```
curl -fsSL https://artifacts.elastic.co/GPG-KEY-elasticsearch | sudo apt-key add -
```

- Add the Elasticsearch repository to the package manager (the list of package sources in `/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
```

- Update the package manager (update the package lists from all configured repositories, including the newly added Elasticsearch repository):

```
sudo apt update -y
```

- Install Elasticsearch. By using this command, the package manager will download and install the necessary files, including Elasticsearch itself and its dependencies:

```
sudo apt install elasticsearch
```

- Opens the Elasticsearch configuration file (`elasticsearch.yml`) in the Nano text editor to configure Elasticsearch.

```
sudo nano /etc/elasticsearch/elasticsearch.yml
```

> **Info**
>
> The Elasticsearch cluster or nodes are used by Graylog to store and search log data, and it needs to be properly configured to ensure that Graylog can access them. That's why, users should edit, using a text editor, the elasticsearch.yml file which is typically located in the `/etc/elasticsearch/` directory on Linux systems.

- You can make changes to the configuration, such as specifying network settings, cluster settings, or enabling/disabling specific features.
  - By default, Elasticsearch is only accessible on localhost. So, we should set a different address here to expose the node on the network.

    ```
    network.hosts= 192.168.1.41
    ```

  - users can modify the port that Elasticsearch will listen on. However, it is recommended to use the default value which is 9200

    ```
    http.port= 9200
    ```

- Reload the `systemd` configuration:

```
sudo systemctl daemon-reload
```

- Enable the `elasticsearch.service` unit, which is responsible for managing the MongoDB service:

```
sudo systemctl enable elasticsearch.service
```

- Start Elasticsearch:

```
sudo systemctl start elasticsearch
```

✓ Success

To check if the Elasticsearch service is currently active, we can use 2 methods:

- Listing all active services on the system and filters the output to show only the services with `elasticsearch` in their names.

```
systemctl --type=service --state=active | grep elasticsearch
```

- Providing detailed information about the service, including whether it is running, any errors or warnings, and the most recent logs.

```
sudo systemctl status elasticsearch
```

✓ Success

To verify Elasticsearch installation and network configuration you can use these commands:

- This command sends an HTTP GET request to the local Elasticsearch instance on port 9200. If Elasticsearch is running correctly, you should see a JSON response containing information about the Elasticsearch cluster:

```
curl 192.168.1.41:9200
```

- This command helps you verify if Elasticsearch is actively listening on the expected port. As we are using the port `9200`, It uses `netstat` to list all active network connections (`-a`), displays numeric addresses (`-n`), and filters the output to show only TCP connections on port `9200`:

```
netstat -a -n | grep tcp | grep 9200
```

- This command is used for network connections. It checks if a TCP connection can be established to the specified IP address (`192.168.1.41`) and port (`9200`).

```
nc -vz 192.168.1.41 9200
```

> ⚠ **Attention**
>
> Remember to restart Elasticsearch after making any changes to the configuration for them to take effect.

1. **Graylog**

- Download Graylog repository configuration package `graylog-4.2-repository_latest.deb` from the specified URL:

```
wget https://packages.graylog2.org/repo/packages/graylog-4.2-repository_latest.deb
```

- Install the Graylog repository package and add it to the system's package sources.:

```
sudo dpkg -i graylog-4.2-repository_latest.deb
```

- Update the repository cache (update the package lists from all configured repositories, including the newly added Graylog repository):

```
sudo apt update -y
```

- Install the Graylog server:

```
sudo apt install -y graylog-server
```

- Generate a secret fto secure the user passwords (The output will be used as the "password_secret" in the Graylog server configuration):

```
pwgen -N 1 -s 96
```

- Open the Graylog server configuration file `server.conf` in the Nano text editor to make the changes listed in the next steps.

```
sudo nano /etc/graylog/server/server.conf
```

- Set the `password_secret` in `server.conf` by copying the generated password and pasting it after the equal sign of the `password_secret` setting in the `server.conf` file.
- Set the `root_username` setting that will be used to log into the Graylog Web-UI. The default root user is named `admin`.
- Generate a hash for the Graylog admin user. This will be used to log into the Graylog Web-UI. For example if the password is `yourpassword`, the command should be:

```
echo -n "yourpassword: " && head -1 </dev/stdin | tr -d '\n' | sha256sum |
cut -d" " -f1` ``
```

- Set the `http_bind_address` setting using your server's IP address and it's recommended to use the `9000` port number.

```
http_bind_adress= 192.168.1.41:9000
```

- The last step is to enable Graylog during the operating system's startup and verify it is running.

```
sudo systemctl daemon-reload
sudo systemctl enable graylog-server.service
sudo systemctl start graylog-server.service
sudo systemctl status graylog-server.service
```

✓ Success

To verify Graylog installation and network configuration you can use these commands:

- This command sends an HTTP GET request to the local Graylog instance on port 9000. If Graylog is running correctly, you should see a HTML response containing information about the Graylog GUI:

```
curl 192.168.1.41:9200
```

- This command helps you verify if Graylog is actively listening on the expected port. As we are using the port 9000, It uses `netstat` to list all active network connections (`-a`), displays numeric addresses (`-n`), and filters the output to show only TCP connections on port 9000:

```
netstat -a -n | grep tcp | grep 9000
```

- This command is used for network connections. It checks if a TCP connection can be established to the specified IP address (`192.168.1.41`) and port (`9000`).

```
nc -vz 192.168.1.41 9000
```

⚠ Attention

you should enable the firewall to open the necessary ports and establish connections to the Graylog GUI.

```
sudo ufw enable
sudo ufw allow 9200
sudo ufw allow 9000
sudo ufw status
```

Through the firewall we can enhance the security

⚠ **Attention**

Remember to restart Elasticsearch after making any changes to the configuration for them to take effect.

✓ **Done**

You will be able to log into the web interface with the username **admin** and the password **yourpassword** after typing the URL http://192.168.1.41:9000 in the search bar of a web browser.

Client Server Side

In this tutorial we will consider that our client virtual machine is Setup with Zimbra server and we will use Filebeat as a log shipper to collect and forward logs to Graylog.

Hardware Requirements

Filebeat Installation & Configuration

- Add the Filebeat GPG key:

```
curl -fsSL https://artifacts.elastic.co/GPG-KEY-elasticsearch | sudo apt-key add -
```

- Add the Filebeat repository:

```
sudo add-apt-repository "deb https://artifacts.elastic.co/packages/7.x/apt stable main"
```

- Update the package manager:

```
sudo apt-get update
```

- Install Filebeat version 7.17.7:

```
sudo apt-get install filebeat=7.17.7
```

- Open the Filebeat configuration file `filebeat.yml` to configure Filebeat:

```
sudo nano /etc/filebeat/filebeat.yml
```

Info

The Filebeat configuration file is used to specify how Filebeat collects, processes, and sends log data:

- **Filebeat inputs:** This section defines the inputs from where Filebeat collects log data.
- **Filebeat modules:** This section enables specific modules and their configurations.
- **Filebeat outputs:** This section specifies the output where Filebeat sends the collected log data.
- **Filebeat logging:** This section defines the logging configuration for Filebeat itself.
- **Other Filebeat settings:** This section is for any additional configurations or settings specific to your use case.

- In the `Filebeat inputs` section, set the `enabled` setting to `true` and the `paths` setting to the path of the log files you want to collect:

```
  enabled: true
  paths:
    - /var/log/zimbra.log
```

- In the `Filebeat outputs` section, configure Filebeat to send data to a Graylog instance running at `192.168.1.41` on port `5044`

```
output.logstash:
  hosts: ["192.168.1.41:5044"]
```

Attention

you should enable the firewall to open the necessary ports and establish connections between Filebeat and Graylog:

```
sudo ufw enable
sudo ufw allow 5044
sudo ufw status
```

- The last step is to enable Graylog during the operating system's startup and verify it is running.

```
sudo systemctl daemon-reload
sudo systemctl enable filebeat.service
sudo systemctl start filebeat
sudo systemctl status filebeat
```

> ⚠ **Attention**
>
> Remember to restart Filebeat after making any changes to the configuration for them to take effect.

Graylog GUI

Create an input

To create an input in Graylog 4.2, you can follow these steps:

1. Log in to your Graylog web interface using your credentials.
2. Click on the "System" menu item in the top navigation bar.
3. From the drop-down menu, select "Inputs."
4. On the Inputs page, click the "Create Input" button.
5. Choose the input type based on the kind of data you want to receive. Graylog supports various input types, such as Syslog, GELF (Graylog Extended Log Format), Beats, and more. Select the appropriate input type for your needs.
6. Configure the input parameters based on the selected input type. The configuration options will differ based on the selected input type, so make sure to provide the necessary details as per your requirements.
7. After configuring the input parameters, click the "Save" button to create the input.

Create pipeline

To create pipelines in Graylog 4.2, you can follow these steps:

1. Log in to your Graylog web interface using your credentials.
2. In the Graylog dashboard, click on the "System" menu item in the top navigation bar.
3. From the drop-down menu, select "Pipelines."
4. On the Pipelines page, click the "Create Pipeline" button.
5. In the "Pipeline Configuration" section, provide a unique name and description for your pipeline.
6. In the "Rule Set" section, you can either select an existing rule set or create a new one. If you don't have an existing rule set, click the "Create Rule Set" button to create a new one.
7. In the Rule Set Editor, you can define the stages that make up your pipeline.

8. Click the "Add Stage" button to create a new stage.
9. Specify the stage priority, which determines the order in which stages are executed. Stages with the same priority run concurrently.
10. Within the stage, click the "Add Rule" button to create a new rule.
11. Define the conditions for the rule, specifying when the rule should be applied. For example, you can define conditions based on message fields, source IP address, or message content.
12. Add the actions for the rule, specifying what should be done with the log messages if the conditions are met. Actions can include modifying fields, setting new field values, dropping or blocking messages, sending notifications, or forwarding messages to another destination.
13. Repeat steps 10-12 to add more rules within the stage if needed.
14. Repeat steps 8-13 to add more stages to your pipeline if needed.
15. Once you have defined all the stages and rules for your pipeline, click the "Save" button in the Rule Set Editor.
16. Back in the pipeline creation page, you will see the selected rule set and the rules that are part of it.
17. Your pipeline is now created and ready to be used. You can associate the pipeline with one or more streams, inputs, or extractors to control the processing of messages from the section “Pipeline connections”.
18. To validate and test your pipeline ,click the "Simulator" tab.
19. Click the "Simulate" button to process the sample message(s) through the pipeline.
20. Analyze the simulated output to ensure the pipeline is processing the messages as expected.

Create streams

To create a stream in Graylog 4.2, you can follow these steps:

1. Log in to your Graylog web interface using your credentials.
2. Click on the "Streams" menu item in the top navigation bar.
3. On the Streams page, click the "Create Stream" button.
4. Provide the following information to configure the stream:
   - Stream Title: Enter a descriptive name for your stream.
   - Stream Description (Optional): Optionally, you can provide a brief description of the stream to provide more context.
   - Stream Index Set: Select the index set that should be used for indexing messages in this stream. If no index set is specified, the default index set will be used.
5. In the "Stream Rules" section, you can define rules to filter and route messages to the stream based on specific criteria. Click the "Add Stream Rule" button to add a rule.
   - Field: Select the log message field that you want to use for the rule.
   - Type: Choose the type of comparison you want to perform on the selected field.
   - Value: Enter the value or pattern to match against the field.
6. You can add multiple stream rules to fine-tune the filtering and routing of messages to the stream.
7. After configuring the stream settings, click the "Save" button to create the stream.

8. Once the stream is created, you can start receiving and routing messages to the stream based on the defined rules.

Create dashboards

To create a dashboard in Graylog, follow these steps:

1. Log in to your Graylog web interface and navigate to the Dashboards section using the link in the top menu bar.
2. Click on the "Create new dashboard" button to create a new empty dashboard.
3. Once the empty dashboard is created, click on the "Save as" button on the right side of the search bar to save the dashboard. A modal will open.
4. In the modal, provide a title for the dashboard. The title should be specific and easily understandable. You can also add a summary and description for more detailed information about the dashboard.
5. After providing the required information, click on the "Save" button to save the dashboard.
6. Now, you can start adding widgets to the dashboard. Widgets allow you to display search result information on the dashboard. To add a widget, click on the "Add widget" button.
7. Select the type of widget you want to add. Graylog provides various types of widgets such as line charts, bar charts, tables, etc. Choose the appropriate widget type for your needs.
8. Configure the widget by specifying the widget-specific search criteria. This includes the time range, search query, and stream selection. Use the search bar inside the widget edit modal to define these options.
9. Customize the widget settings according to your preferences. You can specify the widget title, size, visualization options, and other display settings.
10. Repeat steps 7-9 to add more widgets to the dashboard. You can add multiple widgets of different types to create a comprehensive view of your data.
11. Once you have added and configured all the desired widgets, click on the "Save" button to save the changes to the dashboard.
12. Your dashboard is now ready. You can view it and access the information by navigating to the Dashboards section and selecting the created dashboard.
13. You can share the dashboard with others by providing them with the appropriate permissions. Depending on the permissions granted, you can share the dashboard with co-workers, managers, or specific departments such as sales and marketing.