# Architectural considerations

1.1 Minimum setup ................................................. 3
1.2 Bigger production setup ...................................... 4
1.3 Graylog Architecture Deep Dive ............................ 5

# Getting Started

2.1 Download Graylog ............................................. 7
2.2 Import the VM ................................................... 8
2.3 Start VM ......................................................... 9
2.4 Connect to the Web Console .................................. 10
2.5 Explore Graylog ............................................... 12
2.6 Get Messages In ............................................... 17
2.7 Check If You Have Messages ............................... 19
2.8 If You Don’t Have Messages ............................... 20
2.9 Create Your Dashboard ...................................... 20
2.10 Get Alerted .................................................... 26

# Installing Graylog

3.1 Virtual Machine Appliances ................................. 35
3.2 Operating System Packages .................................. 39
3.3 Chef, Puppet, Ansible ....................................... 48
3.4 Docker .......................................................... 49
3.5 Vagrant ........................................................ 52
3.6 OpenStack ...................................................... 53
3.7 Amazon Web Services ...................................... 54
3.8 Microsoft Windows ......................................... 55
3.9 Manual Setup ................................................ 56
3.10 System requirements ....................................... 60

# Upgrading Graylog

4.1 Upgrading to Graylog 2.0.x .................................. 61
4.2 Upgrading to Graylog 2.1.x .................................. 65
4.3 Upgrading to Graylog 2.2.x .................................. 68

# Configuring Graylog

5.1 The graylog-ctl script ....................................... 71
5.2 Web interface .................................................. 81
5.3 Load balancer integration ................................... 86
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.4</td>
<td>Using HTTPS</td>
<td>87</td>
</tr>
<tr>
<td>5.5</td>
<td>Multi-node Setup</td>
<td>93</td>
</tr>
<tr>
<td>5.6</td>
<td>Elasticsearch</td>
<td>95</td>
</tr>
<tr>
<td>5.7</td>
<td>Index model</td>
<td>105</td>
</tr>
<tr>
<td>5.8</td>
<td>Backup</td>
<td>111</td>
</tr>
<tr>
<td>5.9</td>
<td>Default file locations</td>
<td>111</td>
</tr>
<tr>
<td>5.10</td>
<td>Graylog REST API</td>
<td>114</td>
</tr>
<tr>
<td>6</td>
<td>Securing Graylog</td>
<td>119</td>
</tr>
<tr>
<td>6.1</td>
<td>Default ports</td>
<td>119</td>
</tr>
<tr>
<td>6.2</td>
<td>Logging user activity</td>
<td>119</td>
</tr>
<tr>
<td>7</td>
<td>Sending in log data</td>
<td>121</td>
</tr>
<tr>
<td>7.1</td>
<td>What are Graylog message inputs?</td>
<td>121</td>
</tr>
<tr>
<td>7.2</td>
<td>Content packs</td>
<td>121</td>
</tr>
<tr>
<td>7.3</td>
<td>Syslog</td>
<td>121</td>
</tr>
<tr>
<td>7.4</td>
<td>GELF / Sending from applications</td>
<td>122</td>
</tr>
<tr>
<td>7.5</td>
<td>Using Apache Kafka as transport queue</td>
<td>122</td>
</tr>
<tr>
<td>7.6</td>
<td>Using RabbitMQ (AMQP) as transport queue</td>
<td>123</td>
</tr>
<tr>
<td>7.7</td>
<td>Microsoft Windows</td>
<td>123</td>
</tr>
<tr>
<td>7.8</td>
<td>Heroku</td>
<td>123</td>
</tr>
<tr>
<td>7.9</td>
<td>Ruby on Rails</td>
<td>125</td>
</tr>
<tr>
<td>7.10</td>
<td>Raw/Plaintext inputs</td>
<td>126</td>
</tr>
<tr>
<td>7.11</td>
<td>JSON path from HTTP API input</td>
<td>126</td>
</tr>
<tr>
<td>7.12</td>
<td>Reading from files</td>
<td>127</td>
</tr>
<tr>
<td>8</td>
<td>Graylog Collector Sidecar</td>
<td>129</td>
</tr>
<tr>
<td>8.1</td>
<td>Backends</td>
<td>130</td>
</tr>
<tr>
<td>8.2</td>
<td>Installation</td>
<td>130</td>
</tr>
<tr>
<td>8.3</td>
<td>Configuration</td>
<td>132</td>
</tr>
<tr>
<td>8.4</td>
<td>Secure Sidecar Communication</td>
<td>142</td>
</tr>
<tr>
<td>8.5</td>
<td>Run Sidecar as non-root user</td>
<td>142</td>
</tr>
<tr>
<td>8.6</td>
<td>Sidecar Glossary</td>
<td>143</td>
</tr>
<tr>
<td>8.7</td>
<td>Debug</td>
<td>145</td>
</tr>
<tr>
<td>8.8</td>
<td>Known Problems</td>
<td>146</td>
</tr>
<tr>
<td>9</td>
<td>Graylog Collector (deprecated)</td>
<td>147</td>
</tr>
<tr>
<td>9.1</td>
<td>Installation</td>
<td>147</td>
</tr>
<tr>
<td>9.2</td>
<td>Configuration</td>
<td>151</td>
</tr>
<tr>
<td>9.3</td>
<td>Running Graylog Collector</td>
<td>158</td>
</tr>
<tr>
<td>9.4</td>
<td>Command Line Options</td>
<td>160</td>
</tr>
<tr>
<td>9.5</td>
<td>Troubleshooting</td>
<td>161</td>
</tr>
<tr>
<td>10</td>
<td>Searching</td>
<td>163</td>
</tr>
<tr>
<td>10.1</td>
<td>Search query language</td>
<td>163</td>
</tr>
<tr>
<td>10.2</td>
<td>Time frame selector</td>
<td>165</td>
</tr>
<tr>
<td>10.3</td>
<td>Saved searches</td>
<td>166</td>
</tr>
<tr>
<td>10.4</td>
<td>Histogram</td>
<td>167</td>
</tr>
<tr>
<td>10.5</td>
<td>Analysis</td>
<td>168</td>
</tr>
<tr>
<td>10.6</td>
<td>Decorators</td>
<td>170</td>
</tr>
<tr>
<td>10.7</td>
<td>Export results as CSV</td>
<td>174</td>
</tr>
<tr>
<td>10.8</td>
<td>Search result highlighting</td>
<td>175</td>
</tr>
<tr>
<td>10.9</td>
<td>Search configuration</td>
<td>176</td>
</tr>
<tr>
<td>11</td>
<td>Streams</td>
<td>181</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
<td></td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>11.1 What are streams?</td>
<td>181</td>
<td></td>
</tr>
<tr>
<td>11.2 How do I create a stream?</td>
<td>182</td>
<td></td>
</tr>
<tr>
<td>11.3 Index Sets</td>
<td>182</td>
<td></td>
</tr>
<tr>
<td>11.4 Outputs</td>
<td>183</td>
<td></td>
</tr>
<tr>
<td>11.5 Use cases</td>
<td>184</td>
<td></td>
</tr>
<tr>
<td>11.6 How are streams processed internally?</td>
<td>184</td>
<td></td>
</tr>
<tr>
<td>11.7 Stream Processing Runtime Limits</td>
<td>185</td>
<td></td>
</tr>
<tr>
<td>11.8 Programmatic access via the REST API</td>
<td>186</td>
<td></td>
</tr>
<tr>
<td>11.9 FAQs</td>
<td>188</td>
<td></td>
</tr>
<tr>
<td>12 Alerts</td>
<td>191</td>
<td></td>
</tr>
<tr>
<td>12.1 Alert states</td>
<td>191</td>
<td></td>
</tr>
<tr>
<td>12.2 Alerts overview</td>
<td>191</td>
<td></td>
</tr>
<tr>
<td>12.3 Conditions</td>
<td>193</td>
<td></td>
</tr>
<tr>
<td>12.4 Notifications</td>
<td>195</td>
<td></td>
</tr>
<tr>
<td>13 Dashboards</td>
<td>201</td>
<td></td>
</tr>
<tr>
<td>13.1 Why dashboards matter</td>
<td>201</td>
<td></td>
</tr>
<tr>
<td>13.2 How to use dashboards</td>
<td>202</td>
<td></td>
</tr>
<tr>
<td>13.3 Examples</td>
<td>204</td>
<td></td>
</tr>
<tr>
<td>13.4 Widgets from streams</td>
<td>205</td>
<td></td>
</tr>
<tr>
<td>13.5 Result</td>
<td>205</td>
<td></td>
</tr>
<tr>
<td>13.6 Widget types explained</td>
<td>205</td>
<td></td>
</tr>
<tr>
<td>13.7 Modifying dashboards</td>
<td>206</td>
<td></td>
</tr>
<tr>
<td>13.8 Dashboard permissions</td>
<td>208</td>
<td></td>
</tr>
<tr>
<td>14 Extractors</td>
<td>211</td>
<td></td>
</tr>
<tr>
<td>14.1 The problem explained</td>
<td>211</td>
<td></td>
</tr>
<tr>
<td>14.2 Graylog extractors explained</td>
<td>211</td>
<td></td>
</tr>
<tr>
<td>14.3 Import extractors</td>
<td>213</td>
<td></td>
</tr>
<tr>
<td>14.4 Using regular expressions to extract data</td>
<td>215</td>
<td></td>
</tr>
<tr>
<td>14.5 Using Grok patterns to extract data</td>
<td>215</td>
<td></td>
</tr>
<tr>
<td>14.6 Using the JSON extractor</td>
<td>216</td>
<td></td>
</tr>
<tr>
<td>14.7 Automatically extract all key=value pairs</td>
<td>217</td>
<td></td>
</tr>
<tr>
<td>14.8 Normalization</td>
<td>219</td>
<td></td>
</tr>
<tr>
<td>15 Processing Pipelines</td>
<td>223</td>
<td></td>
</tr>
<tr>
<td>15.1 Pipelines</td>
<td>223</td>
<td></td>
</tr>
<tr>
<td>15.2 Rules</td>
<td>224</td>
<td></td>
</tr>
<tr>
<td>15.3 Stream connections</td>
<td>226</td>
<td></td>
</tr>
<tr>
<td>15.4 Functions</td>
<td>226</td>
<td></td>
</tr>
<tr>
<td>15.5 Usage</td>
<td>238</td>
<td></td>
</tr>
<tr>
<td>16 Message rewriting with Drools</td>
<td>243</td>
<td></td>
</tr>
<tr>
<td>16.1 Getting Started</td>
<td>243</td>
<td></td>
</tr>
<tr>
<td>16.2 Example rules file</td>
<td>243</td>
<td></td>
</tr>
<tr>
<td>16.3 Parsing Message and adding fields</td>
<td>244</td>
<td></td>
</tr>
<tr>
<td>16.4 Blacklisting messages</td>
<td>245</td>
<td></td>
</tr>
<tr>
<td>17 Blacklisting</td>
<td>247</td>
<td></td>
</tr>
<tr>
<td>17.1 How to</td>
<td>247</td>
<td></td>
</tr>
<tr>
<td>17.2 Based on regular expressions</td>
<td>247</td>
<td></td>
</tr>
<tr>
<td>18 Geolocation</td>
<td>249</td>
<td></td>
</tr>
<tr>
<td>18.1 Setup</td>
<td>249</td>
<td></td>
</tr>
</tbody>
</table>
27.57 Graylog2 0.20.2 .................................................. 355

28 Introduction .................................................. 357

29 Setup .................................................. 359
  29.1 Requirements ............................................. 359
  29.2 Installation ................................................ 359
  29.3 Server Restart .......................................... 360
  29.4 Cluster Setup .......................................... 361
  29.5 License Installation ................................... 361

30 Archiving .................................................. 363
  30.1 Setup ..................................................... 363
  30.2 Usage ..................................................... 368

31 Audit Log .................................................. 375
  31.1 Setup ..................................................... 375
  31.2 Usage ..................................................... 379

32 Changelog ................................................ 383
  32.1 Graylog Enterprise 2.2.3 ............................... 383
  32.2 Graylog Enterprise 2.2.2 ............................... 383
  32.3 Graylog Enterprise 2.2.1 ............................... 383
  32.4 Graylog Enterprise 2.2.0 ............................... 383
  32.5 Graylog Enterprise 1.2.1 ............................... 384
  32.6 Graylog Enterprise 1.2.0 ............................... 384
  32.7 Graylog Enterprise 1.1 ................................. 384
  32.8 Graylog Enterprise 1.0.1 ............................... 384
  32.9 Graylog Enterprise 1.0.0 ............................... 385
NOTE: There are multiple options for reading this documentation. See link to the lower left.
There are a few rules of thumb when scaling resources for Graylog:

- Graylog nodes should have a focus on CPU power. These also serve the user interface to the browser.
- Elasticsearch nodes should have as much RAM as possible and the fastest disks you can get. Everything depends on I/O speed here.
- MongoDB is storing meta information and configuration data and doesn’t need many resources.

Also keep in mind that ingested messages are only stored in Elasticsearch. If you have data loss in the Elasticsearch cluster, the messages are gone - except if you have created backups of the indices.

**Minimum setup**

This is a minimum Graylog setup that can be used for smaller, non-critical, or test setups. None of the components is redundant but it is easy and quick to setup.
Our Virtual Machine Appliances are using this design by default, deploying nginx as frontend proxy.

**Bigger production setup**

This is a setup for bigger production environments. It has several Graylog nodes behind a load balancer distributing the processing load.

The load balancer can ping the Graylog nodes via HTTP on the Graylog REST API to check if they are alive and take dead nodes out of the cluster.
How to plan and configure such a setup is covered in our *Multi-node Setup guide*. Some guides on the Graylog Marketplace also offer some ideas how you can use RabbitMQ (AMQP) or Apache Kafka to add some queueing to your setup.

**Graylog Architecture Deep Dive**

If you are really interested in the Graylog architecture at a more detailed level - whether you want to understand more for planning your architecture design, performance tuning, or just because you love stuff like that, our cheeky engineering team has put together this deep architecture guide. It’s not for the faint at heart, but we hope you love it.
Getting Started

To get the first look what Graylog is and how you can work with we have a virtual appliance to run on VMWare or Virtualbox. The following will guide you to start using Graylog.

After you feel comfortable you should make a fresh installation that takes care of your recommendation regarding the sizing and the security settings. The virtual appliance is only created to have a showcase about the features and is not created to run in production.

If you need assistance to plan and build your logging environment we offer professional support that can work with you.

**Download Graylog**

Point your favorite Browser to our Download page and get the latest stable version. If you are unsure what the latest version number is, take a look at our release page.
Import the VM

These screenshots are for VMWare (VirtualBox is nearly the same). Select File -> Import..., choose the graylog.ova file you downloaded, and follow the prompts.
Start VM

This is what you’ll see when you run the virtual machine. This is where all the Graylog server processes (more on that later) and the database will run. We can split them apart later for performance, but there’s no need to do that right now for a quick overview. Don’t close this window just yet, we’re going to need the IP for the next step.
The above steps are also covered in our virtual machine appliance installation page with some additional information. If you do not have DHCP enabled in your network you need to assign a static IP.

Connect to the Web Console

Go to a web browser on your host machine and type the IP address you saw in the previous screen. You should get a Graylog Web Console login page. Enter admin/admin to login.
Logging in will get you to blank screen. Not very fun? No problem, let’s get some stuff in there!
You can dismiss the Guide, and just follow all steps in this section to get started with Graylog.

**Explore Graylog**

By default the appliance is already indexing log messages from all components running inside the virtual machine. If you don’t want to record these messages, you can disable this internal logging using *The graylog-ctl script*. 
When opening the Streams page in the top menu, you will see a list of all predefined streams. Think of them as a filtered view of all incoming messages. You can find more information on the streams page. As you can see, you can define rules for the messages which should be visible in this stream.
If you click on *Sources* in the top menu, you will see a nice overview of which clients are sending data into Graylog and how much. In the showcase setup, you will see “graylog-server” and “nginx” in the beginning. Once you start sending data from other systems, their hostnames or IP addresses will be listed on the *Sources* page.
The most important question is: How will the data actually come into Graylog.

Open the *System > Inputs* page from the top menu and you will see some inputs which are running by default, such as a syslog input.

This means that you can already use the running appliance as a syslog receiver for any device supporting the syslog protocol which is able to connect to the virtual machine on port 514/udp (which happens to be the standard port for syslog).
If you would like to know the supported options for ingesting messages into Graylog, you can find additional information at **Sending in log data**.

To show what’s possible with Graylog, we have already created a few dashboards for you in the virtual appliance. Navigate to the **Dashboards** page in the top menu and click on **nginx overview**. After this you will see the following screen with a dashboard generated from the internal log data of the appliance itself.
To have some more data, we will add the syslog data from the running virtual machine in the next step.

Get Messages In

Log into the VM

We’re going to use rsyslog because we already have it running on the virtual machine. So, go to the image and login with ubuntu/ubuntu.
Modify rsyslog.conf

Go to the /etc directory, and use vi, vim (vim Cheat Sheet), or the editor of your choice to modify the /etc/rsyslog.conf file. There are excellent resources on the web for rsyslog configuration.

At the bottom of the file, add the following so messages will forward:

```
*.@127.0.0.1:514;RSYSLOG_SyslogProtocol23Format
```

In case you wanted to know, @ means UDP, 127.0.0.1 is localhost, and 514 is the port.

You can find out more about ingesting syslog messages with Graylog in our Syslog configuration guide.

Restart rsyslog

Type:

```
$ sudo service rsyslog status
$ sudo service rsyslog restart
```
If you have modified the config file and it is somehow invalid, the service command will not bring rsyslog back up - so don’t worry, you can always delete the line!

**Ingesting more log messages**

Please refer to *Sending in log data* for further instructions about configuring Graylog and ingesting data from external sources.

**Check If You Have Messages**

After that go back to the System > Inputs page.

Click on the Show received messages button next to the appliance-syslog-udp input and you should see only the messages received from your syslog clients.
BOOM! Now that you have messages coming in, this is where the fun starts.

_Skip the next section if you are all good._

**If You Don’t Have Messages**

1. Check to see that you made the proper entries in the rsyslog configuration file.
2. Check the syslog UDP configuration and make sure that is right - remember we changed the default port to 514.
3. Check to see if rsyslog messages are being forwarded to the port. You can use the `tcpdump` command to do this:
   ```
   $ sudo tcpdump -i lo host 127.0.0.1 and udp port 514
   ```
4. Check to see if the server is listening on the host:
   ```
   $ sudo netstat -peanut | grep ":514"
   ```

If you still have issues, connect to our community support or get in touch with us via the professional support offering.

**Create Your Dashboard**

You’ve got data coming in, let’s add information to a dashboard to better visualize the data we want to see.

**Add a Dashboard**

Now we’ll create a new dashboard for these messages by opening the _Dashboards_ page in the top menu and clicking on _Create dashboard_.

---

20 Chapter 2. Getting Started
Give your new dashboard a title and description.

In this given example the title is *Important Stuff* and the description *Look at this, yo.*

### Add a Dashboard Widget

Click the “Show received messages” button next to the `appliance-syslog-udp` input on the System > Inputs page to only show messages received by this input.

We’ll start by adding a widget with the message count: Click on *Add count to dashboard* and select the new created dashboard.
Now it will let you create a widget. In this case, we are creating a widget from our search result of message count in the last 8 hours. I like to put a timeframe in the title, and trends are always a big bowl of sunshine.
When you hit create you will see a small green notification at the bottom of the page. Click Dashboards and then the name of your dashboard.

Smile

And you’ll end up with the widget you created!
Extra Credit - One more

Let’s add a widget for root activity, because that sounds like it may actually be useful. We need to start with a search query for root. Click Search. Type root in the search and select your timeframe. Once the search results come in, click Add count to the dashboard.

Give your chart a title and hit Create.

The new widget is now on the screen. Good job - you’ve got this!
Go play around. If you want to know how to create more exciting charts and graphs, check out the section below.

Extra Credit - Graphs

Let’s start by searching for all messages within the last 1 hour. To do this, click Search, select Search in the last 1 hour, and leave the search bar blank. Once the search results populate, expand the messages field in the Search results sidebar and select Quick Values. Click Add to dashboard to add this entire pie chart and data table to your dashboard.

I like to track password changes, privilege assignments, root activity, system events, user logins, etc. Go knock yourself out and show your co-workers.

Once you have a few widgets in your dashboard, go into unlock/edit mode to quickly edit any widget, rearrange them on your dashboard, or delete. Make sure to click Lock to save!

You can find out why dashboards matter at our Dashboard documentation.
Get Alerted

I know, we’re all lazy, and busy. Nobody wants to just stare at a dashboard all day like it’s the World Cup. That’s for management.

Let’s configure some proactive alerts to let us know when something needs our attention.

Create a Stream

In order to set up an alert, we need to first create a stream. Streams process incoming messages in real time based on conditions that you set. Click Streams.

Let’s create a stream for all incoming security/authentication error messages. Click Create Stream.

Type in a Title and Description.
Create a Stream Rule

Next, we are going to configure the stream to process our Syslog UDP input messages for any security alerts.

Hit the Manage Rules button.

Pick the Syslog UDP Input, and click Add stream rule.
Then, type in the values shown below and hit save.

Then click I’m done!

We have just configured this stream to process in real time all the messages that come in from the security/authorization facility.
Now let’s create the alert.

**Create the Alert**

You can now either output your new stream to a 3rd party application or database, or trigger an alert to ping you in real time when a message that matches our stream rule comes in. Let’s create an alert that will email us when there are more than 2 messages in the last 2 minutes.

Click Alerts in the navigation bar and then **Manage conditions** on the Alerts overview page.

In the Condition section, select the “Security/Auth Errors from Syslogs” stream and the “Message Count Alert Condition” from the Condition type menu.

Configure the rest based on my screenshot (input 2’s in every field). Then click Add alert condition.
Create new Message Count Alert Condition

Message Count Alert Condition description
This condition is triggered when the number of messages is higher/lower than a defined threshold in a given time range.

Title
Some security messages

Time Range
2
Evaluate the condition for all messages received in the given number of minutes

Threshold Type
more than
Select condition to trigger alert: when there are more or less messages than the threshold

Threshold
2
Value which triggers an alert if crossed

Grace Period
2
Number of minutes to wait after an alert is resolved, to trigger another alert

Message Backlog
2
The number of messages to be included in alert notifications
Send a Test Email

On the Alerts overview page, click Manage notifications to setup an email notification.

Click Add new notification to create an email notification for the “Security/Auth Errors from Syslogs” stream.

Enter a title and your email address in the “E-Mail Receivers” section.
Create new Email Alert Callback

**Title**

Security Messages

**E-Mail Subject**

Graylog alert for stream: ${stream.title}: ${check_result.resultDescription}

The subject of sent out mail alerts

**Sender** (optional)

graylog@example.org

The sender of sent out mail alerts

**E-Mail Body** (optional)

```
###########
Alert Description: ${check_result.resultDescription}
Date: ${check_result.triggeredAt}
Stream ID: ${stream.id}
Stream title: ${stream.title}
Stream description: ${stream.description}
Alert Condition Title: ${alertCondition.title}
${if stream_url}Stream URL: ${stream_url}${end}

Triggered condition: ${check_result.triggeredCondition}
```

The template to generate the body from

**User Receivers** (optional)

Select User Receivers

Graylog usernames that should receive this alert

**E-Mail Receivers** (optional)

E-Mail addresses that should receive this alert
After adding the notification, hit the blue “Send test alert” button.

Going Further

If you want to configure an SMTP server, you can refer to *The graylog-ctl script*.

If you want to make this stream active, just go back to Streams and where you see the stream name, click the green *Start Stream* button.

You are done - go grab a Creamsicle, take a deep breath, and chillax. Tomorrow you can configure all your own logs and alerts. To help, go and get some deep knowledge in the official *documentation* friendly people that will help and guide you can be found in our *support community*.

You can find more information on our *Alerts* page.
Installing Graylog

Modern server architectures and configurations are managed in many different ways. Some people still put new software somewhere in `opt` manually for each server while others have already jumped on the configuration management train and fully automated reproducible setups.

Graylog can be installed in many different ways so you can pick whatever works best for you. We recommend to start with the virtual machine appliances for the fastest way to get started and then pick one of the other, more flexible installation methods to build an easier to scale setup.

This chapter is explaining the many ways to install Graylog and aims to help choosing the one that fits your needs.

Virtual Machine Appliances

Pre-Considerations

This is a showcase of Graylog and its cluster mode. Please run this appliance always in a separated network that is isolated from the internet. Read also the notes about production readiness!

Download

Download the OVA image. If you are unsure what the latest version number is, take a look at our release page.

Run the image

You can run the OVA in many systems like VMware or Virtualbox. In this example we will guide you through running the OVA in the free Virtualbox on OSX.

In Virtualbox select File -> Import appliance:
Hit *Continue* and keep the suggested settings on the next page as they are. Make sure that you have enough RAM and CPUs on your local machine. You can lower the resources the virtual machine will get assigned but we recommend to not lower it to ensure a good Graylog experience. In fact you might have to raise it if you plan to scale out later and send more messages into Graylog.

Press *Import* to finish loading the OVA into Virtualbox:
You can now start the VM and should see a login shell like this when the boot completed:
Note: If you don’t have a working DHCP server for your virtual machine, you will get the error message:

“Your appliance came up without a configured IP address. Graylog is probably not running correctly!”

In this case, you have to login and edit `/etc/network/interfaces` in order to setup a fixed IP address. Then manually reconfigure Graylog as shown in the following paragraphs.

Logging in

You can log into the shell of the operating system of the appliance with the user `ubuntu` and the password `ubuntu`. You should of course change those credentials.

The web interface is reachable on port 80 at the IP address of your virtual machine. The login prompt of the shell is showing you this IP address, too. (See screenshot above). DHCP should be enabled in your network otherwise take a look into the `graylog-ctl` command to apply a static IP address to the appliance.

The standard user for the web interface is `admin` with the password `admin`. 
Basic configuration

We are shipping the graylog-ctl tool with the virtual machine appliances to get you started with a customised setup as quickly as possible. Run these (optional) commands to configure the most basic settings of Graylog in the appliance:

```bash
sudo graylog-ctl set-email-config <smtp server> [--port=<smtp port> --user=<username> --password=<password>]
sudo graylog-ctl set-admin-password <password>
sudo graylog-ctl set-timezone <zone acronym>
sudo graylog-ctl reconfigure
```

The graylog-ctl has much more functionality documented. We strongly recommend to learn more about it to ensure smooth operation of your virtual appliance.

VMWare tools

If you are using the appliance on a VMWare host, you might want to install the hypervisor tools:

```bash
sudo apt-get install -y open-vm-tools
```

Update OVA to latest Version

You can update your Appliance to the newest release without deploying a new template.

Production readiness

You can use the Graylog appliances (OVA, Docker, AWS, ...) for small production setups but please consider to harden the security of the box before.

- Set another password for the default ubuntu user
- Disable remote password logins in /etc/ssh/sshd_config and deploy proper ssh keys
- Separate the box network-wise from the outside, otherwise Elasticsearch and MongoDB can be reached by anyone
- add additional RAM to the appliance and raise the java heap!
- add additional HDD to the appliance and extend disk space.
- add the appliance to your monitoring and metric systems.

If you want to create your own customised setup take a look at our other installation methods.

Operating System Packages

Until configuration management systems made their way into broader markets and many datacenters, one of the most common ways to install software on Linux servers was to use operating system packages. Debian has DEB, Red Hat has RPM and many other distributions are based on those or come with own package formats. Online repositories of software packages and corresponding package managers make installing and configuring new software a matter of a single command and a few minutes of time.

Graylog offers official DEB and RPM package repositories. The packages have been tested on the following operating systems:
The repositories can be setup by installing a single package. Once that’s done the Graylog packages can be installed via apt-get or yum. The packages can also be downloaded with a web browser at https://packages.graylog2.org/ if needed.

**Prerequisites**

Make sure to install and configure the following software before installing and starting any Graylog services:

- **Java** (>= 8)
- **MongoDB** (>= 2.4)
- **Elasticsearch** (>= 2.0)

**Caution:** Graylog 2.x does not work with Elasticsearch 5.x!

**DEB / APT**

Download and install `graylog-2.2-repository_latest.deb` via `dpkg(1)` and also make sure that the `apt-transport-https` package is installed:

```
$ sudo apt-get install apt-transport-https
$ wget https://packages.graylog2.org/repo/packages/graylog-2.2-repository_latest.deb
$ sudo dpkg -i graylog-2.2-repository_latest.deb
$ sudo apt-get update
$ sudo apt-get install graylog-server
```

After the installation completed successfully, Graylog can be started with the following commands. Make sure to use the correct command for your operating system.

<table>
<thead>
<tr>
<th>OS</th>
<th>Init System</th>
<th>Command</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ubuntu 14.04, 12.04</td>
<td>upstart</td>
<td><code>sudo start graylog-server</code></td>
</tr>
<tr>
<td>Debian 7</td>
<td>SysV</td>
<td><code>sudo service graylog-server start</code></td>
</tr>
<tr>
<td>Debian 8, Ubuntu 16.04</td>
<td>systemd</td>
<td><code>sudo systemctl start graylog-server</code></td>
</tr>
</tbody>
</table>

The packages are configured to **not** start any Graylog services during boot. You can use the following commands to start Graylog when the operating system is booting.

<table>
<thead>
<tr>
<th>OS</th>
<th>Init System</th>
<th>Command</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ubuntu 14.04, 12.04</td>
<td>upstart</td>
<td><code>sudo rm -f /etc/init/graylog-server.override</code></td>
</tr>
<tr>
<td>Debian 7</td>
<td>SysV</td>
<td><code>sudo update-rc.d graylog-server defaults 95 10</code></td>
</tr>
<tr>
<td>Debian 8, Ubuntu 16.06</td>
<td>systemd</td>
<td><code>sudo systemctl enable graylog-server</code></td>
</tr>
</tbody>
</table>

**Update to latest version**

If you’ve been using the repository package to install Graylog before, it has to be updated first. The new package will replace the repository URL, without which you will only be able to get bugfix releases of your previously installed version of Graylog.

The update basically works like a fresh installation:
Manual Repository Installation

If you don’t like to install the repository DEB to get the repository configuration onto your system, you can do so manually (although we don’t recommend to do that).

First, add the Graylog GPG keyring which is being used to sign the packages to your system.

**Hint:** We assume that you have placed the GPG key into /etc/apt/trusted.gpg.d/.

Now create a file /etc/apt/sources.list.d/graylog.list with the following content:

```
deb https://packages.graylog2.org/repo/debian/ stable 2.2
```

RPM / YUM / DNF

Download and install graylog-2.2-repository_latest.rpm via `rpm(8)`:

```
$ sudo rpm -Uvh https://packages.graylog2.org/repo/packages/graylog-2.2-repository_latest.rpm
$ sudo yum install graylog-server
```

After the installation completed successfully, Graylog can be started with the following commands. Make sure to use the correct command for your operating system.

<table>
<thead>
<tr>
<th>OS</th>
<th>Init System</th>
<th>Command</th>
</tr>
</thead>
<tbody>
<tr>
<td>CentOS 6</td>
<td>SysV</td>
<td><code>sudo service graylog-server start</code></td>
</tr>
<tr>
<td>CentOS 7</td>
<td>systemd</td>
<td><code>sudo systemctl start graylog-server</code></td>
</tr>
</tbody>
</table>

The packages are configured to **not** start any Graylog services during boot. You can use the following commands to start Graylog when the operating system is booting.

<table>
<thead>
<tr>
<th>OS</th>
<th>Init System</th>
<th>Command</th>
</tr>
</thead>
<tbody>
<tr>
<td>CentOS 6</td>
<td>SysV</td>
<td><code>sudo update-rc.d graylog-server defaults 95 10</code></td>
</tr>
<tr>
<td>CentOS 7</td>
<td>systemd</td>
<td><code>sudo systemctl enable graylog-server</code></td>
</tr>
</tbody>
</table>

Update to latest version

If you’ve been using the repository package to install Graylog before, it has to be updated first. The new package will replace the repository URL, without which you will only be able to get bugfix releases of your previously installed version of Graylog.

The update basically works like a fresh installation:

```
$ sudo rpm -Uvh https://packages.graylog2.org/repo/packages/graylog-2.2-repository_latest.rpm
$ sudo yum clean all
$ sudo yum install graylog-server
```

Running `yum clean all` is required because YUM might use a stale cache and thus might be unable to find the latest version of the `graylog-server` package.

---

3.2. Operating System Packages
Manual Repository Installation

If you don’t like to install the repository RPM to get the repository configuration onto your system, you can do so manually (although we don’t recommend to do that).

First, add the Graylog GPG key which is being used to sign the packages to your system.

**Hint:** We assume that you have placed the GPG key into `/etc/pki/rpm-gpg/RPM-GPG-KEY-graylog`.

Now create a file named `/etc/yum.repos.d/graylog.repo` with the following content:

```plaintext
[graylog]
name=graylog
baseurl=https://packages.graylog2.org/repo/el/stable/2.2/$basearch/
gpgcheck=1
repo_gpgcheck=0
gpgkey=file:///etc/pki/rpm-gpg/RPM-GPG-KEY-graylog
```

Step-by-step guides

Ubuntu installation

This guide describes the fastest way to install Graylog on Ubuntu 16.04 LTS. All links and packages are present at the time of writing but might need to be updated later on.

**Warning:** This setup should not be done on publicly exposed servers. This guide does not cover security settings!

Prerequisites

Taking a minimal server setup as base will need this additional packages:

```
$ sudo apt-get update && sudo apt-get upgrade
$ sudo apt-get install apt-transport-https openjdk-8-jre-headless uuid-runtime pwgen
```

MongoDB

The Version included in Ubuntu 16.04 LTS can be used together with Graylog 2.0.0 and higher:

```
$ sudo apt-get install mongodb-server
```

Elasticsearch

Graylog 2.0.0 and higher requires Elasticsearch 2.x, so we took the installation instructions from the Elasticsearch installation guide:

```
$ wget -qO- https://packages.elastic.co/GPG-KEY-elasticsearch | sudo apt-key add -
$ echo "deb https://packages.elastic.co/elasticsearch/2.x/debian stable main" | sudo tee -a /etc/apt/sources.list.d/elasticsearch-2.x.list
$ sudo apt-get update && sudo apt-get install elasticsearch
```
Make sure to modify the Elasticsearch configuration file (/etc/elasticsearch/elasticsearch.yml) and set the cluster name to graylog additionally you need to uncomment (remove the # as first character) the line:

```
cluster.name: graylog
```

After you have modified the configuration, you can start Elasticsearch:

```
$ sudo systemctl daemon-reload
$ sudo systemctl enable elasticsearch.service
$ sudo systemctl restart elasticsearch.service
```

**Graylog**

Now install the Graylog repository configuration and Graylog itself with the following commands:

```
$ wget https://packages.graylog2.org/repo/packages/graylog-2.2-repository_latest.deb
$ sudo dpkg -i graylog-2.2-repository_latest.deb
$ sudo apt-get update && sudo apt-get install graylog-server
```

Follow the instructions in your /etc/graylog/server/server.conf and add `password_secret` and `root_password_sha2`. These settings are mandatory and without them, Graylog will not start!

You need to use the following command to create your `root_password_sha2`:

```
echo -n yourpassword | sha256sum
```

To be able to connect to Graylog you should set `rest_listen_uri` and `web_listen_uri` to the public host name or a public IP address of the machine you can connect to. More information about these settings can be found in Configuring the web interface.

**Note:** If you’re operating a single-node setup and would like to use HTTPS for the Graylog web interface and the Graylog REST API, it’s possible to use NGINX or Apache as a reverse proxy.

The last step is to enable Graylog during the operating system’s startup:

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

The next step is to **ingest messages** into your Graylog and extract the messages with **extractors** or use the **Pipelines** to work with the messages.

**Multiple Server Setup**

If you plan to have multiple server taking care of different roles in your cluster **like we have in this big production setup** you need to modify only a few settings. This is covered in our **Multi-node Setup guide**. The **default file location guide** will give you the file you need to modify in your setup.

**Feedback**

Please file a bug report in the GitHub repository for the operating system packages if you run into any packaging related issues.
Debian installation

This guide describes the fastest way to install Graylog on Debian Linux 8 (Jessie). All links and packages are present at the time of writing but might need to be updated later on.

**Warning:** This setup should not be done on publicly exposed servers. This guide **does not cover** security settings!

### Prerequisites

Not all required dependencies are available in the standard repository, so we need to add Debian Backports to the list of package sources:

```bash
$ sudo echo "deb http://ftp.debian.org/debian jessie-backports main" > /etc/apt/sources.list.d/backports.list
$ sudo apt-get update && sudo apt-get upgrade
```

If you’re starting from a minimal server setup, you will need to install these additional packages:

```bash
$ sudo apt-get install apt-transport-https openjdk-8-jre-headless uuid-runtime pwgen
```

**MongoDB**

The version of MongoDB included in Debian Jessie is recent enough to be used with Graylog 2.0.0 and higher:

```bash
$ sudo apt-get install mongodb-server
```

**Elasticsearch**

Graylog 2.0.0 and higher requires Elasticsearch 2.x, so we took the installation instructions from the Elasticsearch installation guide:

```bash
$ wget -qO - https://packages.elastic.co/GPG-KEY-elasticsearch | sudo apt-key add -
$ echo "deb https://packages.elastic.co/elasticsearch/2.x/debian stable main" | sudo tee -a /etc/apt/sources.list.d/elasticsearch-2.x.list
$ sudo apt-get update && sudo apt-get install elasticsearch
```

Make sure to modify the Elasticsearch configuration file (`/etc/elasticsearch/elasticsearch.yml`) and set the cluster name to `graylog` additionally you need to uncomment (remove the # as first character) the line:

```
cluster.name: graylog
```

After you have modified the configuration, you can start Elasticsearch:

```bash
$ sudo systemctl daemon-reload
$ sudo systemctl enable elasticsearch.service
$ sudo systemctl restart elasticsearch.service
```
Graylog

Now install the Graylog repository configuration and Graylog itself with the following commands:

```
$ wget https://packages.graylog2.org/repo/packages/graylog-2.2-repository_latest.deb
$ sudo dpkg -i graylog-2.2-repository_latest.deb
$ sudo apt-get update && sudo apt-get install graylog-server
```

Follow the instructions in your `/etc/graylog/server/server.conf` and add `password_secret` and `root_password_sha2`. These settings are mandatory and without them, Graylog will not start!

You need to use the following command to create your `root_password_sha2`:

```
echo -n yourpassword | sha256sum
```

To be able to connect to Graylog you should set `rest_listen_uri` and `web_listen_uri` to the public host name or a public IP address of the machine you can connect to. More information about these settings can be found in Configuring the web interface.

**Note:** If you’re operating a single-node setup and would like to use HTTPS for the Graylog web interface and the Graylog REST API, it’s possible to use NGINX or Apache as a reverse proxy.

The last step is to enable Graylog during the operating system’s startup:

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

The next step is to ingest messages into your Graylog and extract the messages with extractors or use the Pipelines to work with the messages.

**Multiple Server Setup**

If you plan to have multiple server taking care of different roles in your cluster like we have in this big production setup you need to modify only a few settings. This is covered in our Multi-node Setup guide. The default file location guide will give you the file you need to modify in your setup.

**Feedback**

Please file a bug report in the GitHub repository for the operating system packages if you run into any packaging related issues.

If you found this documentation confusing or have more questions, please open an issue in the Github repository for the documentation.

**CentOS installation**

This guide describes the fastest way to install Graylog on CentOS 7. All links and packages are present at the time of writing but might need to be updated later on.

**Warning:** This setup should not be done on publicly exposed servers. This guide does not cover security settings!
Prerequisites

Taking a minimal server setup as base will need this additional packages:

```
$ sudo yum install java-1.8.0-openjdk-headless.x86_64
```

If you want to use `pwgen` later on you need to Setup EPEL on your system with `sudo yum install epel-release` and install the package with `sudo yum install pwgen`.

MongoDB

Installing MongoDB on CentOS should follow the tutorial for RHEL and CentOS from the MongoDB documentation. First add the repository file `/etc/yum.repos.d/mongodb-org-3.2.repo` with the following contents:

```
[mongodb-org-3.2]
name=MongoDB Repository
baseurl=https://repo.mongodb.org/yum/redhat/$releasever/mongodb-org/3.2/x86_64/
gpgcheck=1
enabled=1
gpgkey=https://www.mongodb.org/static/pgp/server-3.2.asc
```

After that, install the latest release of MongoDB with `sudo yum install mongodb-org`.

Additionally, run these last steps to start MongoDB during the operating system’s boot and start it right away:

```
$ sudo chkconfig --add mongod
$ sudo systemctl daemon-reload
$ sudo systemctl enable mongod.service
$ sudo systemctl start mongod.service
```

Elasticsearch

Graylog 2.0.0 and higher requires Elasticsearch 2.x, so we took the installation instructions from the Elasticsearch installation guide.

First install the Elastic GPG key with `rpm --import https://packages.elastic.co/GPG-KEY-elasticsearch` then add the repository file `/etc/yum.repos.d/elasticsearch.repo` with the following contents:

```
[elasticsearch-2.x]
name=Elasticsearch repository for 2.x packages
baseurl=https://packages.elastic.co/elasticsearch/2.x/centos
gpgcheck=1
gpgkey=https://packages.elastic.co/GPG-KEY-elasticsearch
enabled=1
```

followed by the installation of the latest release with `sudo yum install elasticsearch`.

Make sure to modify the Elasticsearch configuration file (`/etc/elasticsearch/elasticsearch.yml`) and set the cluster name to `graylog` additionally you need to uncomment (remove the `#` as first character) the line:

```
cluster.name: graylog
```

After you have modified the configuration, you can start Elasticsearch:

```
$ sudo chkconfig --add elasticsearch
$ sudo systemctl daemon-reload
$ sudo systemctl enable elasticsearch.service
$ sudo systemctl restart elasticsearch.service
```
Graylog

Now install the Graylog repository configuration and Graylog itself with the following commands:

$ sudo rpm -Uvh https://packages.graylog2.org/repo/packages/graylog-2.2-repository_latest.rpm
$ sudo yum install graylog-server

Follow the instructions in your /etc/graylog/server/server.conf and add password_secret and root_password_sha2. These settings are mandatory and without them, Graylog will not start!

You need to use the following command to create your root_password_sha2:

```
echo -n yourpassword | sha256sum
```

To be able to connect to Graylog you should set rest_listen_uri and web_listen_uri to the public host name or a public IP address of the machine you can connect to. More information about these settings can be found in Configuring the web interface.

**Note:** If you’re operating a single-node setup and would like to use HTTPS for the Graylog web interface and the Graylog REST API, it’s possible to use NGINX or Apache as a reverse proxy.

The last step is to enable Graylog during the operating system’s startup:

```
$ sudo chkconfig --add graylog-server
$ sudo systemctl daemon-reload
$ sudo systemctl enable graylog-server.service
$ sudo systemctl start graylog-server.service
```

The next step is to ingest messages into your Graylog and extract the messages with extractors or use the Pipelines to work with the messages.

**SELinux information**

**Hint:** We assume that you have policycoreutils-python installed to manage SELinux.

If you’re using SELinux on your system, you need to take care of the following settings:

- Allow the web server to access the network: `sudo setsebool -P httpd_can_network_connect 1`

- If the policy above does not comply with your security policy, you can also allow access to each port individually:
  
  - Graylog REST API and web interface: `sudo semanage port -a -t http_port_t -p tcp 9000`
  
  - Elasticsearch (only if the HTTP API is being used): `sudo semanage port -a -t http_port_t -p tcp 9200`

- Allow using MongoDB’s default port (27017/tcp): `sudo semanage port -a -t mongod_port_t -p tcp 27017`

If you run a single server environment with NGINX or Apache proxy, enabling the Graylog REST API is enough. All other rules are only required in a multi-node setup.
Hint: Depending on your actual setup and configuration, you might need to add more SELinux rules to get to a running setup.

Further reading

- https://www.nginx.com/blog/nginx-se-linux-changes-upgrading-rhel-6-6/
- https://wiki.centos.org/HowTos/SELinux
- https://wiki.centos.org/TipsAndTricks/SelinuxBooleans
- https://www.digitalocean.com/community/tutorials/an-introduction-to-selinux-on-centos-7-part-1-basic-concepts

Multiple Server Setup

If you plan to have multiple servers taking care of different roles in your cluster like we have in this big production setup you need to modify only a few settings. This is covered in our Multi-node Setup guide. The default file location guide will give you the file you need to modify in your setup.

Feedback

Please file a bug report in the GitHub repository for the operating system packages if you run into any packaging related issues.

If you found this documentation confusing or have more questions, please open an issue in the Github repository for the documentation.

Chef, Puppet, Ansible

The DevOps movement turbocharged market adoption of the newest generation of configuration management and orchestration tools like Chef, Puppet or Ansible. Graylog offers official scripts for all three of them:

- https://supermarket.chef.io/cookbooks/graylog2
- https://forge.puppet.com/graylog/graylog
- https://galaxy.ansible.com/Graylog2/graylog-ansible-role
Docker

Requirements

You need a recent `docker` version installed.

This will create three containers with all Graylog services running:

```bash
$ docker run --name some-mongo -d mongo:3
$ docker run --name some-elasticsearch -d elasticsearch:2 elasticsearch -Des.cluster.name="graylog"
```

Testing a beta version

You can also run a pre-release (alpha, beta, or release candidate) version of Graylog using Docker. The pre-releases are included in the `graylog2/server` image. Follow this guide and pick an alpha/beta/rc tag like:

```bash
```

We only recommend running pre-release versions if you are an experienced Graylog user and know what you are doing.

Settings

Graylog comes with a default configuration that works out of the box but you have to set a password for the admin user. Also the web interface needs to know how to connect from your browser to the Graylog API. Both can be done via environment variables:

```bash
-e GRAYLOG_PASSWORD_SECRET=somepasswordpepper
-e GRAYLOG_ROOT_PASSWORD_SHA2=8c6976e5b5410415bde908bd4dee15dfb167a9c873fc4bb8a81f6f2ab48a918
-e GRAYLOG_WEB_ENDPOINT_URI="http://127.0.0.1:9000/api"
```

In this case you can login to Graylog with the username and password `admin`. Generate your own password with this command:

```bash
$ echo -n yourpassword | shasum -a 256
```

This all can be put in a `docker-compose.yml` file, like:

```yaml
version: '2'
services:
  mongo:
    image: "mongo:3"
  elasticsearch:
    image: "elasticsearch:2"
    command: "elasticsearch -Des.cluster.name='graylog'"
  graylog:
    image: graylog2/server:2.2.1-1
    environment:
      GRAYLOG_PASSWORD_SECRET: somepasswordpepper
      GRAYLOG_ROOT_PASSWORD_SHA2: 8c6976e5b5410415bde908bd4dee15dfb167a9c873fc4bb8a81f6f2ab48a918
      GRAYLOG_WEB_ENDPOINT_URI: "http://127.0.0.1:9000/api"
    depends_on:
      - mongo
      - elasticsearch
```

3.4. Docker
After starting the three containers with `docker-compose up` open your browser with the URL `http://127.0.0.1:9000` and login with `admin:admin`.

**How to get log data in**

You can create different kinds of inputs under *System -> Inputs*, however you can only use ports that have been properly mapped to your docker container, otherwise data will not go through.

E.g. to start a raw TCP input on port 5555, stop your container and recreate it, whilst appending `-p 5555:5555` to your `run` argument. Similarly, the same can be done for UDP by appending `-p 5555:5555/udp` option. Then you can send raw text to Graylog like `echo 'first log message' | nc localhost 5555`.

**Persist log data**

In order to make the log data and configuration of Graylog persistent, you can use external volumes to store all data. In case of a container restart simply re-use the existing data from the former instances. Create the configuration directory and copy the default files:

```
mkdir /graylog/config
cd /graylog/config
wget https://raw.githubusercontent.com/Graylog2/graylog2-images/2.2/docker/config/graylog.conf
wget https://raw.githubusercontent.com/Graylog2/graylog2-images/2.2/docker/config/log4j2.xml
```

The `docker-compose.yml` file looks like this:

```
version: '2'
services:
  mongo:
    image: "mongo:3"
    volumes:
      - /graylog/data/mongo/:/data/db
  elasticsearch:
    image: "elasticsearch:2"
    command: "elasticsearch -Des.cluster.name='graylog'"
    volumes:
      - /graylog/data/elasticsearch:/usr/share/elasticsearch/data
  graylog:
    image: graylog2/server:2.2.1-1
    volumes:
      - /graylog/data/journal:/usr/share/graylog/data/journal
      - /graylog/config:/usr/share/graylog/data/config
    environment:
      GRAYLOG_PASSWORD_SECRET: somepasswordpepper
      GRAYLOG_ROOT_PASSWORD_SHA2: 8c6976e5b5410415bde908bd4dee15dfb167a9c873fc4bb8a81f6d2ab448a918
      GRAYLOG_WEB_ENDPOINT_URI: http://127.0.0.1:9000/api/
    depends_on:
      - mongo
      - elasticsearch
    ports:
      - "9000:9000"
      - "12201/udp:12201/udp"
      - "1514/udp:1514/udp"
```
Start all services with exposed data directories:

```bash
$ docker-compose up
```

---

**Configuration**

Every configuration option can be set via environment variables. Simply prefix the parameter name with `GRAYLOG_` and put it all in upper case.

For example setting up a smtp configuration, the `docker-compose.yml` would look like this:

```yaml
environment:
  GRAYLOG_TRANSPORT_EMAIL_ENABLED: "true"
  GRAYLOG_TRANSPORT_EMAIL_HOSTNAME: smtp
  GRAYLOG_TRANSPORT_EMAIL_PORT: 25
  GRAYLOG_TRANSPORT_EMAIL_USE_AUTH: "false"
  GRAYLOG_TRANSPORT_EMAIL_USE_TLS: "false"
  GRAYLOG_TRANSPORT_EMAIL_USE_SSL: "false"
```

Another option would be to store the configuration file outside of the container and edit it directly.

---

**Plugins**

In order to add plugins you can build a new image based on the existing `graylog2/server` image with the needed plugin included. Simply create a new Dockerfile in an empty directory:

```bash
FROM graylog2/server:2.2.1-1
RUN wget -O /usr/share/graylog/plugin/graylog-plugin-beats-1.1.0.jar https://github.com/Graylog2/graylog-plugin-beats/releases/download/1.1.0/graylog-plugin-beats-1.1.0.jar
```

Build a new image from that:

```bash
$ docker build -t graylog-with-beats-plugin .
```

In this example we created a new image with the Beats plugin installed. From now on reference to that image instead of the `graylog2/server` e.g. in a `docker-compose.yml` file:

```yaml
version: '2'
services:
  mongo:
    image: "mongo:3"
    volumes:
      - /graylog/data/mongo:/data/db
  elasticsearch:
    image: "elasticsearch:2"
    command: "elasticsearch -Des.cluster.name='graylog'"
    volumes:
      - /graylog/data/elasticsearch:/usr/share/elasticsearch/data
  graylog:
    image: graylog-with-beats-plugin
    ...  
```

---

**Problems**

- In case you see warnings regarding open file limit, try to set ulimit from the outside of the container:
$ docker run --ulimit nofile=64000:64000 ...

• The devicemapper storage driver can produce problems with Graylogs disk journal on some systems. In this case please pick another driver like aufs or overlay.

Build

To build the image from scratch run:

$ docker build --build-arg GRAYLOG_VERSION=${GRAYLOG_VERSION} -t graylog2/server .

Production readiness

You can use the Graylog appliances (OVA, Docker, AWS, ...) for small production setups but please consider to harden the security of the box before.

• Set another password for the default ubuntu user
• Disable remote password logins in /etc/ssh/sshd_config and deploy proper ssh keys
• Separate the box network-wise from the outside, otherwise Elasticsearch and MongoDB can be reached by anyone
• add additional RAM to the appliance and raise the java heap!
• add additional HDD to the appliance and extend disk space.
• add the appliance to your monitoring and metric systems.

If you want to create your own customised setup take a look at our other installation methods.

Vagrant

Requirements

You need a recent vagrant version.

Installation

These steps will create a Vagrant virtual machine with all Graylog services running:

$ wget https://raw.githubusercontent.com/Graylog2/graylog2-images/2.2/vagrant/Vagrantfile
$ vagrant up

Usage

After starting the virtual machine, your Graylog instance is ready to use. You can reach the web interface by pointing your browser to: http://localhost:8080

The default login is Username: admin, Password: admin.
Configuration

We are shipping the graylog-ctl tool with the virtual machine appliances to get you started with a customised setup as quickly as possible. Run these (optional) commands to configure the most basic settings of Graylog in the appliance:

```
sudo graylog-ctl set-email-config <smtp server> [--port=<smtp port> --user=<username> --password=<password>]
sudo graylog-ctl set-admin-password <password>
sudo graylog-ctl set-timezone <zone acronym>
sudo graylog-ctl reconfigure
```

The graylog-ctl has much more functionality documented. We strongly recommend to learn more about it to ensure smooth operation of your virtual appliance.

If you want to create your own customised setup take a look at our other installation methods.

OpenStack

Installation

Download the Graylog image from the package site, uncompress it and import it into the OpenStack image store:

```
$ wget https://packages.graylog2.org/releases/graylog-omnibus/qcow2/graylog-2.2.3-1.qcow2.gz
$ gunzip graylog-2.2.3-1.qcow2.gz
$ glance image-create --name='graylog' --is-public=true --container-format=bare --disk-format=qcow2 --file graylog-2.2.3-1.qcow2
```

You should now see an image called graylog in the OpenStack web interface under Images

Usage

Launch a new instance of the image, make sure to reserve at least 4GB ram for the instance. After spinning up, login with the username ubuntu and your selected ssh key. Run the reconfigure program in order to setup Graylog and start all services:

```
$ ssh ubuntu@<vm IP>
$ sudo graylog-ctl reconfigure
```

Open http://<vm ip> in your browser to access the Graylog web interface. Default username and password is admin.

Basic configuration

We are shipping the graylog-ctl tool with the virtual machine appliances to get you started with a customised setup as quickly as possible. Run these (optional) commands to configure the most basic settings of Graylog in the appliance:

```
sudo graylog-ctl set-email-config <smtp server> [--port=<smtp port> --user=<username> --password=<password>]
sudo graylog-ctl set-admin-password <password>
sudo graylog-ctl set-timezone <zone acronym>
sudo graylog-ctl reconfigure
```

The graylog-ctl has much more functionality documented. We strongly recommend to learn more about it to ensure smooth operation of your virtual appliance.
Production readiness

You can use the Graylog appliances (OVA, Docker, AWS, ...) for small production setups but please consider to harden the security of the box before.

- Set another password for the default ubuntu user
- Disable remote password logins in /etc/ssh/sshd_config and deploy proper ssh keys
- Separate the box network-wise from the outside, otherwise Elasticsearch and MongoDB can be reached by anyone
- add additional RAM to the appliance and raise the java heap!
- add additional HDD to the appliance and extend disk space.
- add the appliance to your monitoring and metric systems.

If you want to create your own customised setup take a look at our other installation methods.

Amazon Web Services

AMIs

Select your AMI and AWS Region.

Usage

- Click on Launch instance for your AWS region to start Graylog into.
- Choose an instance type with at least 4GB memory.
- Finish the wizard and spin up the VM.
- Login to the instance via SSH as user ubuntu.
- Run sudo graylog-ctl reconfigure.
- Open port 80 and 9000 in the applied security group to access the web interface.
- additionally open more ports for ingesting log data, like 514 for syslog or 12201 for the GELF protocol.

Open http://<private ip> in your browser to access the Graylog web interface. Default username and password is admin.

Networking

Your browser needs access to port 80 or 443 for reaching the web interface. The interface itself creates a connection back to the REST API of the Graylog server on port 9000. As long as you are in a private network like Amazon VPC for instance, this works out of the box. If you want to use the public IP address of your VM, this mechanism doesn’t work automatically anymore. You have to tell Graylog how to reach the API from the users browser perspective:

```
sudo graylog-ctl reconfigure
```

Also make sure that this port is open, even on the public IP.
HTTPS

In order to enable HTTPS for the web interface both ports need to be encrypted. Otherwise the web browser would show an error message. For this reason we created a proxy configuration on the appliance that can be enabled by running:

```
sudo graylog-ctl enforce-ssl
sudo graylog-ctl reconfigure
```

This command combines the Graylog web interface and the API on port 443. The API is accessible via the path `/api`. For this reason you have to set the external IP to an HTTPS address with the appended path `/api`:

```
sudo graylog-ctl reconfigure
```

Basic configuration

We are shipping the `graylog-ctl` tool with the virtual machine appliances to get you started with a customised setup as quickly as possible. Run these (optional) commands to configure the most basic settings of Graylog in the appliance:

```
sudo graylog-ctl set-email-config <smtp server> [--port=<smtp port> --user=<username> --password=<password>]
sudo graylog-ctl set-admin-password <password>
sudo graylog-ctl set-timezone <zone acronym>
sudo graylog-ctl reconfigure
```

The `graylog-ctl` has much more functionality documented. We strongly recommend to learn more about it to ensure smooth operation of your virtual appliance.

Production readiness

You can use the Graylog appliances (OVA, Docker, AWS, ...) for small production setups but please consider to harden the security of the box before.

- Set another password for the default ubuntu user
- Disable remote password logins in `/etc/ssh/sshd_config` and deploy proper ssh keys
- Separate the box network-wise from the outside, otherwise Elasticsearch and MongoDB can be reached by anyone
- add additional RAM to the appliance and raise the `java heap`
- add additional HDD to the appliance and `extend disk space`
- add the appliance to your monitoring and metric systems.

If you want to create your own customised setup take a look at our other installation methods.

Microsoft Windows

Unfortunately there is no supported way to run Graylog on Microsoft Windows operating systems even though all parts run on the Java Virtual Machine. We recommend to run the virtual machine appliances on a Windows host. It should be technically possible to run Graylog on Windows but it is most probably not worth the time to work your way around the cliffs.
Should you require running Graylog on Windows, you need to disable the message journal in `graylog-server` by changing the following setting in the `graylog.conf`:

```plaintext
message_journal_enabled = false
```

Due to restrictions of how Windows handles file locking, the journal will not work correctly. **Please note that this impacts Graylog’s ability to buffer messages, so we strongly recommend running the Linux-based OVsAs on Windows.**

### Manual Setup

#### Graylog server on Linux

**Prerequisites**

Graylog depends on MongoDB and Elasticsearch to operate, please refer to the system requirements for details.

**Downloading and extracting the server**

Download the tar archive from the download pages and extract it on your system:

```bash
~$ tar xvfz graylog-VERSION.tgz
~$ cd graylog-VERSION
```

**Configuration**

Now copy the example configuration file:

```bash
~# cp graylog.conf.example /etc/graylog/server/server.conf
```

You can leave most variables as they are for a first start. All of them should be well documented.

Configure at least the following variables in `/etc/graylog/server/server.conf`:

- `is_master = true`
  - Set only one `graylog-server` node as the master. This node will perform periodical and maintenance actions that slave nodes won’t. Every slave node will accept messages just as the master nodes. Nodes will fall back to slave mode if there already is a master in the cluster.

- `password_secret`
  - You must set a secret that is used for password encryption and salting here. The server will refuse to start if it’s not set. Generate a secret with for example `pwgen -N 1 -s 96`. If you run multiple `graylog-server` nodes, make sure you use the same `password_secret` for all of them!

- `root_password_sha2`
  - A SHA2 hash of a password you will use for your initial login. Set this to a SHA2 hash generated with `echo -n yourpassword | shasum -a 256` and you will be able to log in to the web interface with username `admin` and password `yourpassword`.

- `elasticsearch_shards = 4`
  - The number of shards for your indices. A good setting here highly depends on the number of nodes in your Elasticsearch cluster. If you have one node, set it to 1.
• **elasticsearch_replicas = 0**
  
  – The number of replicas for your indices. A good setting here highly depends on the number of nodes in your Elasticsearch cluster. If you have one node, set it to 0.

• **mongodb_uri**
  
  – Enter your MongoDB connection and authentication information here.

### Starting the server

You need to have Java installed. Running the OpenJDK is totally fine and should be available on all platforms. For example on Debian it is:

```
~$ apt-get install openjdk-8-jre
```

Start the server:

```
~$ cd bin/
~$ ./graylogctl start
```

The server will try to write a node_id to the `graylog-server-node-id` file. It won’t start if it can’t write there because of for example missing permissions.

See the startup parameters description below to learn more about available startup parameters. Note that you might have to be root to bind to the popular port 514 for syslog inputs.

You should see a line like this in the debug output of Graylog successfully connected to your Elasticsearch cluster:

```
```

You can find the logs of Graylog in the directory `logs/`.

**Important:** All systems running Graylog must have synchronised system time. We strongly recommend to use NTP or similar mechanisms on all machines of your Graylog infrastructure.

### Supplying external logging configuration

Graylog is using Apache Log4j 2 for its internal logging and ships with a default log configuration file which is embedded within the shipped JAR.

In case you need to modify Graylog’s logging configuration, you can supply a Java system property specifying the path to the configuration file in your start script (e.g. `graylogctl`).

Append this before the `-jar` parameter:

```
-Dlog4j.configurationFile=file:///path/to/log4j2.xml
```

Substitute the actual path to the file for the `/path/to/log4j2.xml` in the example.

In case you do not have a log rotation system already in place, you can also configure Graylog to rotate logs based on their size to prevent the log files to grow without bounds using the `RollingFileAppender`.

One such example `log4j2.xml` configuration is shown below:

```xml
<?xml version="1.0" encoding="UTF-8"?>
<Configuration packages="org.graylog2.log4j" shutdownHook="disable">
  <Appenders>
    <RollingFile name="RollingFile" fileName="/tmp/logs/graylog.log" filePattern="/tmp/logs/graylog-%d{yyyy-MM-dd}.log.gz">
      <PatternLayout>
```

3.9. Manual Setup
Command line (CLI) parameters

There are a number of CLI parameters you can pass to the call in your graylogctl script:

- `-h, --help`: Show help message
- `-f CONFIGFILE, --configfile CONFIGFILE`: Use configuration file CONFIGFILE for Graylog; default: /etc/graylog/server/server.conf
- `-d, --debug`: Run in debug mode
- `-l, --local`: Run in local mode. Automatically invoked if in debug mode. Will not send system statistics, even if enabled and allowed. Only interesting for development and testing purposes.
- `-p PIDFILE, --pidfile PIDFILE`: Set the file containing the PID of graylog to PIDFILE; default: /tmp/graylog.pid
- `-np, --no-pid-file`: Do not write PID file (overrides -p/--pidfile)
- `--version`: Show version of Graylog and exit
Problems with IPv6 vs. IPv4?

If your Graylog node refuses to listen on IPv4 addresses and always chooses for example a `rest_listen_address` like `:::9000` you can tell the JVM to prefer the IPv4 stack.

Add the `java.net.preferIPv4Stack` flag in your `graylogctl` script or from wherever you are calling the `graylog.jar`:

```
~$ sudo -u graylog java -Djava.net.preferIPv4Stack=true -jar graylog.jar
```

Create a message input and send a first message

Log in to the web interface on port 9000 (e.g. `http://127.0.0.1:9000`) and navigate to `System -> Inputs`.

Launch a new Raw/Plaintext UDP input, listening on `127.0.0.1` on port `9099`. There’s no need to configure anything else for now.

The list of running inputs on that node should show you your new input right away.

Let’s send a message in:

```
echo "Hello Graylog, let's be friends." | nc -w 1 -u 127.0.0.1 9099
```

This has sent a short string to the raw UDP input you just opened. Now search for `friends` using the search bar on the top and you should already see the message you just sent in. Click on it in the table and see it in detail:
You have just sent your first message to Graylog! Why not spawn a syslog input and point some of your servers to it? You could also create some user accounts for your colleagues.

**System requirements**

The Graylog server application has the following prerequisites:

- Some modern Linux distribution (Debian Linux, Ubuntu Linux, or CentOS recommended)
- Elasticsearch 2.x (2.1.0 or later)
- MongoDB 2.4 or later (latest stable version is recommended)
- Oracle Java SE 8 (OpenJDK 8 also works; latest stable update is recommended)

**Caution:** Graylog 2.x does not work with Elasticsearch 5.x!
Upgrading to Graylog 2.0.x

Elasticsearch 2.x

The embedded Elasticsearch node being used by Graylog has been upgraded to Elasticsearch 2.x which includes some breaking changes. Graylog 2.x does not work with Elasticsearch 1.x anymore and cannot communicate with existing Elasticsearch 1.x clusters.

Please see Breaking changes in Elasticsearch 2.x for details.

The blog article Key points to be aware of when upgrading from Elasticsearch 1.x to 2.x also contains interesting information about the upgrade path from Elasticsearch 1.x to 2.x.

Multicast Discovery

Multicast discovery has been removed from Elasticsearch 2.x (although it is still provided as an Elasticsearch plugin for now).

To reflect this change, the `elasticsearch_discovery_zen_ping_unicast_hosts` now has to contain the address of at least one Elasticsearch node in the cluster which Graylog can connect to.

Default network host

The network interface which Elasticsearch binds to (elasticsearch_network_host) has been changed to localhost (i.e. 127.0.0.1 or ::1); see Network changes/Bind to localhost.

If Elasticsearch is not running on the same machine, `elasticsearch_network_host` must be set to a host name or an IP address which can be accessed by the other Elasticsearch nodes in the cluster.

Index range types

Note: This step needs to be performed before the update to Elasticsearch 2.x!

Some Graylog versions stored meta information about indices in elasticsearch, alongside the messages themselves. Since Elasticsearch 2.0 having multiple types with conflicting mappings is no longer possible, which means that the index_range type must be removed before upgrading to Elasticsearch 2.x.
Find out if your setup is affected by running (replace $elasticsearch with the address of one of your Elasticsearch nodes)

curl -XGET $elasticsearch:9200/_all/_mapping/index_range; echo

If the output is {} you are not affected and can skip this step.

Otherwise, you need to delete the index_range type, Graylog does not use it anymore.

As Graylog sets older indices to read-only, first we need to remove the write block on those indices. Since we’ll be working with Elasticsearch’s JSON output, we recommend installing the jq utility which should be available on all popular package managers or directly at GitHub.

```
for i in `curl -s -XGET $elasticsearch:9200/_all/_mapping/index_range | jq -r "keys[]"`; do
    echo -n "Updating index $i: 
    curl -XPUT $elasticsearch:9200/$i/_settings -d '{"index.blocks.read_only":false, "index.blocks.write":false}'
    echo
done
```

The output for each of the curl commands should be {"acknowledged":true}. Next we have to delete the index_range mapping. We can perform this via the next command.

```
for i in `curl -s -XGET $elasticsearch:9200/_all/_mapping/index_range | jq -r "keys[]"`; do
    echo -n "Updating index $i: 
    curl -XDELETE $elasticsearch:9200/$i/index_range
    echo
done
```

It is not strictly necessary to set the indices back to read only, but if you prefer to do that, note the index names and commands during the first step and change the false into true.

**Graylog Index Template**

Graylog applies a custom index template to ensure that the indexed messages adhere to a specific schema.

Unfortunately the index template being used by Graylog 1.x is incompatible with Elasticsearch 2.x and has to be removed prior to upgrading.

In order to delete the index template the following curl command has to be issued against on of the Elasticsearch nodes:

```
curl -X DELETE http://localhost:9200/_template/graylog-internal
```

Graylog will automatically create the new index template on the next startup.

**Dots in field names**

One of the most important breaking changes in Elasticsearch 2.x is that field names may not contain dots anymore.

Using the Elasticsearch Migration Plugin might help to highlight some potential pitfalls if an existing Elasticsearch 1.x cluster should be upgraded to Elasticsearch 2.x.
**MongoDB**

Graylog 2.x requires MongoDB 2.4 or newer. We recommend using MongoDB 3.x and the WiredTiger storage engine. When upgrading from MongoDB 2.0 or 2.2 to a supported version, make sure to read the Release Notes for the particular version.

**Log4j 2 migration**

Graylog switched its logging backend from Log4j 1.2 to Log4j 2. Please refer to the Log4j Migration Guide for information on how to update your existing logging configuration.

**Dead Letters feature removed**

The Dead Letters feature, which stored messages that couldn’t be indexed into Elasticsearch for various reasons, has been removed.

This feature has been disabled by default. If you have enabled the feature the configuration file, please check the dead_letters_enabled collection in MongoDB and remove it afterwards.

**Removed configuration settings**

**Index Retention and Rotation Settings**

In 2.0.0 the index rotation and retention settings have been moved from the Graylog server config file to the database and are now configurable via the web interface.

The old settings from the graylog.conf or /etc/graylog/server/server.conf will be migrated to the database.

**Warning:** When you upgrade from a 1.x version and you modified any rotation/retention settings, please make sure you KEEP your old settings in the config file so the migration process will add your old settings to the database! Otherwise the retention process will use the default settings and might remove a lot of indices.

**Overview**

Some settings, which have been deprecated in previous versions, have finally been removed from the Graylog configuration file.
Table 4.1: Removed configuration settings

<table>
<thead>
<tr>
<th>Setting name</th>
<th>Replacement</th>
</tr>
</thead>
<tbody>
<tr>
<td>mongodb_host</td>
<td>mongodb_uri</td>
</tr>
<tr>
<td>mongodb_port</td>
<td>mongodb_uri</td>
</tr>
<tr>
<td>mongodb_database</td>
<td>mongodb_uri</td>
</tr>
<tr>
<td>mongodb_useauth</td>
<td>mongodb_uri</td>
</tr>
<tr>
<td>mongodb_user</td>
<td>mongodb_uri</td>
</tr>
<tr>
<td>mongodb_password</td>
<td>mongodb_uri</td>
</tr>
<tr>
<td>elasticsearch_node_name</td>
<td>elasticsearch_node_name_prefix</td>
</tr>
<tr>
<td>collectorExpirationThreshold</td>
<td>(moved to collector plugin)</td>
</tr>
<tr>
<td>collectorInactiveThreshold</td>
<td>(moved to collector plugin)</td>
</tr>
<tr>
<td>rotation_strategy</td>
<td>UI in web interface (System/Indices)</td>
</tr>
<tr>
<td>retention_strategy</td>
<td>UI in web interface (System/Indices)</td>
</tr>
<tr>
<td>elasticsearch_maxDocsPerIndex</td>
<td>UI in web interface (System/Indices)</td>
</tr>
<tr>
<td>elasticsearch_maxSizePerIndex</td>
<td>UI in web interface (System/Indices)</td>
</tr>
<tr>
<td>elasticsearch_maxTimePerIndex</td>
<td>UI in web interface (System/Indices)</td>
</tr>
<tr>
<td>elasticsearch_maxNumberOfIndices</td>
<td>UI in web interface (System/Indices)</td>
</tr>
<tr>
<td>deadLettersEnabled</td>
<td>None</td>
</tr>
</tbody>
</table>

**Changed configuration defaults**

For better consistency, the defaults of some configuration settings have been changed after the project has been renamed from *Graylog2* to *Graylog*.

Table 4.2: Configuration defaults

<table>
<thead>
<tr>
<th>Setting name</th>
<th>Old default</th>
<th>New default</th>
</tr>
</thead>
<tbody>
<tr>
<td>elasticsearch_cluster_name</td>
<td>graylog2</td>
<td>graylog</td>
</tr>
<tr>
<td>elasticsearch_node_name</td>
<td>graylog2-server</td>
<td>graylog-server</td>
</tr>
<tr>
<td>elasticsearch_index_prefix</td>
<td>graylog2</td>
<td>graylog</td>
</tr>
<tr>
<td>elasticsearch_discoveryZenPingUnicastHosts</td>
<td>empty</td>
<td>127.0.0.1:9300</td>
</tr>
<tr>
<td>elasticsearch_discoveryZenPingMulticastEnabled</td>
<td>true</td>
<td>false</td>
</tr>
<tr>
<td>mongodb_uri</td>
<td>mongodb://127.0.0.1:9300</td>
<td>mongodb://localhost:9300</td>
</tr>
</tbody>
</table>

**Changed prefixes for configuration override**

In the past it was possible to override configuration settings in Graylog using environment variables or Java system properties with a specific prefix.

For better consistency, these prefixes have been changed after the project has been renamed from *Graylog2* to *Graylog*.

Table 4.3: Configuration override prefixes

<table>
<thead>
<tr>
<th>Override</th>
<th>Old prefix</th>
<th>New prefix</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Environment variables</td>
<td>GRAYLOG2_</td>
<td>GRAYLOG_</td>
<td>GRAYLOG_IS_MASTER</td>
</tr>
<tr>
<td>System properties</td>
<td>graylog2.</td>
<td>graylog.</td>
<td>graylog.is_master</td>
</tr>
</tbody>
</table>

**REST API Changes**

The output ID key for the list of outputs in the /streams/* endpoints has been changed from _id to id.
Web Interface Config Changes

The web interface has been integrated into the Graylog server and was rewritten in React. Therefore configuring it has changed fundamentally since the last version(s). Please consult Web interface for details.

Please take note that the application.context configuration parameter present in Graylog 1.x (and earlier) is not existing anymore. The web interface can currently only be served without a path prefix.

Upgrading to Graylog 2.1.x

HTTPS Setup

Previous versions of Graylog were automatically generating a private key/certificate pair for HTTPS if either the private key or the certificate (or both) for rest_tls_key_file, rest_tls_cert_file, web_tls_key_file, or web_tls_cert_file couldn’t be read. While this feature is very comfortable for inexperienced users, it has lots of serious drawbacks like very weak key sizes (only 1024 bits), being untrusted by all TLS libraries used by web browsers and other client software (because they are self-signed and not included in the system’s CA/trust store), and problems with inter-node communications with other Graylog nodes.
Due to those shortcomings, the feature has been removed completely. Users need to use proper certificates or generate their own self-signed certificates and configure them with the appropriate settings, see Using HTTPS for reference.

**Web Interface Listener**

Graylog 2.0.x has been using separate listeners for the REST API and the web interface by default. The Graylog REST API on http://127.0.0.1:12900, the Graylog web interface on http://127.0.0.1:9000. Beginning with Graylog 2.1.0 it is possible to run both the REST API and the web interface on the same host/port-combination and this is now the default. This means that the REST API is now running on http://127.0.0.1:9000/api/ by default and the web interface is now running on http://127.0.0.1:9000/. Furthermore, all requests going to http://127.0.0.1:9000/api/ requesting a content-type of text/html or application/xhtml+xml are redirected to the web interface, therefore making it even easier to set up Graylog and use it behind proxies, expose it externally etc.

Please take note that you can still run the REST API and the web interface on two separate listeners. If you are running a Graylog 2.0.x configuration specifying web_listen_uri explicitly and you want to keep that, you do not have to change anything.

Please also take note, that when you have configured rest_listen_uri and web_listen_uri to run on the same host/port-combination, the following configuration directives will have no effect:

- web_enable_tls, web_tls_cert_file, web_tls_key_file, web_tls_key_password (These will depend on the TLS configuration of the REST listener).
- web_enable_cors, web_enable_gzip, web_thread_pool_size, web_max_initial_line_length, web_max_header_size (Those will depend on the corresponding settings of the REST listener).

**Internal Metrics to MongoDB**

Previous versions of Graylog included a (long deprecated) metrics reporter for writing internal metrics into MongoDB in a fixed interval of 1 second.

This feature has been removed completely and can be optionally pulled in by using the Graylog Metrics Reporter Plugins.

**Configuration file changes**

**Network settings**

The network settings in the Graylog configuration file (rest_listen_uri, rest_transport_uri, and web_listen_uri) are now using the default ports for the HTTP (80) and HTTPS (443) if no custom port was given. Previously those settings were using the custom ports 12900 (Graylog REST API) and 9000 (Graylog web interface) if no explicit port was given.

Examples:

<table>
<thead>
<tr>
<th>Configuration setting</th>
<th>Old effective URI</th>
<th>New effective URI</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>rest_listen_uri = http://127.0.0.1/</code></td>
<td><a href="http://127.0.0.1:12900/">http://127.0.0.1:12900/</a></td>
<td><a href="http://127.0.0.1:80/">http://127.0.0.1:80/</a></td>
</tr>
</tbody>
</table>
Collector Sidecar

The network changes are reflected in the Sidecar configuration as well and should be adopted. However it’s still possible to use the old API port by setting it explicitly. In case a mass deployment is too hard to change, just run the following to switch back to the old REST API port (OVA based installation):

```
sudo graylog-ctl set-listen-address --service rest --address http://0.0.0.0:12900
sudo graylog-ctl reconfigure
```

Graylog REST API

Removed resources

<table>
<thead>
<tr>
<th>Original resource</th>
<th>Replacement</th>
</tr>
</thead>
<tbody>
<tr>
<td>/system/buffers</td>
<td>/system/metrics/org.graylog2.buffers.input.size</td>
</tr>
<tr>
<td></td>
<td>/system/metrics/org.graylog2.buffers.input.usage</td>
</tr>
<tr>
<td></td>
<td>/system/metrics/org.graylog2.buffers.process.size</td>
</tr>
<tr>
<td></td>
<td>/system/metrics/org.graylog2.buffers.process.usage</td>
</tr>
<tr>
<td></td>
<td>/system/metrics/org.graylog2.buffers.output.size</td>
</tr>
<tr>
<td></td>
<td>/system/metrics/org.graylog2.buffers.output.usage</td>
</tr>
<tr>
<td>/system/buffers/classes</td>
<td>None</td>
</tr>
</tbody>
</table>

Removed index rotation/retention settings from “/system/configuration”

The index rotation and retention settings have been moved to MongoDB in Graylog 2.0.0 but the representation of the old configuration options was still present in the /system/configuration resource.

In order to stay in sync with the actual configuration file, the following values have been removed:

- rotation_strategy
- retention_strategy
- elasticsearch_max_docs_per_index
- elasticsearch_max_size_per_index
- elasticsearch_max_time_per_index
- elasticsearch_max_number_of_indices

The retention and rotation configuration settings can be retrieved using the following resources:

- /system/indices/rotation/config
- /system/indices/retention/config

For Plugin Authors

Between Graylog 2.0.x and 2.1.0 we also made changes to the Plugin API. These include:

- Removing org.graylog2.plugin.streams.Stream#getAlertCondition, as it was faulty and not easily replaceable with a working version without breaking our separation of models and persistence services.

If you are maintaining a plugin that was originally written for Graylog 1.x or 2.0.x, you need to make sure that your plugin is still compiling and working under Graylog 2.1.x or adapt it if necessary.
UI Plugins

The new app prefix feature requires some changes in UI plugins to make them work with that.

- `import webpackEntry from 'webpack-entry';` needs to be added at the very top of the `src/web/index.jsx` file
- The `Routes.pluginRoute()` function needs to be used instead of a literal string to build URLs for links and buttons

Please check the updated plugins documentation for details.

Changed Elasticsearch Cluster Status Behavior

In previous versions Graylog stopped indexing into the current write index if the Elasticsearch cluster status turned RED. Since Graylog 2.1.0 only checks the status of the current write index when it tries to index messages.

If the current write index is GREEN or YELLOW, Graylog will continue to index messages even though the overall cluster status is RED. This avoids Graylog downtimes when doing Elasticsearch maintenance or when older indices have problems.

Changes in message field values trimming

Previous versions of Graylog were trimming message field values inconsistently, depending on the codec used. We have changed that behaviour in Graylog 2.1.0, so all message field values are trimmed by default. This means that leading or trailing whitespace of every field is removed during ingestion.

**Important:** This change will break your existing stream rules, extractors, and Drool rules if you are expecting leading or trailing white spaces in them. Please adapt them so they do not require those white spaces.

Upgrading to Graylog 2.2.x

This file only contains the upgrade note for the upcoming release. Please see our documentation for the complete upgrade notes.

Email Alarm Callback

Previous versions of Graylog created an implicit email alarm callback if no explicit callback existed for a stream.

Due to the extensive rework done in alerting, this behavior has been modified to be explicit, and more consistent with other entities within Graylog: from now on there will not be a default alarm callback.

To simplify the transition for people relying on this behavior, we have added a migration step that will create an email alarm callback for each stream that has alert conditions, has alert receivers, but has no associated alarm callbacks.

With the introduction of email templates in 0.21, the `transport_email_subject_prefix` config setting became unused. It is now being removed completely. In early versions it was used to add a prefix to the generated subject of alerting emails. Since 0.21 it is possible to define a complete template used for the generation of alert email subjects.
Alert Notifications (previously known as Alarm Callbacks)

Graylog 2.2.0 introduces some changes in alerting. Alerts have now states, helping you to know in an easier way if something requires your attention.

These changes also affect the way we send notifications: Starting in Graylog 2.2.0, alert notifications are only executed once, just when a new alert is triggered. As long as the alert is unresolved or in grace period, Graylog will not send further notifications. This will help you reducing the noise and annoyance of getting notified way too often when a problem persists for a while.

If you are using Graylog for alerting, please take a moment to ensure this change will not break any of your processes when an alert occurs.

Default stream/Index Sets

With the introduction of index sets, and the ability to change a stream’s write target, the default stream needs additional information, which is calculated when starting a new Graylog 2.2 master node.

It requires recalculation of the index ranges of the default stream’s index set, which when updating from pre-2.2 versions is stored in the graylog_index. This is potentially expensive, because it has to calculate three aggregations across every open index to detect which streams are stored in which index.

Please be advised that this necessary migration can put additional load on your cluster.

| Warning: Make sure that all rotation and retention strategy plugins you had installed in 2.1 are updated to a version that is compatible with 2.2 before you start the Graylog 2.2 version for the first time. (e.g. Graylog Enterprise) This is needed so the required data migrations will run without problems. |
| Warning: The option to remove a message from the default stream is currently not available when using the pipeline function route_to_stream. This will be fixed in a subsequent bug fix release. Please see the corresponding Github issue. |

RotationStrategy & RetentionStrategy Interfaces

The Java interfaces for RetentionStrategy and RotationStrategy changed in 2.2. The #rotate() and #retain() methods are now getting an IndexSet as first parameter.

This only affects you if you are using custom rotation or retention strategies.

Changes in Exposed Configuration

The exposed configuration settings on the /system/configuration resource of the Graylog REST API doesn’t contain the following (deprecated) Elasticsearch-related settings anymore:

- elasticsearch_shards
- elasticsearch_replicas
- index_optimization_max_num_segments
- disable_index_optimization
Changes in Split & Count Converter

The behavior of the split & count converter has been changed to that it resembles typical `split()` functions. Previously, the split & count converter returned 0, if the split pattern didn’t occur in the string. Now it will return 1.

Examples:

<table>
<thead>
<tr>
<th>String</th>
<th>Split Pattern</th>
<th>Old Result</th>
<th>New Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;empty&gt;</td>
<td>–</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>foo</td>
<td>–</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>foo-bar</td>
<td>–</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

Graylog REST API

Streams API

Due to the introduction of index sets, the payload for creating, updating and cloning of streams now requires the `index_set_id` field. The value for this needs to be the ID of an existing index set.

Affected endpoints:

- POST /streams
- PUT /streams/{streamId}
- POST /streams/{streamId}/clone
CHAPTER 5

Configuring Graylog

The graylog-ctl script

Some packages of Graylog (for example the virtual machine appliances) ship with a pre-installed graylog-ctl script to allow you easy configuration of certain settings.

**Important:** graylog-ctl is only available in the virtual machine appliances, but not in the tar-ball (for manual setup), operating system packages, or configuration management scripts (Puppet, Chef, Ansible).

Configuration commands

The following commands are changing the configuration of Graylog:
<table>
<thead>
<tr>
<th>Command</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>sudo graylog-ctl set-admin-password &lt;password&gt;</code></td>
<td>Set a new admin password</td>
</tr>
<tr>
<td><code>sudo graylog-ctl set-admin-username &lt;username&gt;</code></td>
<td>Set a different username for the admin user</td>
</tr>
<tr>
<td><code>sudo graylog-ctl set-email-config &lt;smtp server&gt; [--port=&lt;smtp port&gt; --user=&lt;username&gt; --password=&lt;password&gt; --from-email=&lt;sender-address&gt; --web-url=&lt;graylog web-interface url&gt; --no-tls --no-ssl]</code></td>
<td>Configure SMTP settings to send alert mails</td>
</tr>
<tr>
<td><code>sudo graylog-ctl set-timezone &lt;zone acronym&gt;</code></td>
<td>Set Graylog’s time zone from a list of valid time zones. Make sure system time is also set correctly with <code>sudo dpkg-reconfigure tzdata</code>.</td>
</tr>
<tr>
<td><code>sudo graylog-ctl enforce-ssl</code></td>
<td>Enforce HTTPS for the web interface</td>
</tr>
<tr>
<td><code>sudo graylog-ctl set-node-id &lt;id&gt;</code></td>
<td>Override random server node id</td>
</tr>
<tr>
<td><code>sudo graylog-ctl set-server-secret &lt;secret&gt;</code></td>
<td>Override server secret used for encryption</td>
</tr>
<tr>
<td><code>sudo graylog-ctl disable-internal-logging</code></td>
<td>Disable sending internal logs (e.g. nginx) from the VM to Graylog. Reboot is needed for activation!</td>
</tr>
<tr>
<td><code>sudo graylog-ctl set-external-ip http[s]://&lt;public IP&gt;:port/</code></td>
<td>Configure an external IP in the Nginx proxy. This is needed to connect the web interface to the REST API e.g. in NAT’d networks or on AWS.</td>
</tr>
<tr>
<td>`sudo graylog-ctl set-listen-address --service &lt;web</td>
<td>rest</td>
</tr>
<tr>
<td><code>sudo graylog-ctl local-connect</code></td>
<td>Bind all services but the web interface to 127.0.0.1</td>
</tr>
<tr>
<td>`sudo graylog-ctl set-mongodb-password [-a</td>
<td>-g] -u &lt;username&gt; -p &lt;password&gt;`</td>
</tr>
</tbody>
</table>
Commands for multi node setups:

<table>
<thead>
<tr>
<th>Command</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>sudo graylog-ctl set-cluster-master &lt;IP of master node&gt;</code></td>
<td>Set IP address of node where others can fetch cluster configuration from</td>
</tr>
<tr>
<td><code>sudo graylog-ctl reconfigure-as-backend</code></td>
<td>Run Graylog server and Elasticsearch on this node</td>
</tr>
<tr>
<td><code>sudo graylog-ctl reconfigure-as-datanode</code></td>
<td>Run Elasticsearch on this node only</td>
</tr>
<tr>
<td><code>sudo graylog-ctl reconfigure-as-server</code></td>
<td>Run Graylog server on this node only</td>
</tr>
</tbody>
</table>

General commands:
<table>
<thead>
<tr>
<th>Command</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>sudo graylog-ctl cleanse</code></td>
<td>Delete all graylog data, and start from scratch</td>
</tr>
<tr>
<td><code>sudo graylog-ctl graceful-kill</code></td>
<td>Attempt a graceful stop, then SIGKILL the entire process group</td>
</tr>
<tr>
<td><code>sudo graylog-ctl hup</code></td>
<td>Send the services a HUP signal</td>
</tr>
<tr>
<td><code>sudo graylog-ctl int</code></td>
<td>Send the services an INT signal</td>
</tr>
<tr>
<td><code>sudo graylog-ctl term</code></td>
<td>Send the services a TERM signal</td>
</tr>
<tr>
<td><code>sudo graylog-ctl kill</code></td>
<td>Send the services a KILL signal</td>
</tr>
<tr>
<td><code>sudo graylog-ctl list-servers</code></td>
<td>List all Graylog servers in your cluster</td>
</tr>
<tr>
<td><code>sudo graylog-ctl status</code></td>
<td>Show the status of all the services</td>
</tr>
<tr>
<td><code>sudo graylog-ctl start</code></td>
<td>Start services if they are down, and restart them if they stop</td>
</tr>
<tr>
<td><code>sudo graylog-ctl stop</code></td>
<td>Stop the services, and do not restart them</td>
</tr>
<tr>
<td><code>sudo graylog-ctl restart</code></td>
<td>Stop the services if they are running, then start them again</td>
</tr>
<tr>
<td><code>sudo graylog-ctl once</code></td>
<td>Start the services if they are down. Do not restart them if they stop</td>
</tr>
<tr>
<td><code>sudo graylog-ctl uninstall</code></td>
<td>Kill all processes and uninstall the process supervisor (data will be preserved)</td>
</tr>
<tr>
<td><code>sudo graylog-ctl tail</code></td>
<td>Watch the service logs of all enabled services</td>
</tr>
<tr>
<td><code>sudo graylog-ctl tail &lt;service name&gt;</code></td>
<td>Watch the logs of just one service, name can be 'server', 'elasticsearch', 'mongodb', 'nginx', 'etcd'</td>
</tr>
<tr>
<td><code>sudo graylog-ctl show-config</code></td>
<td>Show the service configuration</td>
</tr>
<tr>
<td><code>sudo graylog-ctl reconfigure</code></td>
<td>Reconfigure the application</td>
</tr>
</tbody>
</table>
**Important:** After using a command that changes the application configuration re-run `sudo graylog-ctl reconfigure` to actually enable the changes.

## Multi VM setup

At some point it make sense to not run all services on a single VM anymore. For performance reasons you might want to add more Elasticsearch nodes to the cluster or even add a second Graylog server. This can be achieved by changing IP addresses in the Graylog configuration files by hand or use our canned configurations which come with the `graylog-ctl` command.

The idea is to have one VM which is a central point for other VMs to fetch all needed configuration settings to join the cluster. Typically the first VM you spin up is used for this task. Automatically an instance of etcd is started and filled with the necessary settings for other hosts.

For example, to create a small cluster with a dedicated Graylog server node and another for Elasticsearch, spin up two VMs from the same Graylog image. On the first one start only Graylog and MongoDB:

```shell
vm1> sudo graylog-ctl set-admin-password sEcReT
vm1> sudo graylog-ctl reconfigure-as-server
```

On the second VM start only Elasticsearch. Before doing so set the IP of the first VM to fetch the configuration data from there:

```shell
vm2> sudo graylog-ctl set-cluster-master <ip-of-vm1>
vm2> sudo graylog-ctl reconfigure-as-datanode
vm1> sudo graylog-ctl reconfigure-as-server
```

This results in a perfectly fine dual VM setup. However if you want to scale this setup out by adding an additional Elasticsearch node, you can proceed in the same way:

```shell
vm3> sudo graylog-ctl set-cluster-master <ip-of-vm1>
vm3> sudo graylog-ctl reconfigure-as-datanode
vm1> sudo graylog-ctl reconfigure-as-server
vm2> sudo graylog-ctl reconfigure-as-datanode
```

Verify that all nodes are working as a cluster by going to the Kopf plugin on one of the Elasticsearch nodes open [http://vm2:9200/_plugin/kopf/#!/nodes](http://vm2:9200/_plugin/kopf/#!/nodes).

**Important:** In case you want to add a second Graylog server you have to set the same server secret on all machines. The secret is stored in the file `/etc/graylog/graylog-secrets` and can be applied to other hosts with the `set-server-secret` sub-command.

The following configuration modes do exist:

<table>
<thead>
<tr>
<th>Command</th>
<th>Services</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>sudo graylog-ctl reconfigure</code></td>
<td>Run all services on this box</td>
</tr>
<tr>
<td><code>sudo graylog-ctl reconfigure-as-server</code></td>
<td>Run Graylog, web and MongoDB (no Elasticsearch)</td>
</tr>
<tr>
<td><code>sudo graylog-ctl reconfigure-as-backend</code></td>
<td>Run Graylog, Elasticsearch and MongoDB (no nginx for web interface access)</td>
</tr>
<tr>
<td><code>sudo graylog-ctl reconfigure-as-datanode</code></td>
<td>Run only Elasticsearch</td>
</tr>
</tbody>
</table>

A server with only the web interface running is not supported as of Graylog 2.0. The web interface is now included in the server process. But you can create your own service combinations by editing the

---

### 5.1. The graylog-ctl script

75
file `/etc/graylog/graylog-services.json` by hand and enable or disable single services. Just run `graylog-ctl reconfigure` afterwards.

**Extend disk space**

All data of the appliance setup is stored in `/var/opt/graylog/data`. In order to extend the disk space mount a second (virtual) hard drive into this directory.

**Important:** Make sure to move old data to the new drive before and give the graylog user permissions to read and write here.

**Example procedure for the Graylog virtual appliance**

**Note:** These steps require basic knowledge in using Linux and the common shell programs.

- Shutdown the virtual machine as preparation for creating a consistent snapshot.
- Take a snapshot of the virtual machine in case something goes wrong.
  - Understanding VM snapshots in ESXi / ESX
  - VMware vSphere: Managing Snapshots
  - VirtualBox: Snapshots
  - Parallels: Save Snapshots of a Virtual Machine
  - Parallels: Working with snapshots
- Attach an additional hard drive to the virtual machine.
  - VMware Workstation: Adding a New Virtual Disk to a Virtual Machine
  - VMware vSphere: Virtual Disk Configuration
  - VirtualBox: Virtual storage
  - Parallels: Hard Disk
- Start the virtual machine again.
- Stop all services to prevent disk access:
  ```
  $ sudo graylog-ctl stop
  ```
- Check for the *logical name* of the new hard drive. Usually this is `/dev/sdb`:
  ```
  $ sudo lshw -class disk
  ```
- Partition and format new disk:
  ```
  $ sudo parted -a optimal /dev/sdb mklabel gpt
  # A reboot may be necessary at this point so that the updated GPT is being recognized by the operating system
  $ sudo parted -a optimal -- /dev/sdb unit compact mkpart primary ext3 "1" "-1"
  $ sudo mkfs.ext4 /dev/sdb1
  ```
- Mount disk into temporary directory `/mnt/tmp`:
$ sudo mkdir /mnt/tmp
$ sudo mount /dev/sdb1 /mnt/tmp

• Copy current data to new disk:
$ sudo cp -ax /var/opt/graylog/data/* /mnt/tmp/

• Compare both folders:
# Output should be: Only in /mnt/tmp: lost+found
$ sudo diff -qr --suppress-common-lines /var/opt/graylog/data /mnt/tmp

• Delete old data:
$ sudo rm -rf /var/opt/graylog/data/*

• Mount new disk into /var/opt/graylog/data directory:
$ sudo umount /mnt/tmp
$ sudo mount /dev/sdb1 /var/opt/graylog/data

• Make change permanent by adding an entry to /etc/fstab:
$ echo '/dev/sdb1 /var/opt/graylog/data ext4 defaults 0 0' | sudo tee -a /etc/fstab

• Reboot virtual machine:
$ sudo shutdown -r now

Install Graylog plugins
The Graylog plugin directory is located in /opt/graylog/plugin/. Just drop a JAR file there and restart the server with `sudo graylog-ctl restart graylog-server` to load the plugin.

Install Elasticsearch plugins
Elasticsearch comes with a helper program to install additional plugins you can call it like this `sudo JAVA_HOME=/opt/graylog/embedded/jre /opt/graylog/elasticsearch/bin/plugin`

Install custom SSL certificates
During the first reconfigure run self signed SSL certificates are generated. You can replace this certificate with your own to prevent security warnings in your browser. Just drop the key and combined certificate file here: `/opt/graylog/conf/nginx/ca/graylog.crt` respectively `/opt/graylog/conf/nginx/ca/graylog.key`. Afterwards restart `nginx` with `sudo graylog-ctl restart nginx`.

Assign a static IP
Per default the appliance make use of DHCP to setup the network. If you want to access Graylog under a static IP please follow these instructions:
Graylog Documentation, Release 2.2.1

$ sudo ifdown eth0

Edit the file /etc/network/interfaces like this (just the important lines):

```plaintext
auto eth0
  iface eth0 inet static
  address <static IP address>
  netmask <netmask>
  gateway <default gateway>
  pre-up sleep 2
```

Activate the new IP and reconfigure Graylog to make use of it:

```
$ sudo ifup eth0
$ sudo graylog-ctl reconfigure
```

Wait some time until all services are restarted and running again. Afterwards you should be able to access Graylog with the new IP.

Upgrade Graylog

**Warning:** The Graylog omnibus package does *not* support unattended upgrading from Graylog 1.x to Graylog 2.x!

Always perform a full backup or snapshot of the appliance before proceeding. Only upgrade if the release notes say the next version is a drop-in replacement. Choose the Graylog version you want to install from the list of Omnibus packages.

```
$ wget https://packages.graylog2.org/releases/graylog-omnibus/ubuntu/graylog_latest.deb
$ sudo graylog-ctl stop
$ sudo dpkg -G -i graylog_latest.deb
$ sudo graylog-ctl reconfigure
$ sudo reboot
```

**Note:** Reboot the server after the update procedure to make sure that all services are running in the correct version.

Migrate manually from 1.x to 2.x

To update a 1.x appliance to 2.x the administrator has to purge the Graylog installation, migrate the stored log data and install the new version as Omnibus package. Before upgrading read the upgrade notes. This procedure can potentially delete log data or configuration settings. So it’s absolutely necessary to perform a backup or a snapshot before!

Stop all services but Elasticsearch:

```
$ sudo -s
$ graylog-ctl stop graylog-web
$ graylog-ctl stop graylog-server
$ graylog-ctl stop mongodb
$ graylog-ctl stop nginx
$ graylog-ctl stop etcd
```

Check for index range types. The output of this command should be `/`, if not read these notes for how to fix this:
Delete the Graylog index template:

```
$ curl -X DELETE <appliance_IP>:9200/_template/graylog-internal
```

Migrate appliance configuration:

```
$ cd /etc
$ mv graylog graylog2.2
$ vi graylog2.2/graylog-secrets.json

# Remove the graylog_web section
}, << don't forget the comma!
"graylog_web": {
    "secret_token": "3552c87f3e3..."
}

$ vi graylog2.2/graylog-services.json

# Remove the graylog_web section
}, << don't forget the comma!
"graylog_web": {
    "enabled": true
}

$ vi graylog2.2/graylog-settings.json

# Remove "rotation_size", "rotation_time", "indices"
"enforce_ssl": false,
"rotation_size": 1073741824,
"rotation_time": 0,
"indices": 10,
"journal_size": 1,
```

Migrate appliance data:

```
$ cd /var/opt
$ mv graylog graylog2.2
$ mv graylog2.2/data/elasticsearch/graylog2 graylog2.2/data/elasticsearch/graylog
```

Delete old Graylog version and install new Omnibus package:

```
$ wget http://packages.graylog2.org/releases/graylog-omnibus/ubuntu/graylog_2.2.1-1_amd64.deb
$ apt-get purge graylog
$ dpkg -i graylog_2.2.1-1_amd64.deb
```

Move directories back:

```
$ cd /etc
$ mv graylog2.2 graylog
$ cd /var/opt/
$ mv graylog2.2 graylog
```

Reconfigure and Reboot:

```
$ graylog-ctl reconfigure
$ reboot
```

5.1. The graylog-ctl script
Graylog should now be updated and old data still available.

**Important:** The index retention configuration moved from the Graylog configuration file to the web interface. After the first start go to ‘System -> Indices -> Update configuration’ to re-enable your settings.

### Advanced Settings

To change certain parameters used by `graylog-ctl` during a reconfigure run you can override all default parameters found in the `attributes` file.

If you want to change the username used by Graylog for example, edit the file `/etc/graylog/graylog-settings.json` like this:

```json
"custom_attributes": {
    "user": {
        "username": "log-user"
    }
}
```

Afterwards run `sudo graylog-ctl reconfigure` and `sudo graylog-ctl restart`. The first command renders all changed configuration files and the later makes sure that all services restart to activate the change.

There are a couple of other use cases of this, e.g. change the default data directories used by Graylog to `/data` (make sure this is writeable by the graylog user):

```json
"custom_attributes": {
    "elasticsearch": {
        "data_directory": "/data/elasticsearch"
    },
    "mongodb": {
        "data_directory": "/data/mongodb"
    },
    "etcd": {
        "data_directory": "/data/etcd"
    },
    "graylog-server": {
        "journal_directory": "/data/journal"
    }
}
```

Or change the default memory settings used by Graylog or Elasticsearch:

```json
"custom_attributes": {
    "graylog-server": {
        "memory": "1700m"
    },
    "elasticsearch": {
        "memory": "2200m"
    }
}
```

Again, run `reconfigure` and `restart` afterwards to activate the changes.
Securing an appliance

Even though the Graylog appliances are not meant for production use there are still two commands you can use to increase the security of an installation. With `graylog-ctl local-connect` only the web interface is reachable from the outside. All other services are listening on the local loopback device. This is only useful when you run the appliance as a single node. Clustered setups are not possible anymore. But data stored in MongoDB or Elasticsearch are protected from direct external access.

The other one is `graylog-ctl set-mongodb-password`. This command enables authentication for MongoDB and creates or updates a database user. First an admin user should be created. This user is needed for database maintenance and future password changes. Afterwards an unprivileged service user can be created for Graylog. The procedure works like this:

```
$ graylog-ctl set-mongodb-password -a -u admin -p someAdminPassword123
$ graylog-ctl set-mongodb-password -g -u graylog -p someGraylogServicePassword
$ graylog-ctl reconfigure
```

MongoDB and the Graylog server will be restarted with activated authentication. The username and password needs to be set on every Graylog node to make a cluster work. Login to another Graylog server and only set the service user:

```
$ graylog-ctl set-cluster-master 1.1.1.2
$ graylog-ctl set-mongodb-password -g -u graylog -p someGraylogServicePassword
$ graylog-ctl reconfigure-as-server
```

Since the pre-build appliances are based on standard Ubuntu-Linux, tools like iptables/SELinux/AppArmor can be used additionally. But to explain all available countermeasurements would go beyond this documentation.

Web interface

When your Graylog instance/cluster is up and running, the next thing you usually want to do is check out our web interface, which offers you great capabilities for searching and analyzing your indexed data and configuring your Graylog environment. Per default you can access it using your browser on `http://<graylog-server>:9000/api/`.

Overview

The Graylog web interface was rewritten in JavaScript for 2.0 to be a client-side single-page browser application. This means its code is running solely in your browser, fetching all data via HTTP(S) from the REST API of your Graylog server.

Note: Both the web interface URI (see `web_listen_uri`) and the REST API (see `rest_listen_uri` and `rest_transport_uri`) must be accessible by everyone using the web interface. This means that Graylog must listen on a public network interface or be exposed to one using a proxy or NAT!

Single or separate listeners for web interface and REST API?

Since Graylog 2.1 you have two options when it comes to exposing its web interface:

- Running both on the same port, using different paths (defaulting to `http://localhost:9000/api/` for the REST API and `http://localhost:9000/` for the web interface), this is the default since 2.1 and is assumed for most parts of the documentation.
• Running on two different ports (for example http://localhost:12900/ for the REST API and http://localhost:9000/ for the web interface)

**Note:** When you are using the first option and you want to run the REST API and the web interface on the same host and port, the path part of both URIs (rest_listened_uri & web_listened_uri) must be different and the path part of web_listened_uri must be non-empty and different than /.

### Configuration Options

If our default settings do not work for you, there is a number of options in the Graylog server configuration file which you can change to influence its behavior:

<table>
<thead>
<tr>
<th>Setting</th>
<th>Default</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>web_enable</td>
<td>true</td>
<td>Determines if the web interface endpoint is started or not.</td>
</tr>
<tr>
<td>web_listened_uri</td>
<td><a href="http://127.0.0.1:9000/">http://127.0.0.1:9000/</a></td>
<td>Default address the web interface listener binds to.</td>
</tr>
<tr>
<td>web_endpoint_uri</td>
<td>if not set, rest_transport_uri will be used.</td>
<td>This is the external address of the REST API of the Graylog server. Web interface clients need to be able to connect to this for the web interface to work.</td>
</tr>
<tr>
<td>web_enable_cors</td>
<td>false</td>
<td>Support Cross-Origin Resource Sharing for the web interface assets. Not required, because no REST calls are made to this listener. This setting is ignored, if the host and port parts of web_listened_uri and rest_listened_uri are identical.</td>
</tr>
<tr>
<td>web_enable_gzip</td>
<td>true</td>
<td>Serve web interface assets using compression. This setting is ignored, if the host and port parts of web_listened_uri and rest_listened_uri are identical.</td>
</tr>
<tr>
<td>web_enable_tls</td>
<td>false</td>
<td>Should the web interface serve assets using encryption or not. This setting is ignored, if the host and port parts of web_listened_uri and rest_listened_uri are identical.</td>
</tr>
<tr>
<td>web_tls_cert_file</td>
<td>(no default)</td>
<td>Path to TLS certificate file, if TLS is enabled. This setting is ignored, if the host and port parts of web_listened_uri and rest_listened_uri are identical.</td>
</tr>
<tr>
<td>web_tls_key_file</td>
<td>(no default)</td>
<td>Path to private key for certificate, used if TLS is enabled. This setting is ignored, if the host and port parts of web_listened_uri and rest_listened_uri are identical.</td>
</tr>
<tr>
<td>web_tls_key_password</td>
<td>(no default)</td>
<td>Password for TLS key (if it is encrypted). This setting is ignored, if the host and port parts of web_listened_uri and rest_listened_uri are identical.</td>
</tr>
<tr>
<td>web_thread_pool_size</td>
<td>16</td>
<td>Number of threads used for web interface listener. This setting is ignored, if the host and port parts of web_listened_uri and rest_listened_uri are identical.</td>
</tr>
</tbody>
</table>

**How does the web interface connect to the Graylog server?**

The web interface is fetching all information it is showing from the REST API of the Graylog server. Therefore it needs to connect to it using HTTP(S). There are several ways how you can define which way the web interface connects to the Graylog server. The URI used by the web interface is determined in this exact order:

- If the HTTP(S) client going to the web interface port sends a X-Graylog-Server-URL header, which contains a valid URL, then this is overriding everything else.
- If web_endpoint_uri is defined in the Graylog configuration file, this is used if the aforementioned header is not set.
• If both are not defined, `rest_transport_uri` is used.

**Browser Compatibility**

Writing the web interface as a single-page application is a challenging task. We want to provide the best possible experience to everyone, which often means using modern web technology only available in recent browsers, while keeping a reasonable compatibility with old and less-capable browsers. These browsers are officially supported in Graylog 2.0:

<table>
<thead>
<tr>
<th>Browser</th>
<th>OS</th>
<th>Minimum Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chrome</td>
<td>Windows, OS X, Linux</td>
<td>50</td>
</tr>
<tr>
<td>Firefox</td>
<td>Windows, OS X, Linux</td>
<td>45 / 38 ESR</td>
</tr>
<tr>
<td>Internet Explorer</td>
<td>Windows</td>
<td>11</td>
</tr>
<tr>
<td>Microsoft Edge</td>
<td>Windows</td>
<td>25</td>
</tr>
<tr>
<td>Safari</td>
<td>OS X</td>
<td>9</td>
</tr>
</tbody>
</table>

Please take into account that you need to enable JavaScript in order to use Graylog web interface.

**Making the web interface work with load balancers/proxies**

If you want to run a load balancer/reverse proxy in front of Graylog, you need to make sure that:

• The REST API port is accessible for clients
• The address for the Graylog server’s REST API is properly set (as explained in *How does the web interface connect to the Graylog server?*), so it is resolvable and accessible for any client of the web interface.
• You are either using only HTTP or only HTTPS (no mixed content) for both the web interface endpoint and the REST API endpoint.
• If you use SSL, your certificates must be valid and trusted by your clients.

**Note:** To help you with your specific environment, we have some example configurations. We take the following assumption in all examples. Your Graylog server.conf has the following settings set `rest_listen_uri` = `http://127.0.0.1:9000/api/` and `web_listen_uri` = `http://127.0.0.1:9000/`. Your URL will be `graylog.example.org` with the IP `192.168.0.10`.

**Using a Layer 3 load balancer (forwarding TCP Ports)**

1. Configure your load balancer to forward connections going to `192.168.0.10:80` to `127.0.0.1:9000` (web_listen_uri) and `192.168.0.10:9000/api/` to `127.0.0.1:9000/api/` (rest_listen_uri).
3. Start the Graylog server as usual.
5. Read up on *Using HTTPS*.
NGINX

REST API and Web Interface on one port (using HTTP):

```plaintext
server {
    listen 80 default_server;
    listen [::]:80 default_server ipv6only=on;
    server_name graylog.example.org;

    location / {
        proxy_set_header Host $http_host;
        proxy_set_header X-Forwarded-Host $host;
        proxy_set_header X-Forwarded-Server $host;
        proxy_set_header X-Forwarded-For $proxy_add_x_forwarded_for;
        proxy_set_header X-Graylog-Server-URL http://graylog.example.org/api;
        proxy_pass http://127.0.0.1:9000;
    }
}
```

NGINX can be used for SSL Termination, you would only need to modify the `server listen` directive and add all information about your certificate.

If you are running multiple Graylog Server you might want to use HTTPS/SSL to connect to the Graylog Servers (on how to Setup read Using HTTPS) and use HTTPS/SSL on NGINX. The configuration for TLS certificates, keys and ciphers is omitted from the sample config for brevity’s sake.

REST API and Web Interface on one port (using HTTPS/SSL):

```plaintext
server {
    listen 443 ssl spdy;
    server_name graylog.example.org;
    # <- your SSL Settings here!

    location / {
        proxy_set_header X-Forwarded-For $proxy_add_x_forwarded_for;
        proxy_set_header Host $http_host;
        proxy_set_header X-Graylog-Server-URL https://graylog.example.org/api;
        proxy_pass http://127.0.0.1:9000;
    }
}
```

Apache httpd 2.x

REST API and Web Interface on one port (using HTTP):

```plaintext
<VirtualHost *:80>
    ServerName graylog.example.org
    ProxyRequests Off
    <Proxy »>
        Order deny,allow
        Allow from all
    </Proxy>
</VirtualHost>

<Location />
    RequestHeader set X-Graylog-Server-URL "http://graylog.example.org/api/"
</Location>
```
REST API and Web Interface on one port (using HTTPS/SSL):

```xml
<VirtualHost *:443>
    ServerName graylog.example.org
    ProxyRequests Off
    SSLEngine on
    # <- your SSL Settings here!

    <Proxy>*</Proxy>
    Order deny,allow
    Allow from all
</VirtualHost>

<Location>
    RequestHeader set X-Graylog-Server-URL "https://graylog.example.org/api/"
    ProxyPass http://127.0.0.1:9000/
    ProxyPassReverse http://127.0.0.1:9000/
</Location>
</VirtualHost>
```

HAProxy 1.6

REST API and Web Interface on one port (using HTTP):

```yaml
frontend http
    bind 0.0.0.0:80
    option forwardfor
    http-request add-header X-Forwarded-Host %[req.hdr(host)]
    http-request add-header X-Forwarded-Server %[req.hdr(host)]
    http-request add-header X-Forwarded-Port %[dst_port]
    acl is_graylog hdr_dom(host) -i -m str graylog.example.org
    use_backend graylog

backend graylog
    description The Graylog Web backend.
    http-request set-header X-Graylog-Server-URL http://graylog.example.org/api
    use-server graylog_1
    server graylog_1 127.0.0.1:9000 maxconn 20 check

Multiple Backends (roundrobin) with Health-Check (using HTTP):

frontend graylog_http
    bind *:80
    option forwardfor
    http-request add-header X-Forwarded-Host %[req.hdr(host)]
    http-request add-header X-Forwarded-Server %[req.hdr(host)]
    http-request add-header X-Forwarded-Port %[dst_port]
    acl is_graylog hdr_dom(host) -i -m str graylog.example.org
    use_backend graylog
```

5.2. Web interface
Load balancer integration

When running multiple Graylog servers a common deployment scenario is to route the message traffic through an IP load balancer. By doing this we can achieve both a highly available setup, as well as increasing message processing throughput, by simply adding more servers that operate in parallel.

Load balancer state

However, load balancers usually need some way of determining whether a backend service is reachable and healthy or not. For this purpose Graylog exposes a load balancer state that is reachable via its REST API.

There are two ways the load balancer state can change:

- due to a lifecycle change (e.g. the server is starting to accept messages, or shutting down)
- due to manual intervention via the REST API

**Note:** In the following examples we assume that the Graylog REST API is available on the URI path `/api/` (e.g. `http://graylog.example.com/api/`).

To query the current load balancer status of a Graylog instance, all you need to do is to issue a HTTP call to its REST API:

```
GET /api/system/lbstatus
```

The status knows two different states, **ALIVE** and **DEAD**, which is also the `text/plain` response of the resource. Additionally, the same information is reflected in the HTTP status codes: If the state is **ALIVE** the return code will be `200 OK`, for **DEAD** it will be `503 Service unavailable`. This is done to make it easier to configure a wide range of load balancer types and vendors to be able to react to the status.

The resource is accessible without authentication to make it easier for load balancers to access it.

To programmatically change the load balancer status, an additional endpoint is exposed:

```
PUT /api/system/lbstatus/override/alive
PUT /api/system/lbstatus/override/dead
```

Only authenticated and authorized users are able to change the status, in the currently released Graylog version this means only admin users can change it.

Graceful shutdown

Often, when running a service behind a load balancer, the goal is to be able to perform zero-downtime upgrades, by taking one of the servers offline, upgrading it, and then bringing it back online. During that time the remaining servers can take the load seamlessly.
By using the load balancer status API described above one can already perform such a task. However, it would still be guesswork when the Graylog server is done processing all the messages it already accepted.

For this purpose Graylog supports a graceful shutdown command, also accessible via the web interface and API. It will set the load balancer status to DEAD, stop all inputs, turn on messages processing (should it have been disabled manually previously), and flush all messages in memory to Elasticsearch. After all buffers and caches are processed, it will shut itself down safely.

**Web Interface**

It is possible to use the Graylog web interface behind a load balancer for high availability purposes.

**Note:** Take care of the configuration you need with a proxy setup, as it will not work out of the box.

As of version 2.0 you do not need a sticky session (previously required under 1.x). It is still important, however, that the API URL /api/system/deflector/cycle hits the configured Graylog master node, although this might change in a future release.

Please refer to your vendor’s documentation to learn how to route this URL to a specific host.

**Using HTTPS**

We highly recommend securing your Graylog installation using SSL/TLS to make sure that no sensitive data is sent over the wire in plain text. To make this work, you need to do two things:

- Enable TLS for the Graylog REST API (rest_enable_tls)
- Enable TLS for the web interface endpoint (web_enable_tls)

You also need to make sure that you have proper certificates in place, which are valid and trusted by the clients. Not enabling TLS for either one of them will result in a browser error about mixed content and the web interface will cease to work.

**Note:** If you’re operating a single-node setup and would like to use HTTPS for the Graylog web interface and the Graylog REST API, it’s possible to use NGINX or Apache as a reverse proxy.
Certificate/Key file format

When you are configuring TLS, you need to make sure that your certificate/key files are in the right format, which is X.509 for certificates and PKCS#8 for the private keys. Both must to be stored in PEM format.

Creating a self-signed private key/certificate

Create a file named openssl-graylog.cnf with the following content (customized to your needs):

```
[req]
distinguished_name = req_distinguished_name
x509_extensions = v3_req
prompt = no

# Details about the issuer of the certificate
[req_distinguished_name]
C = US
ST = Some-State
L = Some-City
O = My Company
OU = My Division
CN = graylog.example.com

[v3_req]
keyUsage = keyEncipherment, dataEncipherment
extendedKeyUsage = serverAuth
subjectAltName = @alt_names

# IP addresses and DNS names the certificate should include
# Use IP.### for IP addresses and DNS.### for DNS names,
# with "###" being a consecutive number.
[alt_names]
IP.1 = 203.0.113.42
DNS.1 = graylog.example.com
```

Create PKCS#5 private key and X.509 certificate:

```
$ openssl version
OpenSSL 0.9.8zh 14 Jan 2016
$ openssl req -x509 -days 365 -nodes -newkey rsa:2048 -config openssl-graylog.cnf -keyout pkcs5-plain.pem -out cert.pem
Generating a 2048 bit RSA private key
............................+++.
++
writing new private key to 'pkcs5-plain.pem'

# IP addresses and DNS names the certificate should include
# Use IP.### for IP addresses and DNS.### for DNS names,
# with "###" being a consecutive number.

Create PKCS#5 private key into an unencrypted PKCS#8 private key:

```
$ openssl pkcs8 -in pkcs5-plain.pem -topk8 -nocrypt -out pkcs8-plain.pem
```

Convert PKCS#5 private key into an encrypted PKCS#8 private key (using the passphrase secret):

```
$ openssl pkcs8 -in pkcs5-plain.pem -topk8 -out pkcs8-encrypted.pem -passout pass:secret
```

Converting a PKCS #12 (PFX) file to private key and certificate pair

PKCS #12 key stores (PFX files) are commonly used on Microsoft Windows.
In this example, the PKCS #12 (PFX) file is named `keystore.pfx`:

```
$ openssl pkcs12 -in keystore.pfx -nokeys -out graylog-certificate.pem
$ openssl pkcs12 -in keystore.pfx -nocerts -out graylog-pkcs5.pem
$ openssl pkcs8 -in graylog-pkcs5.pem -topk8 -out graylog-key.pem
```

The resulting `graylog-certificate.pem` and `graylog-key.pem` can be used in the Graylog configuration file.

**Converting an existing Java Keystore to private key/certificate pair**

This section describes how to export a private key and certificate from an existing Java KeyStore in JKS format.

The starting point is an existing Java KeyStore in JKS format which contains a private key and certificate which should be used in Graylog:

```
$ keytool -list -v -keystore keystore.jks -alias graylog.example.com
Enter keystore password:
Alias name: graylog.example.com
Creation date: May 10, 2016
Entry type: PrivateKeyEntry
Certificate chain length: 1
Certificate[1]:
  Owner: CN=graylog.example.com, OU=Unknown, O="Graylog, Inc.", L=Hamburg, ST=Hamburg, C=DE
  Issuer: CN=graylog.example.com, OU=Unknown, O="Graylog, Inc.", L=Hamburg, ST=Hamburg, C=DE
  Serial number: 2b33832d
  Valid from: Tue May 10 10:02:34 CEST 2016 until: Mon Aug 08 10:02:34 CEST 2016
  Certificate fingerprints:
  Signature algorithm name: SHA256withRSA
  Version: 3

Extensions:
#1: ObjectId: 2.5.29.14 Criticality=false
  SubjectKeyIdentifier [ KeyIdentifier [ 0000: AC 79 64 9F A1 60 14 B9 51 F4 F5 0B B3 B5 02 A5 .yd..`..Q....... 0010: B8 07 DC 7B ] ]

The Java KeyStore in JKS format has to be converted to a PKCS#12 keystore, so that OpenSSL can work with it:

```
$ keytool -importkeystore -srckeystore keystore.jks -destkeystore keystore.p12 -deststoretype PKCS12
Enter destination keystore password:
Re-enter new password:
Enter source keystore password:
Entry for alias graylog.example.com successfully imported.
Import command completed: 1 entries successfully imported, 0 entries failed or cancelled
```

After the keystore has been successfully converted into PKCS#12 format, OpenSSL can export the X.509 certificate with PEM encoding:

```
$ openssl pkcs12 -in keystore.p12 -nokeys -out graylog-certificate.pem
Enter Import Password:
MAC verified OK
```

5.4. Using HTTPS
The private key can only be exported in PKCS#5 format with PEM encoding:

```bash
$ openssl pkcs12 -in keystore.p12 -nocerts -out graylog-pkcs5.pem
Enter Import Password:
MAC verified OK
Enter PEM pass phrase:
Verifying - Enter PEM pass phrase:

Graylog currently only supports PKCS#8 private keys with PEM encoding, so OpenSSL has to convert it into the correct format:

```bash
$ openssl pkcs8 -in graylog-pkcs5.pem -topk8 -out graylog-key.pem
Enter pass phrase for graylog-pkcs5.pem:
Enter Encryption Password:
Verifying - Enter Encryption Password:
```

The working directory should now contain the PKCS#8 private key (graylog-key.pem) and the X.509 certificate (graylog-certificate.pem) to be used with Graylog:

```bash
$ head graylog-key.pem graylog-certificate.pem
===> graylog-key.pem <==
-----BEGIN ENCRYPTED PRIVATE KEY-----
MIIE6TAbBgkqhkiG9w0BBMQuMdwQ1wMhLa5bw9vgCAaggABIIEyN42AeYJNBEiqhI
mWqJDot4Jokw2vB4abcIJSDo4+71tjMrecVRCDsV8ZzkXjnbumBHEoxexe5f0/z
wgqgf/UDyTM3uKYQGTG91fcqTyMDU103wBQsQsNehOAQzA7hMehqgNJOH0Zfny
EFxvKXh67j4eA9vLVrUp86dbm4wp3o0pjm0LduXiegCvtm5fd7XfL5crFS8k8S
bSFH4y8bxDxLaNjMkKj19gPCACDr8909Gk/Z9q9N8zkj+xG9x1NRM51Srz02E0
nyyK7b3t9F3J35zjroB2HfiWyrPQ1u4FyGoroGvu0L3eWosjBldRseO8gjCM5P
ic9zVsVmh6/4CPKJqB87vP4QhpYcr9jJJsbn6zg40EiLpM00VIv0yU9tqTuRR
TDYAI3MNZLzRrV52Esh32O21WHM7r187x4W3prDFnjkXf02DrFbgCswkEQnb3
vj86q13RHzqoXT4Wzugvvc2/NBLMV0HNQBAfEK3X1YBmtQPvEJhwSxeszA17CpU

===> graylog-certificate.pem <==
Bag Attributes
    friendlyName: graylog.example.com
    localKeyId: 54 69 65 20 31 34 36 32 38 32 38 37 38 32 33 30 39 32
    subject=/C=DE/ST=Hamburg/L=Hamburg/O=Graylog, Inc./OU=Unknown/CN=graylog.example.com
    issuer=/C=DE/ST=Hamburg/L=Hamburg/O=Graylog, Inc./OU=Unknown/CN=graylog.example.com
-----BEGIN CERTIFICATE-----
```

The resulting PKCS#8 private key (graylog-key.pem) and the X.509 certificate (graylog-certificate.pem) can now be used to enable encrypted connections with Graylog by enabling TLS for the Graylog REST API and the web interface in the Graylog configuration file:

```bash
# Enable HTTPS support for the REST API. This secures the communication with the REST API
# using TLS to prevent request forgery and eavesdropping.
rest_enable_tls = true

# The X.509 certificate chain file in PEM format to use for securing the REST API.
rest_tls_cert_file = /path/to/graylog-certificate.pem

# The PKCS#8 private key file in PEM format to use for securing the REST API.
rest_tls_key_file = /path/to/graylog-key.pem

# The password to unlock the private key used for securing the REST API.
```

90 Chapter 5. Configuring Graylog
rest_tls_key_password = secret

# Enable HTTPS support for the web interface. This secures the communication the web interface
# using TLS to prevent request forgery and eavesdropping.
web_enable_tls = true

# The X.509 certificate chain file in PEM format to use for securing the web interface.
web_tls_cert_file = /path/to/graylog-certificate.pem

# The PKCS#8 private key file in PEM format to use for securing the web interface.
web_tls_key_file = /path/to/graylog-key.pem

# The password to unlock the private key used for securing the web interface.
web_tls_key_password = secret

### Sample files

This section shows the difference between following private key formats with samples.

**PKCS#5 plain private key:**

```
-----BEGIN RSA PRIVATE KEY-----
MIIBowIBAAJBANxtmQ1Kccdp7HNtBzqTal48Vv617bj4SnhkcmMN99sCQ2Naj/sp
[...]
NiCYNLiCawBbp2nYw/ztPVACK4EwOpUy+u19cMB0JA==
-----END RSA PRIVATE KEY-----
```

**PKCS#8 plain private key:**

```
-----BEGIN PRIVATE KEY-----
MIIBVAIBADANBgkqhkiG9w0BAQEFAASCAT4wggE6AgEAAkEA6GZN0rQFKRIVaPOz
[...]
LaLGd9G63kLg85eldSy55uIAxsvqQLgfSYaliVtSbAgyx1Yfs3hJ+CTpNKzTNv/
Fx80EltYV6k=
-----END PRIVATE KEY-----
```

**PKCS#5 encrypted private key:**

```
-----BEGIN RSA PRIVATE KEY-----
Proc-Type: 4,ENCRYPTED
DEK-Info: DES-EDE3-CBC,E83B4019057F55E9

iIPs59Qn4Rsd7ppch9/vNE7PfRSHLoQFmaAjaF0DxjV9oucznUjJq2ghABZ2E2H
[...]
y5IT1M2PgN3LNkVSSLpWko0BuQdfu0JTKc7NPyRc=
-----END RSA PRIVATE KEY-----
```

**PKCS#8 encrypted private key:**

```
-----BEGIN ENCRYPTED PRIVATE KEY-----
MIIBpjbABgkqhkiG9w0BBQowMzAbBgkqhkiG9w0BBQwwDgQIUI9Y9p2EfWucCAggA
[...]
Ijs2NZp6zmlqf/RXnETsJjGd0TRWaEdu+x0OyVyPskX2177X9DUJoD31
-----END ENCRYPTED PRIVATE KEY-----
```
Adding a self-signed certificate to the JVM trust store

Graylog nodes inside a cluster need to communicate with each other using the Graylog REST API. When using HTTPS for the Graylog REST API, the X.509 certificate must be trusted by the JVM trust store (similar to the trusted CA bundle in an operating system), otherwise communication will fail.

**Important:** If you are using different X.509 certificates for each Graylog node, you have to add all of them into the JVM trust store of each Graylog node.

The default trust store of an installed Java runtime environment can be found at $JAVA_HOME/jre/lib/security/cacerts. In order not to “pollute” the official trust store, we make a copy of it which we will use with Graylog instead:

```bash
$ cp -a "$JAVA_HOME/jre/lib/security/cacerts" /path/to/cacerts.jks
```

After the original key store file has been copied, we can add the self-signed certificate (cert.pem, see Creating a self-signed private key/certificate) to the key store (the default password is changeit):

```bash
$ keytool -importcert -keystore /path/to/cacerts.jks -storepass changeit -alias graylog-self-signed -file cert.pem
```

Owner: CN=graylog.example.com, O="Graylog, Inc.", L=Hamburg, ST=Hamburg, C=DE
Issuer: CN=graylog.example.com, O="Graylog, Inc.", L=Hamburg, ST=Hamburg, C=DE
Serial number: 8c80134cee556734
Certificate fingerprints:
  SHA1: F0:64:D0:1B:3B:6B:C8:01:D5:4D:33:36:87:F0:FB:10:E1:36:21:9E
Signature algorithm name: SHA1withRSA
Version: 3

Extensions:

[Other details about the certificate...]

Trust this certificate? [no]: yes
Certificate was added to keystore

To verify that the self-signed certificate has indeed been added, it can be listed with the following command:

```bash
$ keytool -keystore /path/to/cacerts.jks -storepass changeit -list | grep graylog-self-signed -A1
```

Graylog-self-signed, Jun 14, 2016, trustedCertEntry,

The printed certificate fingerprint (SHA1) should match the one printed when importing the self-signed certificate.

In order for the JVM to pick up the new trust store, it has to be started with the JVM parameter -Djavax.net.ssl.trustStore=/path/to/cacerts.jks. If you’ve been using another password to encrypt the JVM trust store than the default changeit, you additionally have to set the JVM parameter -Djavax.net.ssl.trustStorePassword=secret.

Most start and init scripts for Graylog provide a JAVA_OPTS variable which can be used to pass the javax.net.ssl.trustStore and (optionally) javax.net.ssl.trustStorePassword system properties.

Disabling specific TLS ciphers and algorithms

Since Java 7u76 it is possible to disable specific TLS algorithms and ciphers for secure connections.
In order to disable specific TLS algorithms and ciphers, you need to provide a properties file with a list of disabled algorithms and ciphers. Take a look at the example security.properties in the Graylog source repository.

For example, if you want to disable all algorithms except for TLS 1.2, the properties file has to contain the following line:

```
jdk.tls.disabledAlgorithms=SSLv2Hello, SSLv3, TLSv1, TLSv1.1
```

If additionally you want to disable DSA/RSA key sizes lower than 2048 bits and EC key sizes lower than 160 bits, the properties file has to contain the following line:

```
jdk.tls.disabledAlgorithms=SSLv2Hello, SSLv3, TLSv1, TLSv1.1, EC keySize < 160, RSA keySize < 2048, DSA keySize < 2048
```

To load the properties file into a JVM, you have to pass it to Java using the `java.security.properties` system property:

```
java -Djava.security.properties=/path/to/security.properties -jar /path/to/graylog.jar server
```

Most start and init scripts for Graylog provide a `JAVA_OPTS` variable which can be used to pass the `java.security.properties` system property.

### Further reading

- [https://docs.oracle.com/javase/8/docs/technotes/guides/security/jsse/JSSERefGuide.html#DisabledAlgorithms](https://docs.oracle.com/javase/8/docs/technotes/guides/security/jsse/JSSERefGuide.html#DisabledAlgorithms)
- [http://www.oracle.com/technetwork/java/javase/7u76-relnotes-2389087.html](http://www.oracle.com/technetwork/java/javase/7u76-relnotes-2389087.html)

### Multi-node Setup

This guide doesn’t provide a step-by-step tutorial for building a multi-node Graylog cluster but does simply give some advice for questions that might arise during the setup.

It’s important for such a project that you understand each step in the setup process and do some planning upfront. Without a proper roadmap of all the things you want to achieve with a Graylog cluster, you will be lost on the way.

Graylog should be the last component you install in this setup. Its dependencies, namely MongoDB and Elasticsearch, have to be up and running first.

**Important:** This guide doesn’t include instructions for running a multi-node Graylog cluster in an untrusted network. We assume that the connection between the hosts is trusted and doesn’t have to be secured individually.

### Prerequisites

Every server which is part of this setup should have the software requirements installed to run the targeted software. All software requirements can be found in the installation manual.

We highly recommend that the system time on all systems is kept in sync via NTP or a similar mechanism. Needless to say that DNS resolution must be working, too. Because everything is a freaking DNS problem.

In order to simplify the installation process, the servers should have a working Internet connection.
**MongoDB replica set**

We recommend to deploy a MongoDB replica set.

MongoDB doesn’t have to run on dedicated servers for the workload generated by Graylog, but you should follow the recommendations given in the MongoDB documentation about architecture. Most important is that you have an odd number of MongoDB servers in the replica set.

In most setups, each Graylog server will also host an instance of MongoDB which is part of the same replica set and shares the data with all other nodes in the cluster.

**Note:** To avoid unauthorized access to your MongoDB database, the MongoDB replica set should be setup with authentication.

The correct order of working steps should be as follows:

1. Create the replica set (`rs01`)
2. Create the database (`graylog`)
3. Create a user account for accessing the database, which has the roles `readWrite` and `dbAdmin`.

If your MongoDB needs to be reachable over network you should set the IP with `bind_ip` in the configuration.

**Elasticsearch cluster**

The Elasticsearch setup documentation should help you to install Elasticsearch with a robust base configuration.

It is important to name the Elasticsearch cluster not simply named `elasticsearch` to avoid accidental conflicts with Elasticsearch nodes using the default configuration. Just choose anything else (we recommend `graylog`), because this is the default name and any Elasticsearch instance that is started in the same network will try to connect to this cluster.

The Elasticsearch servers need one IP that can be reached over network set in `network.host` and some participants of the cluster in `discovery.zen.ping.unicast.hosts`. That is enough to have a minimal cluster setup.

**Note:** Graylog currently doesn’t work with Elasticsearch clusters using the Shield plugin. Your Elasticsearch cluster need to be secured by design (e.g. segregating network access).

**Graylog Multi-node**

After the installation of Graylog, you should take care that only one Graylog node is configured to be master with the configuration setting `is_master = true`.

The URI configured in `rest_listen_uri` (or `rest_transport_uri`) must be accessible for all Graylog nodes of the cluster.

**Graylog to MongoDB connection**

The `mongodb_uri` configuration setting must include all MongoDB nodes forming the replica set, the name of the replica set, as well as the previously configured user account with access to the replica set. The configuration setting is a normal MongoDB connection string.

Finally, the MongoDB connection string in the Graylog configuration file should look like this:
mongodb_uri = mongodb://USERNAME:PASSWORD@mongodb-node01:27017,mongodb-node02:27017,mongodb-node03:27017/graylog?replicaSet=rs01

Graylog to Elasticsearch connection

Graylog will connect as client node to the Elasticsearch Cluster.

To avoid issues with the connection to the Elasticsearch cluster you should add some of the network addresses of the Elasticsearch nodes to elasticsearch_discovery_zen_ping_unicast_hosts.

Additionally, elasticsearch_network_host must be set to a network interface which can be accessed by the other Elasticsearch nodes in the Elasticsearch cluster.

Graylog web interface

By default, the web interface can be used on every instance of Graylog which hasn’t disabled it with the configuration setting web_enable = false.

It’s possible to use a loadbalancer in front of all Graylog servers, please refer to Making the web interface work with load balancers/proxies for more details.

Depending on your setup, it’s possible to either use a hardware loadbalancer for TLS/HTTPS termination, a reverse proxy, or to simply enable it in the Graylog node.

Scaling

Each component in this multi-node setup can be scaled on the individual needs.

Depending on the amount of messages ingested and how long messages should be available for direct search, the Elasticsearch cluster will need most of the resources on your setup.

Keep an eye on your Elasticsearch cluster with plugins like Elastic HQ or Kopf. Those will help you to understand the Elasticsearch cluster health and behavior.

Graylog Metrics should be monitored with the Graylog Metrics Reporter plugins which are able to send the internal Graylog metrics to your favorite metrics collector (e. g. Graphite or Prometheus).

Up until today, we have almost never faced the issue that the MongoDB replica set needed special attention. But of course you should still monitor it and store its metrics - just to be sure.

Troubleshooting

- After every configuration change or service restart, watch the log file of the applications you have worked on. Sometimes other log files can also give you hints about what went wrong. For example if you’re configuring Graylog and try to find out why the connection to the MongoDB isn’t working, the MongoDB logs can help to identify the problem.

- If HTTPS has been enabled for the Graylog REST API, it need to be setup for the Graylog web interface, too.

Elasticsearch

We strongly recommend to use a dedicated Elasticsearch cluster for your Graylog setup.

If you are using a shared Elasticsearch setup, a problem with indices unrelated to Graylog might turn the cluster status to YELLOW or RED and impact the availability and performance of your Graylog setup.
Important: Graylog currently does not work with Elasticsearch clusters using the License or Shield plugin.

Elasticsearch versions

Graylog hosts an embedded Elasticsearch node which is joining the Elasticsearch cluster as a client node.

The following table provides an overview over the Elasticsearch version in Graylog:

<table>
<thead>
<tr>
<th>Graylog version</th>
<th>Elasticsearch version</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.2.0-1.2.1</td>
<td>1.7.1</td>
</tr>
<tr>
<td>1.3.0-1.3.3</td>
<td>1.7.3</td>
</tr>
<tr>
<td>1.3.4</td>
<td>1.7.5</td>
</tr>
<tr>
<td>2.0.0</td>
<td>2.3.1</td>
</tr>
<tr>
<td>2.0.1-2.0.3</td>
<td>2.3.2</td>
</tr>
<tr>
<td>2.1.0-2.1.3</td>
<td>2.3.5</td>
</tr>
<tr>
<td>2.2.0-2.2.3</td>
<td>2.4.4</td>
</tr>
</tbody>
</table>

Caution: Graylog 2.x does not work with Elasticsearch 5.x!

Configuration

Graylog

The most important settings to make a successful connection are the Elasticsearch cluster name, one or more addresses of Elasticsearch master nodes, and the local network bind address.

Graylog needs to know the address of at least one other Elasticsearch master node given in the elasticsearch_discovery_zen_ping_unicast_hosts setting. Vice versa, the Elasticsearch nodes need to be able to access the embedded Elasticsearch node in Graylog via the interface given in the elasticsearch_network_host setting.

Cluster Name

You need to tell Graylog which Elasticsearch cluster to join. The Elasticsearch default cluster name is elasticsearch and configured for every Elasticsearch node in the elasticsearch.yml configuration file with the cluster.name name.

Configure the same cluster name in every Graylog configuration file (e.g. graylog.conf) with the elasticsearch_cluster_name setting (default: graylog).

We recommend to call the cluster graylog-production or graylog, but not elasticsearch to prevent accidental cluster name collisions.

The Elasticsearch configuration file is typically located at /etc/elasticsearch/elasticsearch.yml.

Network setup

Graylog is using unicast discovery to find all the Elasticsearch nodes in the cluster.
In order for this to work, Graylog has to know some master nodes of the Elasticsearch cluster which can be provided in the `elasticsearch_discovery_zen_ping_unicast_hosts` configuration setting.

For example, add the following lines to your Graylog configuration file for an Elasticsearch cluster which includes the 2 Elasticsearch master nodes `es-node-1.example.org` and `es-node-2.example.org`:

```plaintext
# List of Elasticsearch master nodes to connect to
elasticsearch_discovery_zen_ping_unicast_hosts = es-node-1.example.org:9300,es-node-2.example.org:9300
```

Additionally, Graylog has to use a network interface for the embedded Elasticsearch node which the other Elasticsearch nodes in the cluster can connect to:

```plaintext
# Public IP address or host name of the Graylog node, accessible for the other Elasticsearch nodes
elasticsearch_network_host = 198.51.100.23
```

Also make sure to configure Zen unicast discovery in the Elasticsearch configuration file by adding the `discovery.zen.ping.multicast.enabled` and `discovery.zen.ping.unicast.hosts` settings with the list of Elasticsearch nodes to `elasticsearch.yml`:

```plaintext
discovery.zen.ping.multicast.enabled: false
discovery.zen.ping.unicast.hosts: ["es-node-1.example.org:9300", "es-node-2.example.org:9300"]
```

The Elasticsearch default communication port is `9300/tcp` (not to be confused with the HTTP interface running on port `9200/tcp` by default).

The communication port can be changed in the Elasticsearch configuration file (`elasticsearch.yml`) with the configuration setting `transport.tcp.port`.

Last but not least, make sure that Elasticsearch is binding to a network interface that Graylog can connect to (see `network.host` and Commonly Used Network Settings).

### Configuration of Elasticsearch nodes

**Disable dynamic scripting**

Elasticsearch prior to version 1.2 had an insecure default configuration which could lead to a remote code execution. Make sure to add the following settings to the `elasticsearch.yml` file to disable the dynamic scripting feature and prevent possible remote code executions:

```plaintext
script.inline: false
script.indexed: false
script.file: false
```

Details about dynamic scripting can be found in the reference documentation of Elasticsearch.

**Control access to Elasticsearch ports**

Since Elasticsearch has no authentication mechanism at time of this writing, make sure to restrict access to the Elasticsearch ports (default: `9200/tcp` and `9300/tcp`). Otherwise the data is readable by anyone who has access to the machine over network.

**Open file limits**

Because Elasticsearch has to keep a lot of files open simultaneously it requires a higher open file limit that the usual operating system defaults allow. **Set it to at least 64000 open file descriptors.**
Graylog will show a notification in the web interface when there is a node in the Elasticsearch cluster which has a too low open file limit. Read about how to raise the open file limit in the corresponding Elasticsearch documentation page.

**Heap size**

It is strongly recommended to raise the standard size of heap memory allocated to Elasticsearch. Just set the `ES_HEAP_SIZE` environment variable to for example `24g` to allocate 24GB. We recommend to use around 50% of the available system memory for Elasticsearch (when running on a dedicated host) to leave enough space for the system caches that Elasticsearch uses a lot. But please take care that you don’t cross 32 GB!

**Merge throttling**

Elasticsearch is throttling the merging of Lucene segments to allow extremely fast searches. This throttling however has default values that are very conservative and can lead to slow ingestion rates when used with Graylog. You would see the message journal growing without a real indication of CPU or memory stress on the Elasticsearch nodes. It usually goes along with Elasticsearch INFO log messages like this:

```
now throttling indexing
```

When running on fast IO like SSDs or a SAN we recommend to increase the value of the `indices.store.throttle.max_bytes_per_sec` in your `elasticsearch.yml` to 150MB:

```
indices.store.throttle.max_bytes_per_sec: 150mb
```

Play around with this setting until you reach the best performance.

**Tuning Elasticsearch**

Graylog is already setting specific configuration for every index it is managing. This is enough tuning for a lot of use cases and setups.

A more detailed guide about tuning Elasticsearch will be published at a later time.

**Avoiding split-brain and shard shuffling**

**Split-brain events**

Elasticsearch sacrifices consistency in order to ensure availability, and partition tolerance. The reasoning behind that is that short periods of misbehaviour are less problematic than short periods of unavailability. In other words, when Elasticsearch nodes in a cluster are unable to replicate changes to data, they will keep serving applications such as Graylog. When the nodes are able to replicate their data, they will attempt to converge the replicas and to achieve eventual consistency.

Elasticsearch tackles the previous by electing master nodes, which are in charge of database operations such as creating new indices, moving shards around the cluster nodes, and so forth. Master nodes coordinate their actions actively with others, ensuring that the data can be converged by non-masters. The cluster nodes that are not master nodes are not allowed to make changes that would break the cluster.

The previous mechanism can in some circumstances fail, causing a **split-brain event**. When an Elasticsearch cluster is split into two sides, both thinking they are the master, data consistency is lost as the masters work independently on the data. As a result the nodes will respond differently to same queries. This is considered a catastrophic event,
because the data from two masters can not be rejoined automatically, and it takes quite a bit of manual work to remedy the situation.

Avoiding split-brain events

Elasticsearch nodes take a simple majority vote over who is master. If the majority agrees that they are the master, then most likely the disconnected minority has also come to conclusion that they can not be the master, and everything is just fine. This mechanism requires at least 3 nodes to work reliably however, because one or two nodes can not form a majority.

The minimum amount of master nodes required to elect a master must be configured manually in `elasticsearch.yml`

```
# At least NODES/2+1 on clusters with NODES > 2, where NODES is the number of master nodes in the cluster
discovery.zen.minimum_master_nodes: 2
```

The configuration values should typically for example:

<table>
<thead>
<tr>
<th>Master nodes</th>
<th>minimum_master_nodes</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>With 2 the other node going down would stop the cluster from working!</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>

Some of the master nodes may be dedicated master nodes, meaning they are configured just to handle lightweight operational (cluster management) responsibilities. They will not handle or store any of the cluster’s data. The function of such nodes is similar to so called witness servers on other database products, and setting them up on dedicated witness sites will greatly reduce the chance of Elasticsearch cluster instability.

A dedicated master node has the following configuration in `elasticsearch.yml`:

```
node.data: false
node.master: true
```

Shard shuffling

When cluster status changes, for example because of node restarts or availability issues, Elasticsearch will start automatically rebalancing the data in the cluster. The cluster works on making sure that the amount of shards and replicas will conform to the cluster configuration. This is a problem if the status changes are just temporary. Moving shards and replicas around in the cluster takes considerable amount of resources, and should be done only when necessary.

Avoiding unnecessary shuffling

Elasticsearch has couple configuration options, which are designed to allow short times of unavailability before starting the recovery process with shard shuffling. There are 3 settings that may be configured in `elasticsearch.yml`:

```
# Recover only after the given number of nodes have joined the cluster. Can be seen as "minimum number of nodes to be reachable before automatically trying to recover"
gateway.recover_after_nodes: 8

# Time to wait for additional nodes after recover_after_nodes is met.
gateway.recover_after_time: 5m
```

5.6. Elasticsearch
# Inform Elasticsearch how many nodes form a full cluster. If this number is met, start up immediately.

```
# gateway.expected_nodes: 10
```

The configuration options should be set up so that only _minimal_ node unavailability is tolerated. For example server restarts are common, and should be done in managed manner. The logic is that if you lose large part of your cluster, you probably should start re-shuffling the shards and replicas without tolerating the situation.

## Custom index mappings

Sometimes it’s useful to not rely on Elasticsearch’s dynamic mapping but to define a stricter schema for messages.

**Note:** If the index mapping is conflicting with the actual message to be sent to Elasticsearch, indexing that message will fail.

Graylog itself is using a default mapping which includes settings for the _timestamp, message, full_message, and source_ fields of indexed messages:

```
$ curl -X GET 'http://localhost:9200/_template/graylog-internal?pretty'
{
   "graylog-internal" : {
      "order" : -2147483648,
      "template" : "graylog_\*",
      "settings" : { },
      "mappings" : {
         "message" : {
            "_ttl" : {
               "enabled" : true
            },
            "_source" : {
               "enabled" : true
            },
            "dynamic_templates" : [ {
               "internal_fields" : {
                  "mapping" : {
                     "index" : "not_analyzed",
                     "type" : "string"
                  },
                  "match" : "gl2_\*"
               }
            }, {
               "store_generic" : {
                  "mapping" : {
                     "index" : "not_analyzed"
                  },
                  "match" : "\*"
               }
            }],
            "properties" : {
               "full_message" : {
                  "analyzer" : "standard",
                  "index" : "analyzed",
                  "type" : "string"
               },
               "streams" : {
                  "index" : "not_analyzed"
               }
            }
         }
      }
   }
}
```
In order to extend the default mapping of Elasticsearch and Graylog, you can create one or more custom index mappings and add them as index templates to Elasticsearch.

Let's say we have a schema for our data like the following:

<table>
<thead>
<tr>
<th>Field Name</th>
<th>Field Type</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>http_method</td>
<td>string</td>
<td>GET</td>
</tr>
<tr>
<td>http_response_code</td>
<td>long</td>
<td>200</td>
</tr>
<tr>
<td>ingest_time</td>
<td>date</td>
<td>2016-06-13T15:00:51.927Z</td>
</tr>
<tr>
<td>took_ms</td>
<td>long</td>
<td>56</td>
</tr>
</tbody>
</table>

This would translate to the following additional index mapping in Elasticsearch:

```json
"mappings" : {
    "message" : {
        "properties" : {
            "http_method" : {
                "type" : "string",
                "index" : "not_analyzed"
            },
            "http_response_code" : {
                "type" : "long"
            },
            "ingest_time" : {
                "type" : "date",
                "format": "strict_date_time"
            },
            "took_ms" : {
                "type" : "long"
            }
        }
    }
}
```

The format of the `ingest_time` field is described in the Elasticsearch documentation about the `format_mapping` parameter. Also make sure to check the Elasticsearch documentation about Field datatypes.
In order to apply the additional index mapping when Graylog creates a new index in Elasticsearch, it has to be added to an index template. The Graylog default template (graylog-internal) has the lowest priority and will be merged with the custom index template by Elasticsearch.

**Warning:** If the default index mapping and the custom index mapping cannot be merged (e.g., because of conflicting field datatypes), Elasticsearch will throw an exception and won’t create the index. So be extremely cautious and conservative about the custom index mappings!

### Creating a new index template

Save the following index template for the custom index mapping into a file named `graylog-custom-mapping.json`:

```json
{
    "template": "graylog_*",
    "mappings": {
        "message": {
            "properties": {
                "http_method": {
                    "type": "string",
                    "index": "not_analyzed"
                },
                "http_response_code": {
                    "type": "long"
                },
                "ingest_time": {
                    "type": "date",
                    "format": "strict_date_time"
                },
                "took_ms": {
                    "type": "long"
                }
            }
        }
    }
}
```

Finally, load the index mapping into Elasticsearch with the following command:

```bash
{
    "acknowledged": true
}
```

Every Elasticsearch index created from that time on, will have an index mapping consisting of the original graylog-internal index template and the new graylog-custom-mapping template:

```bash
$ curl -X GET 'http://localhost:9200/graylog_deflector/_mapping?pretty'
{
    "graylog_2": {
        "mappings": {
            "message": {
                "_ttl": {
                    "enabled": true
                },
                "dynamic_templates": [ {
                    "internal_fields": {
```
"mapping" : {
   "index" : "not_analyzed",
   "type" : "string"
},
"match" : "gl2_*"
}, {
   "store_generic" : {
      "mapping" : {
         "index" : "not_analyzed"
      },
      "match" : "*"
   }
}, {
   "properties" : {
      "full_message" : {
         "type" : "string",
         "analyzer" : "standard"
      },
      "http_method" : {
         "type" : "string",
         "index" : "not_analyzed"
      },
      "http_response_code" : {
         "type" : "long"
      },
      "ingest_time" : {
         "type" : "date",
         "format" : "strict_date_time"
      },
      "message" : {
         "type" : "string",
         "analyzer" : "standard"
      },
      "source" : {
         "type" : "string",
         "analyzer" : "analyzer_keyword"
      },
      "streams" : {
         "type" : "string",
         "index" : "not_analyzed"
      },
      "timestamp" : {
         "type" : "date",
         "format" : "yyyy-MM-dd HH:mm:ss.SSS"
      },
      "took_ms" : {
         "type" : "long"
      }
   }
}

Note: When using different index sets every index set can have its own mapping.
Deleting custom index templates

If you want to remove an existing index template from Elasticsearch, simply issue a DELETE request to Elasticsearch:

```bash
$ curl -X DELETE 'http://localhost:9200/_template/graylog-custom-mapping?pretty'
{
   "acknowledged" : true
}
```

After you’ve removed the index template, new indices will only have the original index mapping:

```bash
$ curl -X GET 'http://localhost:9200/graylog_deflector/_mapping?pretty'
{
   "graylog_3" : {
      "mappings" : {
         "message" : {
            "_ttl" : {
               "enabled" : true
            },
            "dynamic_templates" : [ {
               "internal_fields" : {
                  "mapping" : {
                     "index" : "not_analyzed",
                     "type" : "string"
                  },
                  "match" : "gl2_*"
               }
            }, {
               "store_generic" : {
                  "mapping" : {
                     "index" : "not_analyzed"
                  },
                  "match" : "*"
               }
            }]
         },
         "properties" : {
            "full_message" : {
               "type" : "string",
               "analyzer" : "standard"
            },
            "message" : {
               "type" : "string",
               "analyzer" : "standard"
            },
            "source" : {
               "type" : "string",
               "analyzer" : "analyzer_keyword"
            },
            "streams" : {
               "type" : "string",
               "index" : "not_analyzed"
            },
            "timestamp" : {
               "type" : "date",
               "format" : "yyyy-MM-dd HH:mm:ss.SSS"
            }
         }
      }
   }
}
```
Cluster Status explained

Elasticsearch provides a classification for the cluster health.
The cluster status applies to different levels:
- **Shard level** - see status descriptions below
- **Index level** - inherits the status of the worst shard status
- **Cluster level** - inherits the status of the worst index status

That means that the Elasticsearch cluster status can turn red if a single index or shard has problems even though the rest of the indices/shards are okay.

**Note:** Graylog checks the status of the current write index while indexing messages. If that one is GREEN or YELLOW, Graylog will continue to write messages into Elasticsearch regardless of the overall cluster status.

Explanation of the different status levels:

**RED**

The RED status indicates that some or all of the primary shards are not available.

In this state, no searches can be performed until all primary shards have been restored.

**YELLOW**

The YELLOW status means that all of the primary shards are available but some or all shard replicas are not.

With only one Elasticsearch node, the cluster state cannot become green because shard replicas cannot be assigned.

In most cases, this can be solved by adding another Elasticsearch node to the cluster or by reducing the replication factor of the indices (which means less resiliency against node outages, though).

**GREEN**

The cluster is fully operational. All primary and replica shards are available.

Index model

Overview

Graylog is transparently managing one or more sets of Elasticsearch indices to optimize search and analysis operations for speed and low resource consumption.

To enable managing indices with different mappings, analyzers, and replication settings Graylog is using so-called index sets which are an abstraction of all these settings.
Each index set contains the necessary settings for Graylog to create, manage, and fill Elasticsearch indices and handle index rotation and data retention for specific requirements.

Graylog is maintaining an index alias per index set which is always pointing to the current write-active index from that
index set. There is always exactly one index to which new messages are written until the configured rotation criterion (number of documents, index size, or index age) has been met.

A background task continuously checks if the rotation criterion of an index set has been met and a new index is created and prepared when that happens. Once the index is ready, the index alias is atomically switched to it. That means that all Graylog nodes can write messages to the into alias without even knowing what the currently write-active index of the index set is.

Write Path

Almost every read operation is performed with a given time range. Because Graylog is writing messages sequentially into Elasticsearch it can keep information about the time range each index covers. It selects a lists of indices to query when having a time range provided. If no time range was provided, it will search in all indices it knows.
Eviction of indices and messages

There are configuration settings for the maximum number of indices Graylog is managing in a given index set. Depending on the configured retention strategy, the oldest indices of an index set will automatically be closed, deleted, or exported when the configured maximum number of indices has been reached.

The deletion is performed by the Graylog master node in a background thread which is continuously comparing the number of indices with the configured maximum:

```
INFO : org.graylog2.indexer.rotation.strategies.AbstractRotationStrategy - Deflector index <graylog_95> should be rotated, Pointing deflector to new index now!
INFO : org.graylog2.indexer.MongoIndexSet - Cycling from <graylog_95> to <graylog_96>.
INFO : org.graylog2.indexer.MongoIndexSet - Index <graylog_96> has been successfully allocated.
INFO : org.graylog2.indexer.MongoIndexSet - Pointing index alias <graylog_deflector> to new index <graylog_96>.
INFO : org.graylog2.system.jobs.SystemJobManager - Submitted SystemJob <f1018ae0-dcaa-11e6-97c3-6c4008b8fc28> [org.graylog2.indexer.indices.jobs.SetIndexReadOnlyAndCalculateRangeJob]
INFO : org.graylog2.indexer.MongoIndexSet - Successfully pointed index alias <graylog_deflector> to index <graylog_96>.
```

Index Set Configuration

Index sets have a variety of different settings related to how Graylog will store messages into the Elasticsearch cluster.
• **Title**: A descriptive name of the index set.

• **Description**: A description of the index set for human consumption.

• **Index prefix**: A unique prefix used for Elasticsearch indices managed by the index set. The prefix must start with a letter or number, and can only contain letters, numbers, _, and -. The index alias will be named accordingly, e.g. `graylog_deflector` if the index prefix was `graylog`.

• **Analyzer**: (default: `standard`) The Elasticsearch analyzer for the index set.

• **Index shards**: (default: 4) The number of Elasticsearch shards used per index.

• **Index replicas**: (default: 0) The number of Elasticsearch replicas used per index.

• **Max. number of segments**: (default: 1) The maximum number of segments per Elasticsearch index after index optimization (force merge), see Segment Merging for details.

• **Disable index optimization after rotation**: Disable Elasticsearch index optimization (force merge) after index rotation. Only activate this if you have serious problems with the performance of your Elasticsearch cluster during the optimization process.

### Index rotation

• **Message count**: Rotates the index after a specific number of messages have been written.

• **Index size**: Rotates the index after an approximate size on disk (before optimization) has been reached.

• **Index time**: Rotates the index after a specific time (e.g. 1 hour or 1 week).
Index Rotation Configuration
Graylog uses multiple indices to store documents in. You can configure the strategy it uses to determine when to rotate the currently active write index.

Select rotation strategy
- Index Message Count

Max documents per index
- 20000000

Maximum number of documents in an index before it gets rotated

Index retention

- **Delete**: Delete indices in Elasticsearch to minimize resource consumption.
- **Close**: Close indices in Elasticsearch to reduce resource consumption.
- **Do nothing**
- **Archive**: Commercial feature, see Archiving.

Index Retention Configuration
Graylog uses a retention strategy to clean up old indices.

Select retention strategy
- Delete Index

Max number of indices
- 20

Maximum number of indices to keep before deleting the oldest ones

Maintenance

Keeping the index ranges in sync

Graylog will take care of calculating index ranges automatically as soon as a new index has been created.

In case the stored metadata about index time ranges has run out of sync, Graylog will notify you in the web interface. This can happen if an index was deleted manually or messages from already “closed” indices were removed.

The system will offer you to just re-generate all time range information. This may take a few seconds but is an easy task for Graylog.

You can easily re-build the information yourself after manually deleting indices or doing other changes that might cause synchronization problems:

```bash
$ curl -XPOST http://127.0.0.1:9000/api/system/indices/ranges/rebuild
```

This will trigger a system job:
Manually rotating the active write index

Sometimes you might want to rotate the active write index manually and not wait until the configured rotation criterion for in the latest index has been met, for example if you’ve changed the index mapping or the number of shards per index.

You can do this either via an HTTP request against the REST API of the Graylog master node or via the web interface:

$ curl -XPOST http://127.0.0.1:9000/api/system/deflector/cycle

Triggering this job produces log output similar to the following lines:

INFO : org.graylog2.indexer.MongoIndexSet - Cycling from <graylog_97> to <graylog_98>.
INFO : org.graylog2.indexer.MongoIndexSet - Index <graylog_98> has been successfully allocated.
INFO : org.graylog2.indexer.MongoIndexSet - Pointing index alias <graylog_deflector> to new index <graylog_98>.

Backup

When it comes to backup in a Graylog setup it is not easy to answer. You need to consider what type of backup will suit your needs.

Your Graylog Server setup and settings are easy to backup with a MongoDB dump and a filesystem backup of all configuration files.

The data within your Elasticsearch Cluster can take the advantage of the Snapshot and Restore function that are offered by Elasticsearch.

Default file locations

Each installation flavor of Graylog will place the configuration files into a specific location on the local files system. The goal of this section is to provide a short overview about the most common and most important default file locations.
DEB package

This paragraph covers Graylog installations on Ubuntu Linux, Debian Linux, and Debian derivates installed with the DEB package.

<table>
<thead>
<tr>
<th>File system path</th>
</tr>
</thead>
<tbody>
<tr>
<td>Configuration</td>
</tr>
<tr>
<td>Logging configuration</td>
</tr>
<tr>
<td>Plugins</td>
</tr>
<tr>
<td>JVM settings</td>
</tr>
<tr>
<td>Message journal files</td>
</tr>
<tr>
<td>Log Files</td>
</tr>
</tbody>
</table>

Elasticsearch

Note: These are only the most common file locations. Please refer to the Elasticsearch documentation for a comprehensive list of default file locations.

<table>
<thead>
<tr>
<th>File system path</th>
</tr>
</thead>
<tbody>
<tr>
<td>Configuration</td>
</tr>
<tr>
<td>JVM settings</td>
</tr>
<tr>
<td>Data files</td>
</tr>
<tr>
<td>Log files</td>
</tr>
</tbody>
</table>

MongoDB

<table>
<thead>
<tr>
<th>File system path</th>
</tr>
</thead>
<tbody>
<tr>
<td>Configuration</td>
</tr>
<tr>
<td>Data files</td>
</tr>
<tr>
<td>Log files</td>
</tr>
</tbody>
</table>

RPM package

This paragraph covers Graylog installations on Fedora Linux, Red Hat Enterprise Linux, CentOS Linux, and other Red Hat Linux derivates installed with the RPM package.

<table>
<thead>
<tr>
<th>File system path</th>
</tr>
</thead>
<tbody>
<tr>
<td>Configuration</td>
</tr>
<tr>
<td>Logging configuration</td>
</tr>
<tr>
<td>Plugins</td>
</tr>
<tr>
<td>JVM settings</td>
</tr>
<tr>
<td>Message journal files</td>
</tr>
<tr>
<td>Log Files</td>
</tr>
</tbody>
</table>
**Elasticsearch**

*Note:* These are only the most common file locations. Please refer to the Elasticsearch documentation for a comprehensive list of default file locations.

<table>
<thead>
<tr>
<th>File system path</th>
</tr>
</thead>
<tbody>
<tr>
<td>Configuration</td>
</tr>
<tr>
<td>JVM settings</td>
</tr>
<tr>
<td>Data files</td>
</tr>
<tr>
<td>Log files</td>
</tr>
</tbody>
</table>

**MongoDB**

<table>
<thead>
<tr>
<th>File system path</th>
</tr>
</thead>
<tbody>
<tr>
<td>Configuration</td>
</tr>
<tr>
<td>Data files</td>
</tr>
<tr>
<td>Log files</td>
</tr>
</tbody>
</table>

**Omnibus package**

This paragraph covers Graylog installations via OVA, on AWS (via AMI), and on OpenStack using the Graylog Omnibus package.

**Graylog**

<table>
<thead>
<tr>
<th>File system path</th>
</tr>
</thead>
<tbody>
<tr>
<td>Configuration</td>
</tr>
<tr>
<td>Logging configuration</td>
</tr>
<tr>
<td>Plugins</td>
</tr>
<tr>
<td>JVM settings</td>
</tr>
<tr>
<td>Message journal files</td>
</tr>
<tr>
<td>Log files</td>
</tr>
</tbody>
</table>

**Elasticsearch**

*Note:* These are only the most common file locations. Please refer to the Elasticsearch documentation for a comprehensive list of default file locations.

<table>
<thead>
<tr>
<th>File system path</th>
</tr>
</thead>
<tbody>
<tr>
<td>Configuration</td>
</tr>
<tr>
<td>JVM settings</td>
</tr>
<tr>
<td>Data files</td>
</tr>
<tr>
<td>Log files</td>
</tr>
</tbody>
</table>
**MongoDB**

<table>
<thead>
<tr>
<th>Configuration</th>
<th>File system path</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data files</td>
<td>/var/opt/graylog/data/mongodb</td>
</tr>
<tr>
<td>Log files</td>
<td>/var/log/graylog/mongodb/</td>
</tr>
</tbody>
</table>

**Graylog REST API**

The functionality Graylog REST API is very comprehensive; even the Graylog web interface is exclusively using Graylog REST API to interact with the Graylog cluster.

To connect to the Graylog REST API with a web browser, just add `api-browser` to your current `rest_listen_uri` setting or use the **API browser** button on the nodes overview page (System / Nodes in the web interface).

For example if your Graylog REST API is listening on `http://192.168.178.26:9000/api/`, the API browser will be available at `http://192.168.178.26:9000/api/api-browser/`.

**Note:** The customized version of Swagger UI used by Graylog does currently only work in Google Chrome and Firefox.

**Using the API browser**

After providing the credentials (username and password), you can browse all available HTTP resources of the Graylog REST API.
Interacting with the Graylog REST API

While having a graphical UI for the Graylog REST API is perfect for interactive usage and exploratory learning, the real power unfolds when using the Graylog REST API for automation or integrating Graylog into another system, such as monitoring or ticket systems.

Naturally, the same operations the API browser offers can be used on the command line or in scripts. A very common HTTP client being used for this kind of interaction is curl.

Note: In the following examples, the username `GM` and password `superpower` will be used to demonstrate how to...

The following command displays Graylog cluster information as JSON, exactly the same information the web interface is displaying on the System / Nodes page:


The Graylog REST API will respond with the following information:

```
{
  "71ab6aaa-cb39-46be-9dac-4ba99fed3d66" : {
    "facility" : "graylog-server",
    "codename" : "Smuttynose",
    "node_id" : "71ab6aaa-cb39-46be-9dac-4ba99fed3d66",
    "cluster_id" : "3adaf799-1551-4239-84e5-6ed939b56f62",
    "version" : "2.1.1+01d50e5",
    "started_at" : "2016-09-23T10:39:00.179Z",
    "hostname" : "gm-01-c.fritz.box",
    "lifecycle" : "running",
    "lb_status" : "alive",
    "timezone" : "Europe/Berlin",
    "operating_system" : "Linux 3.10.0-327.28.3.el7.x86_64",
    "is_processing" : true
  },
  "ed0ad32d-8776-4d25-be2f-a8956ecebdcf" : {
    "facility" : "graylog-server",
    "codename" : "Smuttynose",
    "node_id" : "ed0ad32d-8776-4d25-be2f-a8956ecebdcf",
    "cluster_id" : "3adaf799-1551-4239-84e5-6ed939b56f62",
    "version" : "2.1.1+01d50e5",
    "started_at" : "2016-09-23T10:40:07.325Z",
    "hostname" : "gm-01-d.fritz.box",
    "lifecycle" : "running",
    "lb_status" : "alive",
    "timezone" : "Europe/Berlin",
    "operating_system" : "Linux 3.16.0-4-amd64",
    "is_processing" : true
  },
  "58c57924-808a-4fa7-be09-63ca551628cd" : {
    "facility" : "graylog-server",
    "codename" : "Smuttynose",
    "node_id" : "58c57924-808a-4fa7-be09-63ca551628cd",
    "cluster_id" : "3adaf799-1551-4239-84e5-6ed939b56f62",
    "version" : "2.1.1+01d50e5",
    "started_at" : "2016-09-30T13:31:39.051Z",
    "hostname" : "gm-01-u.fritz.box",
    "lifecycle" : "running",
    "lb_status" : "alive",
    "timezone" : "Europe/Berlin",
    "operating_system" : "Linux 4.4.0-36-generic",
    "is_processing" : true
  }
}
```

**Creating and using Access Token**

For security reasons, using the username and password directly on the command line or in some third party application is undesirable.
To prevent having to use the clear text credentials, Graylog allows to create access tokens which can be used for authentication instead.

In order to create a new access token, you need to send a POST request to the Graylog REST API which includes the username and the name of the new access token.

**Note:** Users require the permissions `users:tokenlist`, `users:tokencreate`, and `users:tokenremove` to manage their access tokens. Please check the documentation on Permission system for more information. Also note that users, even administrators, may only manage their own tokens.

The following example will create an access token named `icinga` for the user `GM`:

```
```

The response will include the access token in the `token` field:

```
{
   "name" : "icinga",
   "token" : "htgi84ut7jpivsrcldd6l4lmcigvfauldm99ofcb4hsfcvdsru",
   "last_access" : "1970-01-01T00:00:00.000Z"
}
```

The received access token can now be used as username in a request to the Graylog REST API using Basic Auth together with the literal password `token`.

Now the first `curl` example would look as follows:

```
```

If you need to know which access tokens have already been created by a user, just use GET `/users/{username}/tokens/` on the Graylog REST API to request a list of all access tokens that are present for this user.

The following example will request all access tokens of the user `GM`:

```
```

When an access token is no longer needed, it can be delete on the Graylog REST API via DELETE `/users/{username}/tokens/{token}`.

The following example deletes the previously created access token `htgi84ut7jpivsrcldd6l4lmcigvfauldm99ofcb4hsfcvdsru` of the user `GM`:

```
curl -u GM:superpower -H 'Accept: application/json' -X DELETE 'http://192.168.178.26:9000/api/users/GM/tokens/ap84p4jehbf2jddva8rdmjr3k7m3kdnuqbai5s0h5a48e7069po?pretty=true'
```

### Creating and using Session Token

While access tokens can be used for permanent access, session tokens will expire after a certain time. The expiration time can be adjusted in the user’s profile.

Getting a new session token can be obtained via POST request to the Graylog REST API. Username and password are required to get a valid session ID. The following example will create an session token for the user `GM`:

```
```

The response will include the session token in the field `session_id` and the time of expiration:
The received token can now be used as username in a request to the Graylog REST API using Basic Auth together with the literal password `session`.

Now a `curl` command to get a list of access tokens would look as follows:

```
Securing Graylog

To secure your Graylog setup, you should not use one of our pre-configured images, create your own unique installation where you understand each component and secure the environment by design. Expose only the services that are needed and secure them whenever possible with TLS/SSL and some kind of authentication. Do not use the pre-created appliances for critical production environments.

On the Graylog appliances MongoDB and Elasticsearch is listening on the external interface. This makes the creation of a cluster easier and demonstrates the way Graylog works. Never run this in an insecure network.

When using Amazon Web Services and our pre-configured AMI, never open all ports in the security group. Do not expose the server to the internet. Access Graylog only from within your VPC. Enable encryption for the communication.

Default ports

All parts of one Graylog installation will communicate over network sockets. Depending on your setup and number of nodes this might be exposed or can be bound to localhost. The SELinux configuration is already covered in our step-by-step guide for CentOS Linux.

Table 6.1: Default network communication ports

<table>
<thead>
<tr>
<th>Component</th>
<th>Port</th>
</tr>
</thead>
<tbody>
<tr>
<td>Graylog (web interface / API)</td>
<td>9000 (tcp)</td>
</tr>
<tr>
<td>Graylog to Elasticsearch</td>
<td>9350 (tcp)</td>
</tr>
<tr>
<td>Elasticsearch node communication</td>
<td>9300 (tcp)</td>
</tr>
<tr>
<td>MongoDB</td>
<td>27017 (tcp)</td>
</tr>
</tbody>
</table>

Each setup is unique in the requirements and ports might be changed by configuration, but you should limit who is able to connect to which service. In the architecture description you can see what components need to be exposed and communicate with each other.

Logging user activity

Graylog has been built using a client-server architecture model in which the user interface retrieves all data via a collection of REST APIs. Thus logging relevant user activity, in other words an access log, is simply a matter of enabling a built-in feature. It logs all requests to the Graylog REST API and produces an access log augmented by additional information, like the user name, the remote address, and the user agent.
Configuring the Access Log

The Access Log is configured by adding an appender and logger to the Log4j2 configuration file (log4j2.xml). The following example demonstrates required additions on top of the normal logging configuration:

```xml
<?xml version="1.0" encoding="UTF-8"?>
<Configuration packages="org.graylog2.log4j" shutdownHook="disable">
  <Appenders>
    <!-- Simple appender that writes access log to specified file -->
    <File name="RestAccessLog" fileName="/var/log/graylog/server/restaccess.log" append="true">
      <PatternLayout pattern="%d %-5p: %c - %m%n"/>
    </File>
  </Appenders>
  <Loggers>
    <!-- RestAccessLogFilter -->
    <Logger name="org.graylog2.rest.accesslog" level="debug" additivity="false">
      <AppenderRef ref="RestAccessLog" level="debug"/>
      <AppenderRef ref="STDOUT" level="info"/>
    </Logger>
  </Loggers>
</Configuration>
```

The resulting log entries will look similar to the following messages:

```
2016-06-08 18:21:55,651 DEBUG: org.graylog2.rest.accesslog - 192.168.122.1 admin [-] "GET streams" Mozilla/5.0 (X11; Fedora; Linux x86_64; rv:46.0) Gecko/20100101 Firefox/46.0 200 -1
2016-06-08 18:21:55,694 DEBUG: org.graylog2.rest.accesslog - 192.168.122.1 admin [-] "GET system/fields" Mozilla/5.0 (X11; Fedora; Linux x86_64; rv:46.0) Gecko/20100101 Firefox/46.0 200 -1
2016-06-08 18:21:55,698 DEBUG: org.graylog2.rest.accesslog - 192.168.122.1 admin [-] "GET system/fields" Mozilla/5.0 (X11; Fedora; Linux x86_64; rv:46.0) Gecko/20100101 Firefox/46.0 200 -1
2016-06-08 18:21:55,780 DEBUG: org.graylog2.rest.accesslog - 192.168.122.1 admin [-] "GET system/inputs" Mozilla/5.0 (X11; Fedora; Linux x86_64; rv:46.0) Gecko/20100101 Firefox/46.0 200 -1
2016-06-08 18:21:56,021 DEBUG: org.graylog2.rest.accesslog - 192.168.122.1 admin [-] "GET search/unparsed" Mozilla/5.0 (X11; Fedora; Linux x86_64; rv:46.0) Gecko/20100101 Firefox/46.0 200 -1
```

X-Forwarded-For HTTP header support

If there is a proxy server, load balancer, or a network device which hides the client’s IP address from Graylog, it can read the information from a supplied X-Forwarded-For HTTP request header. Most HTTP-capable devices support setting such a (semi-) standard HTTP request header.

Since overriding the client address from a externally supplied HTTP request header opens the door for spoofing, the list of trusted proxy servers which are allowed to provide the X-Forwarded-For HTTP request header, can be configured using the trusted_proxies setting in the Graylog configuration file (graylog.conf):

```
# Comma separated list of trusted proxies that are allowed to set
# the client address with X-Forwarded-For HTTP request header.
# May be subnets, or hosts.
trusted_proxies = 127.0.0.1/32, 0:0:0:0:0:0:0:1/128
```
A Graylog setup is pretty worthless without any data in it. This page explains the basic principles of getting your data into the system and also explains common fallacies.

What are Graylog message inputs?

Message inputs are the Graylog parts responsible for accepting log messages. They are launched from the web interface (or the REST API) in the System -> Inputs section and are launched and configured without the need to restart any part of the system.

Content packs

Content packs are bundles of Graylog input, extractor, stream, dashboard, and output configurations that can provide full support for a data source. Some content packs are shipped with Graylog by default and some are available from the website. Content packs that were downloaded from the Graylog Marketplace can be imported using the Graylog web interface.

You can load and even create own content packs from the System -> Content Packs section of your Graylog web interface.

Syslog

Graylog is able to accept and parse RFC 5424 and RFC 3164 compliant syslog messages and supports TCP transport with both the octet counting or termination character methods. UDP is also supported and the recommended way to send log messages in most architectures.

Many devices, especially routers and firewalls, do not send RFC compliant syslog messages. This might result in wrong or completely failing parsing. In that case you might have to go with a combination of raw/plaintext message inputs that do not attempt to do any parsing and Extractors.

Rule of thumb is that messages forwarded by rsyslog or syslog-ng are usually parsed flawlessly.

Sending syslog from Linux hosts

Sending syslog data from Linux hosts is described on the Graylog Marketplace.
Sending syslog from MacOS X hosts

Sending log messages from MacOS X syslog daemons is easy. Just define a `graylog-server` instance as UDP log target by adding this line in your `/etc/syslog.conf`:

```
*.@graylog.example.org:514
```

Now restart `syslogd`:

```
$ sudo launchctl unload /System/Library/LaunchDaemons/com.apple.syslogd.plist
$ sudo launchctl load /System/Library/LaunchDaemons/com.apple.syslogd.plist
```

**Important:** If `syslogd` was running as another user you might end up with multiple `syslogd` instances and strange behavior of the whole system. Please check that only one `syslogd` process is running:

```
$ ps aux | grep syslog
lennart  58775  0.0  0.0  2432768  592   s004  S+   6:10PM     0:00.00 grep syslog
root    58769  0.0  0.0  2478772 1020  ??    Ss  6:09PM     0:00.01 /usr/sbin/syslogd
```

That’s it! Your MacOS X syslog messages should now appear in your Graylog system.

GELF / Sending from applications

The Graylog Extended Log Format (GELF) is a log format that avoids the shortcomings of classic plain syslog and is perfect to logging from your application layer. It comes with optional compression, chunking and most importantly a clearly defined structure. There are dozens of GELF libraries for many frameworks and programming languages to get you started.

Read more about GELF in the specification.

GELF via HTTP

You can send in all GELF types via HTTP, including uncompressed GELF that is just a plain JSON string.

After launching a GELF HTTP input you can use the following endpoints to send messages:

```
http://graylog.example.org:[port]/gelf (POST)
```

Try sending an example message using curl:

```
curl -XPOST http://graylog.example.org:12202/gelf -p0 -d '{"short_message":"Hello there", "host":"example.org", "facility":"test", "_foo":"bar"}'
```

Both keep-alive and compression are supported via the common HTTP headers. The server will return a 202 Accepted when the message was accepted for processing.

Using Apache Kafka as transport queue

Graylog supports Apache Kafka as a transport for various inputs such as GELF, syslog, and Raw/Plaintext inputs. The Kafka topic can be filtered by a regular expression and depending on the input, various additional settings can be configured.

Learn how to use rsyslog and Apache Kafka in the Sending syslog via Kafka into Graylog guide.
Using RabbitMQ (AMQP) as transport queue

Graylog supports AMQP as a transport for various inputs such as GELF, syslog, and Raw/Plaintext inputs. It can connect to any AMQP broker supporting AMQP 0-9-1 such as RabbitMQ.

Learn how to use rsyslog and RabbitMQ in the Sending syslog via AMQP into Graylog guide.

Microsoft Windows

Sending syslog data from Windows is described on the Graylog Marketplace.

Heroku

Heroku allows you to forward the logs of your application to a custom syslog server by creating a so called Syslog drain. The drain sends all logs to the configured server(s) via TCP. Following example shows you how to configure Graylog to receive the Heroku logs and extract the different fields into a structured log message.

Configuring Graylog to receive Heroku log messages

The Graylog Marketplace contains a content pack for Heroku logs, including extractors to parse the Heroku log format. You can download and use that content pack to configure Graylog to be able to receive Heroku logs.

Go to System -> Content packs, and click on Import content pack. Select the content pack downloaded from the Graylog Marketplace, and click Upload.
On the same page, select *Heroku* on the *SaaS* category on the left column, and click on *Apply*.

![Heroku on the SaaS category](image)

That’s it! You can verify that there is a new input for Heroku, containing a set of extractors to parse your log messages. Make sure your firewall setup allows incoming connections on the inputs port!

![Configuring Heroku to send data to your Graylog setup](image)

**Configuring Heroku to send data to your Graylog setup**

Heroku has a detailed documentation regarding the Syslog drains feature. The following example shows everything that is needed to setup the drain for your application:

```
$ cd path/to/your/heroku/app
$ heroku drains
No drains for this app
$ heroku drains:add syslog://graylog.example.com:5556
Successfully added drain syslog://graylog.example.com:5556
$ heroku drains syslog://graylog.example.com:5556 (d.8cf52d32-7d79-4653-baad-8cb72bb23ee1)
```

The *Heroku CLI tool* needs to be installed for this to work.
You Heroku application logs should now show up in the search results of your Graylog instance.

**Ruby on Rails**

This is easy: You just need to combine a few components.

**Log all requests and logger calls into Graylog**

The recommended way to send structured information (i.e. HTTP return code, action, controller, ... in additional fields) about every request and explicit `Rails.logger` calls is easily accomplished using the GELF gem and lograge. Lograge builds one combined log entry for every request (instead of several lines like the standard Rails logger) and has a Graylog output since version 0.2.0.

Start by adding Lograge and the GELF gem to your Gemfile:

```ruby
gem "gelf"
gem "lograge"
```

Now configure both in your Rails application. Usually `config/environments/production.rb` is a good place for that:

```ruby
config.lograge.enabled = true
config.lograge.formatter = Lograge::Formatters::Graylog2.new
config.logger = GELF::Logger.new("graylog.example.org", 12201, "WAN", { :host => "hostname-of-this-app", :facility => "heroku" })
```

This configuration will also send all explicit `Rails.logger` calls (e.g. `Rails.logger.error "Something went wrong"`) to Graylog.

**Log only explicit logger calls into Graylog**

If you don’t want to log information about every request, but only explicit `Rails.logger` calls, it is enough to only configure the Rails logger.

Add the GELF gem to your Gemfile:

```ruby
gem "gelf"
```

...and configure it in your Rails application. Usually `config/environments/production.rb` is a good place for that:

```ruby
config.logger = GELF::Logger.new("graylog.example.org", 12201, "WAN", { :host => "hostname-of-this-app" })
```

**Heroku**

You need to apply a workaround if you want custom logging on Heroku. The reason for this is that Heroku injects an own logger (`rails_log_stdout`), that overwrites your custom one. The workaround is to add a file that makes Heroku think that the logger is already in your application:

```bash
$ touch vendor/plugins/rails_log_stdout/heroku_fix
```
Raw/Plaintext inputs

The built-in raw/plaintext inputs allow you to parse any text that you can send via TCP or UDP. No parsing is applied at all by default until you build your own parser using custom Extractors. This is a good way to support any text-based logging format.

You can also write Plugins if you need extreme flexibility.

JSON path from HTTP API input

The JSON path from HTTP API input is reading any JSON response of a REST resource and stores a field value of it as a Graylog message.

Example

Let’s try to read the download count of a release package stored on GitHub for analysis in Graylog. The call looks like this:

```bash
$ curl -X GET https://api.github.com/repos/YourAccount/YourRepo/releases/assets/12345
```

```json
{
    "url": "https://api.github.com/repos/YourAccount/YourRepo/releases/assets/12345",
    "id": 12345,
    "name": "somerelease.tgz",
    "label": "somerelease.tgz",
    "content_type": "application/octet-stream",
    "state": "uploaded",
    "size": 3817925,
    "download_count": 9937,
    "created_at": "2013-09-30T20:05:01Z",
    "updated_at": "2013-09-30T20:05:46Z"
}
```

The attribute we want to extract is download_count so we set the JSON path to $.download_count.

This will result in a message in Graylog looking like this:

![Graylog message example]

You can use Graylog to analyze your download counts now.
**JSONPath**

JSONPath can do much more than just selecting a simple known field value. You can for example do this to select the first `download_count` from a list of releases where the field `state` has the value `uploaded`:

```
$.releases[?(@.state == 'uploaded')][0].download_count
```

...or only the first download count at all:

```
$.releases[0].download_count
```

You can learn more about JSONPath here.

**Reading from files**

Log files come in a lot of different flavors and formats, much more than any single program could handle.

To support this use case, we provide the `Collector Sidecar` which acts as a supervisor process for other programs, such as nxlog and Filebeats, which have specifically been built to collect log messages from local files and ship them to remote systems like Graylog.

Of course you can still use any program supporting the GELF or syslog protocol (among others) to send your logs to Graylog.
Graylog Collector Sidecar

Graylog Collector Sidecar is a lightweight configuration management system for different log collectors, also called Backends. The Graylog node(s) act as a centralized hub containing the configurations of log collectors. On supported message-producing devices/hosts, Sidecar can run as a service (Windows host) or daemon (Linux host).

These configurations are centrally managed through the Graylog web interface, in a graphical way. For specific needs, raw backend configurations, called Snippets, may optionally be directly stored into Graylog.

Periodically, the Sidecar daemon will fetch all relevant configurations for the target, using the REST API. Which configurations are actually fetched depends on ‘tags’ defined in the host’s Sidecar configuration file. For instance, a Web server host may include the linux and nginx tags.

On its first run, or when a configuration change has been detected, Sidecar will generate (render) relevant backend configuration files. Then it will start, or restart, those reconfigured log collectors.
Graylog Collector Sidecar (written in Go) and backends (written in various languages, such as C and Go) are meant as a small-footprint replacement for the deprecated, Java-based deprecated Graylog Collector.

**Backends**

Currently the Sidecar is supporting NXLog, Filebeat and Winlogbeat. They all share the same web interface. Switch the tab on a configuration page to create resources for the used collector. The supported features are almost the same. For all collectors a GELF output with SSL encryption is available. The most used input options like file tailing or windows event logging do exist. On the server side you can share inputs with multiple collectors. E.g. All Filebeat and Winlogbeat instances can send logs into a single Graylog-Beats input.

**Installation**

Currently we provide pre-compiled packages on the Github releases page of the project. Once the Sidecar project is settled and matured we will add the packages to the DEB and YUM online repositories. To get the Sidecar working Download a package and install it on the target system.

Please follow the version matrix to pick the right package:

<table>
<thead>
<tr>
<th>Sidecar version</th>
<th>Graylog server version</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0.9</td>
<td>2.1.x</td>
</tr>
<tr>
<td>0.1.x</td>
<td>2.2.x</td>
</tr>
</tbody>
</table>

All following commands should be executed on the remote machine where you want to collect log data from.

**Beats backend**

**Ubuntu**

The Beats binaries (Filebeat and Winlogeventbeat) are included in the Sidecar package. So installation is just one command:

```
$ sudo dpkg -i collector-sidecar_0.0.9-1_amd64.deb
```

Edit `/etc/graylog/collector-sidecar/collector_sidecar.yml`, you should set at least the correct URL to your Graylog server and proper tags. The tags are used to define which configurations the host should receive.

Create a system service and start it:

```
$ sudo graylog-collector-sidecar --service install
```

[Ubuntu 14.04 with Upstart]

```
$ sudo start collector-sidecar
```

[Ubuntu 16.04 with Systemd]

```
$ sudo systemctl start collector-sidecar
```

**CentOS**

Install the RPM package on RedHat based systems
$ sudo rpm -i collector-sidecar-0.0.9-1.x86_64.rpm

Activate the Sidecar as a system service:

$ sudo graylog-collector-sidecar -service install
$ sudo systemctl start collector-sidecar

Windows

Use the Windows installer, it can be run interactively:

$ collector_sidecar_installer.exe

or in silent mode with:

$ collector_sidecar_installer.exe /S -SERVERURL=http://10.0.2.2:9000/api -TAGS="windows,iis"

Edit C:\Program Files\graylog\collector-sidecar\collector_sidecar.yml and register the system service:

$ "C:\Program Files\graylog\collector-sidecar\graylog-collector-sidecar.exe" -service install
$ "C:\Program Files\graylog\collector-sidecar\graylog-collector-sidecar.exe" -service start

NXLog backend

Ubuntu

Install the NXLog package from the official download page. Because the Sidecar takes control of stopping and starting NXlog it’s necessary to stop all running instances of NXlog and deconfigure the default system service. Afterwards we can install and setup the Sidecar:

$ sudo /etc/init.d/nxlog stop
$ sudo update-rc.d -f nxlog remove
$ sudo gpasswd -a nxlog adm
$ sudo chown -R nxlog.nxlog /var/spool/collector-sidecar/nxlog

$ sudo dpkg -i collector-sidecar_0.0.9-1_amd64.deb

Edit /etc/graylog/collector-sidecar/collector_sidecar.yml accordingly and register the Sidecar as a service:

$ sudo graylog-collector-sidecar -service install

[Ubuntu 14.04 with Upstart]
$ sudo start collector-sidecar

[Ubuntu 16.04 with Systemd]
$ sudo systemctl start collector-sidecar

CentOS

The same on a RedHat based system:

$ sudo service nxlog stop
$ sudo chkconfig --del nxlog
$ sudo gpasswd -a nxlog root
$ sudo chown -R nxlog.nxlog /var/spool/collector-sidecar/nxlog

8.2. Installation
$ sudo rpm -i collector-sidecar-0.0.9-1.x86_64.rpm

Activate the Sidecar as a system service:

$ sudo graylog-collector-sidecar -service install
$ sudo systemctl start collector-sidecar

Windows

Install the NXLog package from the official download page and deactivate the system service. We just need the binaries installed on the system:

$ C:\Program Files (x86)\nxlog\nxlog -u
$ collector_sidecar_installer.exe

Edit C:\Program Files\graylog\collector-sidecar\collector_sidecar.yml, you should set at least the correct URL to your Graylog server and proper tags. Register the system service:

$ C:\Program Files\graylog\collector-sidecar\graylog-collector-sidecar.exe -service install
$ C:\Program Files\graylog\collector-sidecar\graylog-collector-sidecar.exe -service start

To perform an uninstall on Windows:

$ C:\Program Files\graylog\collector-sidecar\graylog-collector-sidecar.exe -service stop
$ C:\Program Files\graylog\collector-sidecar\graylog-collector-sidecar.exe -service uninstall

Notice that the NXLog file input is currently not able to do a SavePos for file tailing, this will be fixed in a future version.

Configuration

On the command line you can provide a path to the configuration file with the -c switch. If no path is specified it looks on Linux systems for:

/etc/graylog/collector-sidecar/collector_sidecar.yml

and on Windows machines under:

C:\Program Files\graylog\collector-sidecar\collector_sidecar.yml

The configuration file is separated into global options and backend specific options. Global options are:
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>server_url</td>
<td>URL to the Graylog API, e.g. <a href="http://127.0.0.1:9000/api/">http://127.0.0.1:9000/api/</a></td>
</tr>
<tr>
<td>update_interval</td>
<td>The interval in seconds the sidecar will fetch new configurations from the Graylog server</td>
</tr>
<tr>
<td>tls_skip_verify</td>
<td>Ignore errors when the REST API was started with a self-signed certificate</td>
</tr>
<tr>
<td>send_status</td>
<td>Send the status of each backend back to Graylog and display it on the status page for the host</td>
</tr>
<tr>
<td>list_log_files</td>
<td>Send a directory listing to Graylog and display it on the host status page, e.g. /var/log. This can also be a list of directories</td>
</tr>
<tr>
<td>node_id</td>
<td>Name of the Sidecar instance, will also show up in the web interface. Hostname will be used if not set.</td>
</tr>
<tr>
<td>collector_id</td>
<td>Unique ID (UUID) of the instance. This can be a string or a path to an ID file</td>
</tr>
<tr>
<td>log_path</td>
<td>A path to a directory where the Sidecar can store the output of each running collector backend</td>
</tr>
<tr>
<td>log_rotation_time</td>
<td>Rotate the stdout and stderr logs of each collector after X seconds</td>
</tr>
<tr>
<td>log_max_age</td>
<td>Delete rotated log files older than Y seconds</td>
</tr>
<tr>
<td>tags</td>
<td>List of configuration tags. All configurations on the server side that match the tag list will be fetched and merged by this instance</td>
</tr>
<tr>
<td>backends</td>
<td>A list of collector backends the user wants to run on the target host</td>
</tr>
</tbody>
</table>

Currently NXLog and Beats are supported as collector backend, to make it work the Sidecar needs to know where the binary is installed and where it can write a configuration file for it.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>name</td>
<td>Which backend to use (must be ‘nxlog’, ‘filebeat’ or ‘winlogbeat)</td>
</tr>
<tr>
<td>enabled</td>
<td>Whether this backend should be started by the Sidecar or not</td>
</tr>
<tr>
<td>binary_path</td>
<td>Path to the actual collector binary</td>
</tr>
<tr>
<td>configuration_path</td>
<td>Path to the configuration file for this collector</td>
</tr>
</tbody>
</table>

An example configuration for NXlog looks like this:

```
server_url: http://10.0.2.2:9000/api/
update_interval: 30
tls_skip_verify: true
send_status: true
list_log_files:
  - /var/log
node_id: graylog-collector-sidecar
collector_id: file:/etc/graylog/collector-sidecar/collector-id
log_path: /var/log/graylog/collector-sidecar
log_rotation_time: 86400
log_max_age: 604800
tags:
  - linux
  - apache
  - redis
backends:
  - name: nxlog
    enabled: true
    binary_path: /usr/bin/nxlog
    configuration_path: /etc/graylog/collector-sidecar/generated/nxlog.conf
```

For the Beats platform you can enable each Beat individually, e.g on a Windows host with Filebeat and Winlogbeat enabled use a configuration like this:

```
server_url: http://10.0.2.2:9000/api/
update_interval: 30
tls_skip_verify: true
send_status: true
list_log_files:
```

8.3. Configuration
On the server side the collector plugin is caching the requested configuration in memory. By default up-to 100 entries are stored for 1 hour. If you wish to change that, add to your server configuration:

```yaml
collector_sidecar_cache_time = 2h
collector_sidecar_cache_max_size = 500
```

### First start

Once you installed the Sidecar package you are ready to start the service for the first time. Decide which backend you want to use. Enable or disable the single backends by setting `enabled: true` or respectively to `false`. Now start the Sidecar, depending on your operating system you can do this with:

<table>
<thead>
<tr>
<th>Operating System</th>
<th>Command</th>
</tr>
</thead>
<tbody>
<tr>
<td>Debian/Ubuntu</td>
<td><code>sudo start collector-sidecar</code></td>
</tr>
<tr>
<td>RedHat/CentOS</td>
<td><code>sudo systemctl start collector-sidecar</code></td>
</tr>
<tr>
<td>Windows</td>
<td><code>C:\Program Files\graylog\collector-sidecar\graylog-collector-sidecar.exe -service start</code></td>
</tr>
</tbody>
</table>

Afterwards you will most likely see an error like this in the log file:

```
INFO[0006] [RequestConfiguration] No configuration found for configured tags!
```

This simply means that there is no configuration with the same tag that the Sidecar was started with. So we have to create a new configuration. Define outputs and inputs and tag it in order to collect log files. Take the *Step-by-step guide* to create your first configuration.

When the Sidecar can find a configuration that matches its own tags, it will write for each collector backend a configuration file into the `/generated` directory. E.g. if you enabled the Filebeat collector you will find a `filebeat.yml` file in that directory. All changes have to be made in the Graylog web interface. Everytime the Sidecar detects an update to its configuration it will rewrite the corresponding collector configuration file. So it doesn’t make sense to manually edit those files.

Everytime a collector configuration file is changed the collector process is restarted. The Sidecar takes care of the collector processes and reports the status back to the web interface.
Sidecar Status

Each Sidecar instance is able to send status informations back to Graylog. By enabling the option `send_status` metrics like the configured tags or the IP address of the host Sidecar is running on are send. Also metrics that are relevant for a stable operation e.g. disk volumes over 75% utilization are included. Additionally with the `list_log_files` option a directory listing is displayed in the Graylog web interface. In that way an administrator can see which files are available for collecting. The list is periodically updated and files with write access are highlighted for easy identification. After enabling `send_status` or `send_status + list_log_files` go to the collector overview and click on one of them, a status page with the configured information will be displayed.

Step-by-step guide

We have prepared an example on how to configure the Sidecar using the Graylog Webinterface. The assumption is that we want to collect Apache logfiles and ship them with a Filebeat collector to a Beats input that is listening on Port 5044 on your Graylog Server.

- The first step is to create a Beats input where collectors can send data to. Click on System → Inputs and start a global Beats input on the listening address 0.0.0.0 and port 5044.

- Navigate to the collector configurations. In your Graylog Webinterface click on System → Collectors → Manage configurations.
• Next we create a new configuration
• Give the configuration a name
• Click on the new configuration and create e.g. a Filebeat output. For a first test just change the IP to your Graylog server.
• Create a Filebeat file input to collect the Apache access logs.
• Tag the configuration with the **apache** tag. Just write the tag name in the field press enter followed by the **Update tags** button.
• When you now start the Sidecar with the *apache* tag the output should look like this
• Congratulations your collector setup is working now!

**Secure Sidecar Communication**

The Communication between Sidecar and Graylog will be secured if your API *use SSL*.

To secure the communication between the Collector and Graylog you just need to mark Enable TLS in your Beats Input. Without giving additional Information, Graylog will now create a self-signed certificate for this Input. Now in the Sidecar Beats Output Configuration you just mark Enable TLS Support and Insecure TLS connection. After this is saved, the communication between Beats and Graylog will use TLS.

If you prefer NXLog you need to mark Allow untrusted certificate in the NXLog Outputs configuration and Enable TLS for your GELF Input.

**Certificate based client authentication**

If you want to allow Graylog only to accept data from certificated clients you will need to build your own certificate authority and provide this to the Input and the Client Output configuration.

**Run Sidecar as non-root user**

The default is that the Sidecar is started with the root user to allow access to all log files. But this is not mandatory. If you like to start it with a daemon user, proceede like the following:
• Create a daemon user e.g. collector

The Sidecar itself is accessing the following files and directories:

• collector_sidecar.yml - /etc/graylog/collector-sidecar/collector_sidecar.yml
• backend_configuration_path - /etc/graylog/collector-sidecar/generated/
• collector_id - /etc/graylog/collector-sidecar/collector-id
• cache_path - /var/cache/graylog/collector-sidecar/
• log_path - /var/log/graylog/collector-sidecar/

So to make these directories readable for the collector user, use:

• chown -R collector /etc/graylog
• chown -R collector /var/cache/graylog
• chown -R collector /etc/graylog

You can change all paths to different places in the filesystem. If you prefer to store all Sidecar data in the home directory of the collector user, just change the paths accordingly.

Now systemd needs to know that the Sidecar should be started with a non-root user. Open /etc/systemd/system/collector-sidecar.service with an editor and navigate to the [Service] section, add:

| User=collector  
| Group=collector |

To make use of these settings reload systemd:

```
$ sudo systemctl daemon-reload
$ sudo systemctl restart collector-sidecar
```

Check the log files in /var/log/graylog/collector-sidecar for any errors. Understand that not only the Sidecar but also all backends, like filebeat, will be started as collector user after these changes. So all log files that the backend should observe also need to be readable by the collector user. Depending on the Linux distribution there is usually an administrator group which has access to most log files. By adding the collector user to that group you can grant access fairly easy. For example on Debian/Ubuntu systems this group is called adm (see System Groups in Debian Wiki or Security/Privileges - Monitor system logs in Ubuntu wiki).

## Sidecar Glossary

To understand the different parts of the Graylog Sidecar they are explained in the following section.

### Configuration

A collector configuration is an abstract representation of a collector configuration file. It contains one or many Outputs, Inputs and Snippets. Based on the selected backend the Sidecar will then render a working configuration file for the particular collector. To match a configuration for a Sidecar instance both sides need to be started with the same tag. If the tags of a Sidecar instance match multiple configurations all Out-,Inputs and Snippets are merged together to a single configuration.
Tags

Tags are used to match Sidecar instances with configurations on the Graylog server side. E.g. a user can create a configuration for Apache access log files. The configuration gets the tag `apache`. On all web servers running the Apache daemon the Sidecar can also be started with the `apache` tag to fetch this configuration and to collect web access log files. There can be multiple tags on both sides the Sidecar and the Graylog server side. But to keep the overview the administrator should use at least on one side discrete tags that the assignment is always 1:1 or 1:n.

Outputs

Outputs are used to send data from a collector back to the Graylog server. E.g. NXLog is able to send directly messages in the GELF format. So the natural fit is to create a GELF output in a NXLog configuration. Instructing NXLog to send GELF messages is of course just half the way, we also need a receiver for that. So an administrator needs to create a proper receiver under System → Inputs.

Inputs

Inputs are the way how collectors ingest data. An input can be a log file that the collector should continuous read or a connection to the Windows event system that emits log events. An input is connected to an output, otherwise there would be no way of sending the data to the next hop. So first create an output and then associate one or many inputs with it.

Snippets

Snippets are simply plain text configuration fragments. Sometimes it’s not possible to represent the needed configuration through the provided system. E.g. a user would like to load a special collector module. She could put the directive into a snippet which will be added to the final collector configuration without any modification. It’s also conceivable to put a full configuration file into a snippet and skip all of the input and output mechanism. Before the snippet is actually rendered into the configuration file the Sidecar is sending it through a template engine. It’s using Go’s own text template engine for that. A usage of that can be seen in the `nxlog-default` snippet. It detects which operating the Sidecar is running on and depending on the result, paths for some collector settings change.

Actions

Resources like inputs, output or snippets have all the same actions: create, edit, clone Usually there are only little differences between certain configurations so you can create a resource once, clone it and modify only the fields you need. In this way it’s possible to manage a fairly large amount of configurations.
The Sidecar is writing log files to the directory configured in `log_path`. One file for each backend, there you can check for general issues like file permissions or log transmission problems. The Sidecar itself is writing to `collector_sidecar.log` problems like failed connection to the Graylog API can be found there.

You can also start the Sidecar in foreground and monitor the output of the process:

```
$ graylog-collector-sidecar -debug -c /etc/graylog/collector-sidecar/collector_sidecar.yml
```
Known Problems

Currently we know of two problems with NXLog:

- Since version 2.9.17 timestamps are transmitted without millisecond precision
- On Windows machines NXlog is not able to store its collector state so features like file tailing don’t work correctly in combination with Sidecar. Use Sidecar version 0.1.0-alpha.1 or newer.

Known issue if you use a loadbalancer or firewall in front of Graylog’s API:

- The Sidecar is using a persistent connection for API requests. Therefore it logs 408 Request Time-out if the loadbalancer session or http timeout is lower than the configured update_interval.
Graylog Collector (deprecated)

**Warning:** The Graylog Collector is deprecated and can be replaced with the *Graylog Collector Sidecar*.

Graylog Collector is a lightweight Java application that allows you to forward data from log files to a Graylog cluster. The collector can read local log files and also Windows Events natively, it then can forward the log messages over the network using the *GELF format*.

**Installation**

**Linux/Unix**

You need to have Java $\geq 7$ installed to run the collector.

**Operating System Packages**

We offer official package repositories for the following operating systems.

- Ubuntu 12.04, 14.04
- Debian 8
- CentOS 7

Please open an issue in the Github repository if you run into any packaging related issues. Thank you!

**Ubuntu 14.04**

Download and install `graylog-collector-latest-repository-ubuntu14.04_latest.deb` via `dpkg(1)` and also make sure that the `apt-transport-https` package is installed:

```bash
$ sudo apt-get install graylog-collector
```

**Ubuntu 12.04**

Download and install `graylog-collector-latest-repository-ubuntu12.04_latest.deb` via `dpkg(1)` and also make sure that the `apt-transport-https` package is installed:  

```bash
$ wget https://packages.graylog2.org/repo/packages/graylog-collector-latest-repository-ubuntu12.04_latest.deb
$ sudo apt-get install graylog-collector
```
Debian 8

Download and install graylog-collector-latest-repository-debian8_latest.deb via dpkg(1) and also make sure that the
apt-transport-https package is installed:

```
wget https://packages.graylog2.org/repo/packages/graylog-collector-latest-repository-debian8_latest.deb
sudo dpkg -i graylog-collector-latest-repository-debian8_latest.deb
sudo apt-get install apt-transport-https
sudo apt-get update
sudo apt-get install graylog-collector
```

CentOS 7

Download and install graylog-collector-latest-repository-el7_latest.rpm via rpm(8):

```
sudo rpm -Uvh https://packages.graylog2.org/repo/packages/graylog-collector-latest-repository-el7_latest.rpm
sudo yum install graylog-collector
```

Manual Setup

1. Download the latest collector release. (find download links in the collector repository README)
2. Unzip collector tgz file to target location
3. cp config/collector.conf.example to config/collector.conf
4. Update server-url in collector.conf to correct Graylog server address (required for registration)
5. Update file input configuration with the correct log files
6. Update outputs->gelf-tcp with the correct Graylog server address (required for sending GELF messages)

Note: The collector will not start properly if you do not set the URL or the correct input log files and GELF output configuration

Windows

You need to have Java >= 7 installed to run the collector.

Download a release zip file from the collector repository README. Unzip the collector zip file to target location.
Change into the extracted collector directory and create a collector configuration file in config\collector.conf.
The following configuration file shows a good starting point for Windows systems. It collects the Application, Security, and System event logs. Replace the `<your-graylog-server-ip>` with the IP address of your Graylog server.

Example:

```plaintext
server-url = "http://<your-graylog-server-ip>:9000/api/

inputs {
  win-eventlog-application {
    type = "windows-eventlog"
    source-name = "Application"
    poll-interval = "1s"
  }
  win-eventlog-system {
    type = "windows-eventlog"
    source-name = "System"
    poll-interval = "1s"
  }
  win-eventlog-security {
    type = "windows-eventlog"
    source-name = "Security"
    poll-interval = "1s"
  }
}

outputs {
  gelf-tcp {
    type = "gelf"
    host = "<your-graylog-server-ip>"
    port = 12201
  }
}
```

Start a cmd.exe, change to the collector installation path and execute the following commands to install the collector as Windows service.

Commands:
C:\> cd graylog-collector-0.2.2
C:\graylog-collector-0.2.2> bin\graylog-collector-service.bat install GraylogCollector
C:\graylog-collector-0.2.2> bin\graylog-collector-service.bat start GraylogCollector

Configuration

You will need a configuration file before starting the collector. The configuration file is written in the HOCON format which is a human-optimized version of JSON.

If you choose the operating system installation method, the configuration file defaults to /etc/graylog/collector/collector.conf. For the manual installation method you have to pass the path to the configuration to the start script. (see Running Graylog Collector)

Here is a minimal configuration example that collects logs from the /var/log/syslog file and sends them to a Graylog server:

```plaintext
server-url = "http://10.0.0.1:9000/api/"

inputs {
  syslog {
    type = "file"
    path = "/var/log/syslog"
  }
}

outputs {
  graylog-server {
    type = "gelf"
    host = "10.0.0.1"
    port = 12201
  }
}
```

There are a few global settings available as well as several sections which configure different subsystems of the

9.2. Configuration
Global Settings

server-url - The API URL of the Graylog server  Used to send a heartbeat to the Graylog server.
  (default: "http://localhost:9000/api/")

enable-registration - Enable heartbeat registration  Enables the heartbeat registration with the Graylog server. The collector will not contact the Graylog server API for heartbeat registration if this is set to false.
  (default: true)

collector-id - Unique collector ID setting  The ID used to identify this collector. Can be either a string which is used as ID, or the location of a file if prefixed with file:. If the file does not exist, an ID will be generated and written to that file. If it exists, it is expected to contain a single string without spaces which will be used for the ID.
  (default: "file:config/collector-id")

Input Settings

The input settings need to be nested in a input { } block. Each input has an ID and a type:

inputs {
  syslog {
    type = "file"  // => The input ID
    // => The input type
    ...
  }
}

An input ID needs to be unique among all configured inputs. If there are two inputs with the same ID, the last one wins.

The following input types are available.

File Input

The file input follows files in the file system and reads log data from them.

type  This needs to be set to "file".

path  The path to a file that should be followed.
  Please make sure to escape the \ character in Windows paths:  path = "C:\ProgramFiles\Apache2\logs\www.example.com.access.log"
  (default: none)

path-glob-root  The globbing root directory that should be monitored. See below for an explanation on globbing.
  Please make sure to escape the \ character in Windows paths:  path = "C:\ProgramFiles\Apache2\logs\www.example.com.access.log"
  (default: none)

path-glob-pattern  The globbing pattern. See below for an explanation on globbing.
  (default: none)
content-splitter The content splitter implementation that should be used to detect the end of a log message.

Available content splitters: NEWLINE, PATTERN

See below for an explanation on content splitters.

(default: "NEWLINE")

content-splitter-pattern The pattern that should be used for the PATTERN content splitter.

(default: none)

charset Charset of the content in the configured file(s).

Can be one of the Supported Charsets of the JVM.

(default: "UTF-8")

reader-interval The interval in which the collector tries to read from every configured file. You might set this to a higher value like 1s if you have files which do not change very often to avoid unnecessary work.

(default: "100ms")

Globbing / Wildcards

You might want to configure the collector to read from lots of different files or files which have a different name each time they are rotated. (i.e. time/date in a filename) The file input supports this via the path-glob-root and path-glob-pattern settings.

A usual glob/wildcard string you know from other tools might be /var/log/apache2/**/*.{access,error}.log. This means you are interested in all log files which names end with .access.log or .error.log and which are in a sub directory of /var/log/apache2. Example: /var/log/apache2/example.com/www.example.com.access.log

For compatibility reasons you have to split this string into two parts. The root and the pattern.

Examples:

```
// /var/log/apache2/**/*.{access,error}.log
path-glob-root = "/var/log/apache2"
path-glob-pattern = "**/*.{access,error}.log"

// C:\Program Files\Apache2\logs\*.access.log
path-glob-root = "C:\Program Files\Apache2\logs" // Make sure to escape the \ character in Windows
path-glob-pattern = "*.access.log"
```

The file input will monitor the path-glob-root for new files and checks them against the path-glob-pattern to decide if they should be followed or not.

All available special characters for the glob pattern are documented in the Java docs for the getPathMatcher() method.

Content Splitter

One common problem when reading from plain text log files is to decide when a log message is complete. By default, the file input considers each line in a file to be a separate log message:

```
Jul 15 10:27:08 tumbler anacron[32426]: Job `cron.daily' terminated
Jul 15 10:27:08 tumbler anacron[32426]: Normal exit (1 job run)
```

But there are several cases where this is not correct. Java stack traces are a good example:

```
2015-07-10T11:16:34.486+01:00 WARN [InputBufferImpl] Unable to process event RawMessageEvent{raw=null, uuid=bde580a0-26ec-11e5-9a46-005056b26ca9, encodedLength=350}, sequence 19847516
   java.lang.NullPointerException
```

9.2. Configuration
This should be one message but using a newline separator here will not work because it would generate one log message for each line.

To solve this problem, the file input can be configured to use a PATTERN content splitter. It creates separate log messages based on a regular expression instead of newline characters. A configuration for the stack trace example above could look like this:

```yaml
inputs {
  graylog-server-logs {
    type = "file"
    path = "/var/log/graylog-server/server.log"
    content-splitter = "PATTERN"
    content-splitter-pattern = "^\d{4}-\d{2}-\d{2}T" // Make sure to escape the \ character!
  }
}
```

This instructs the file input to split messages on a timestamp at the beginning of a line. So the first stack trace in the message above will be considered complete once a new timestamp is detected.

### Windows Eventlog Input

The Windows eventlog input can read event logs from Windows systems.

**type** This needs to be set to "windows-eventlog".

**source-name** The Windows event log system has several different sources from which events can be read.

- Common source names: Application, System, Security
  (default: "Application")

**poll-interval** This controls how often the Windows event log should be polled for new events.
  (default: "1s")

Example:

```yaml
inputs {
  win-eventlog-application {
    type = "windows-eventlog"
    source-name = "Application"
    poll-interval = "1s"
  }
}
```
Output Settings

The output settings need to be nested in a `output { }` block. Each output has an ID and a type:

```
outputs {
  graylog-server {  // => The output ID
    type = "gelf"  // => The output type
    ...
  }
}
```

An output ID needs to be unique among all configured outputs. If there are two outputs with the same ID, the last one wins.

The following output types are available.

**GELF Output**

The GELF output sends log messages to a GELF TCP input on a Graylog server.

- **Type** This needs to be set to "gelf".
- **Host** Hostname or IP address of the Graylog server.
  - (default: none)
- **Port** Port of the GELF TCP input on the Graylog server host.
  - (default: none)
- **Client-TLS** Enables TLS for the connection to the GELF TCP input. Requires a TLS-enabled GELF TCP input on the Graylog server. (default: false)
- **Client-TLS-Cert-Chain-File** Path to a TLS certificate chain file. If not set, the default certificate chain of the JVM will be used.
  - (default: none)
- **Client-TLS-Verify-Cert** Verify the TLS certificate of the GELF TCP input on the Graylog server.
  - You might have to disable this if you are using a self-signed certificate for the GELF input and do not have any certificate chain file.
  - (default: true)
- **Client-Queue-Size** The GELF client library that is used for this output has an internal queue of messages. This option configures the size of this queue.
  - (default: 512)
- **Client-Connect-Timeout** TCP connection timeout to the GELF input on the Graylog server.
  - (default: 5000)
- **Client-Reconnect-Delay** The delay before the output tries to reconnect to the GELF input on the Graylog server.
  - (default: 1000)
client-tcp-no-delay  Sets the TCP_NODELAY option on the TCP socket that connects to the GELF input.
(default: true)

client-send-buffer-size  Sets the TCP send buffer size for the connection to the GELF input.
  It uses the JVM default for the operating system if set to -1.
  (default: -1)

STDOUT Output

The STDOUT output prints the string representation of each message to STDOUT. This can be useful for debugging purposes but should be disabled in production.

type  This needs to be set to "stdout".

Static Message Fields

Sometimes it is useful to be able to add some static field to a message. This can help selecting extractors to run on the server, simplify stream routing and can make searching/filtering for those messages easier.

Every collector input can be configured with a message-fields option which takes key-value pairs. The key needs to be a string, the value can be a string or a number.

Example:

```yaml
inputs {
  apache-logs {
    type = "file"
    path = "/var/log/apache2/access.log"
    message-fields = {
      "program" = "apache2"
      "priority" = 3
    }
  }
}
```

Each static message field will end up in the GELF message and shows up in the web interface as a separate field.

An input might overwrite a message field defined in the input configuration. For example the file input always sets a source_file field with the path to the file where the message has been read from. If you configure a source_file message field, it will be overwritten by the input.

Input/Output Routing

Every message that gets read by the configured inputs will be routed to every configured output. If you have two file inputs and two GELF outputs, every message will be received by both outputs. You might want to send some logs to only one output or have one output only accept logs from a certain input, though.

The collector provides two options for inputs and outputs which can be used to influence the message routing.

Inputs have a outputs option and outputs have a inputs option. Both take a comma separated list of input/output IDs.

Example:
Routing for this config:

- **apache-logs** messages will only go to gelf-1 and gelf-2 outputs.
- **auth-log** messages will go to gelf-1 and gelf-2 outputs.
- **syslog** messages will go to all outputs.
- **console** output will only receive messages from syslog input.

<table>
<thead>
<tr>
<th>inputs</th>
<th>outputs</th>
<th>gelf-1</th>
<th>gelf-2</th>
<th>console</th>
</tr>
</thead>
<tbody>
<tr>
<td>apache-logs</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>auth-log</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>syslog</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This is pretty powerful but might get confusing when inputs and outputs have the routing fields. This is how it is implemented in pseudo-code:

```javascript
var message = Object(message)
var output = Object(gelf-output)

if empty(output.inputs) AND empty(message.outputs)
    // No output routing configured, write the message to the output.
    output.write(message)
```

9.2. Configuration
else if output.inputs.contains(message.inputId) OR message.outputs.contains(output.id)

    // Either the input that generated the message has the output ID in its "outputs" field
    // or the output has the ID of the input that generated the message in its "inputs" field.
    output.write(message)

end

Running Graylog Collector

You will need a configuration file before starting the collector. See the configuration documentation above for detailed instructions on how to configure it.

Linux/Unix

The start method for the collector depends on the installation method you choose.

Operating System Package

We ship startup scripts in our OS packages that use the startup method of the particular operating system.

<table>
<thead>
<tr>
<th>OS</th>
<th>Init System</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ubuntu</td>
<td>upstart</td>
<td>sudo start graylog-collector</td>
</tr>
<tr>
<td>Debian</td>
<td>systemd</td>
<td>sudo systemctl start graylog-collector</td>
</tr>
<tr>
<td>CentOS</td>
<td>systemd</td>
<td>sudo systemctl start graylog-collector</td>
</tr>
</tbody>
</table>

Manual Setup

If you use the manual setup, the location of the start script depends on where you extracted the collector.

Example:

$ bin/graylog-collector run -f config/collector.conf

Windows

You probably want to run the collector as Windows service as described in the Windows installation section above. If you want to run it from the command line, run the following commands.

Make sure you have a valid configuration file in config\collector.conf.

Commands:

C:\> cd graylog-collector-0.2.2
C:\graylog-collector-0.2.2> bin\graylog-collector.bat run -f config\collector.conf
Collector Status

Once the collector has been deployed successfully, you can check on the status from the Graylog UI.

You can reach the collector status overview page this way:

1. Log into Graylog Web Interface
2. Navigate to System / Collectors
3. Click Collectors

Troubleshooting

Check the standard output of the collector process for any error messages or warnings. Messages not arriving in your Graylog cluster? Check possible firewalls and the network connection.

Command Line Options

Linux/Unix

The collector offers the following command line options:

```
usage: graylog-collector <command> [args]
```

The most commonly used graylog-collector commands are:

- `help` Display help information
- `run` Start the collector
- `version` Show version information on STDOUT

See `graylog-collector help <command>` for more information on a specific command.

NAME

graylog-collector run - Start the collector

SYNOPSIS

graylog-collector run -f <configFile>

OPTIONS

- `-f <configFile>` Path to configuration file.

Correctly Configured Collector Log Sample

This is the `STDOUT` output of a healthy collector starting:

```
2015-05-12T16:00:10.841+0200 INFO [main] o.graylog.collector.cli.commands.Run - Starting Collector v0.2.0-SNAPSHOT (commit a2ad8c8)
2015-05-12T16:00:11.489+0200 INFO [main] o.g.collector.utils.CollectorId - Collector ID: cf4734f7-01d6-4974-a957-cb71bbd826b7
2015-05-12T16:00:11.513+0200 INFO [main] o.graylog.collector.cli.commands.Run - Service RUNNING: MetricService [RUNNING]
2015-05-12T16:00:11.515+0200 INFO [main] o.graylog.collector.cli.commands.Run - Service RUNNING: FileInput{id='local-syslog', path='/var/log/syslog', charset='UTF-8', outputs='', content-splitter='NEWLINE'}
2015-05-12T16:00:11.516+0200 INFO [main] o.graylog.collector.cli.commands.Run - Service RUNNING: ...
```

2015-05-12T16:00:11.516+0200 INFO [main] o.graylog.collector.cli.commands.Run - Service RUNNING: HeartbeatService [RUNNING]

2015-05-12T16:00:11.516+0200 INFO [main] o.graylog.collector.cli.commands.Run - Service RUNNING: StdoutOutput{id='console', inputs=''}
Troubleshooting

Unable to send heartbeat

The collector registers with your Graylog server on a regular basis to make sure it shows up on the Collectors page in the Graylog web interface. This registration can fail if the collector cannot connect to the server via HTTP on port 9000:

```
2015-06-06T10:45:14.964+0200 WARN [HeartbeatService RUNNING] collector.heartbeat.HeartbeatService - Unable to send heartbeat to Graylog server: ConnectException: Connection refused
```

Possible solutions

- Make sure the server REST API is configured to listen on a reachable IP address. Change the “rest_listen_uri” setting in the Graylog server config to this: `rest_listen_uri = http://0.0.0.0:9000/api/`
- Correctly configure any firewalls between the collector and the server to allow HTTP traffic to port 9000.
Search query language

Syntax

The search syntax is very close to the Lucene syntax. By default all message fields are included in the search if you don’t specify a message field to search in.

Messages that include the term ssh:

```
ssh
```

Messages that include the term ssh or login:

```
ssh login
```

Messages that include the exact phrase ssh login:

```
"ssh login"
```

Messages where the field type includes ssh:

```
type:ssh
```

Messages where the field type includes ssh or login:

```
type:(ssh login)
```

Messages where the field type includes the exact phrase ssh login:

```
type:"ssh login"
```

Messages that do not have the field type:

```
_missing_:type
```

Messages that have the field type:

```
_exists_:type
```

By default all terms or phrases are OR connected so all messages that have at least one hit are returned. You can use **Boolean operators and groups** for control over this:

```
"ssh login" AND source:example.org
("ssh login" AND (source:example.org OR source:another.example.org)) OR _exists_:always_find_me
```
You can also use the NOT operator:

```
"ssh login" AND NOT source:example.org
NOT example.org
```

Note that AND, OR, and NOT are case sensitive and must be typed in all upper-case.

**Wildcards:** Use ? to replace a single character or * to replace zero or more characters:

```
source:*.org
source:exam?le.org
source:exam?le.*
```

Note that leading wildcards are disabled to avoid excessive memory consumption! You can enable them in your Graylog configuration file:

```
allow_leading_wildcard_searches = true
```

Also note that message, full_message, and source are the only fields that are being analyzed by default. While wildcard searches (using * and ?) work on all indexed fields, analyzed fields will behave a little bit different. See wildcard and regexp queries for details.

**Fuzziness:** You can search for similar terms:

```
ssh logni~
source:exmaple.org~
```

This example is using the Damerau–Levenshtein distance with a default distance of 2 and will match “ssh login” and “example.org” (intentionally misspelled in the query).

You can change the distance like this:

```
source:exmaple.org-1
```

You can also use the fuzzyness operator to do a **proximity** search where the terms in a phrase can have different/fuzzy distances from each other and don’t have to be in the defined order:

```
"foo bar"-5
```

Numeric fields support **range queries**. Ranges in square brackets are inclusive, curly brackets are exclusive and can even be combined:

```
http_response_code:[500 TO 504]
http_response_code:{400 TO 404}
bytes:{0 TO 64]
http_response_code:[0 TO 64}
```

You can also do searches with one side unbounded:

```
http_response_code:>400
http_response_code:<400
http_response_code:>=400
http_response_code:<=400
```

It is also possible to combine unbounded range operators:

```
http_response_code:>({=400 AND <500)
```

**Escaping**

The following characters must be escaped with a backslash:
The time frame selector defines in what time range to search in. It offers three different ways of selecting a time range and is vital for search speed: If you know you are only interested in messages of the last hour, only search in that time frame. This will make Graylog search in relevant indices only and greatly reduce system load and required resources.

**Relative time frame selector**

The relative time frame selector lets you look for messages from the selected option to the time you hit the search button. The selector offers a wide set of relative time frames that fit most of your search needs.

**Absolute time frame selector**

When you know exactly the boundaries of your search, you want to use the absolute time frame selector. Simply introduce the dates and times for the search manually or click in the input field to open up a calendar where you can choose the day with your mouse.

**Keyword time frame selector**

Graylog offers a keyword time frame selector that allows you to specify the time frame for the search in natural language like last hour or last 90 days. The web interface shows a preview of the two actual timestamps that will be used for the search.

Here are a few examples for possible values.
• “last month” searches between one month ago and now
• “4 hours ago” searches between four hours ago and now
• “1st of april to 2 days ago” searches between 1st of April and 2 days ago
• “yesterday midnight +0200 to today midnight +0200” searches between yesterday midnight and today midnight in timezone +0200 - will be 22:00 in UTC

The time frame is parsed using the natty natural language parser. Please consult its documentation for details.

Saved searches

Sometimes you may want to search a specific search configuration to be used later. Graylog provides a saved search functionality to accomplish exactly that.

Once you submitted your search, selected the fields you want to show from the search sidebar, and chosen a resolution for the histogram, click on the Save search criteria button on the sidebar.

Give a name to the current search and click on save. When you want to use the saved search later on, you only need to select it from the saved search selector.

Of course, you can always update the selected fields or name of your saved search. To do so, select the saved search from the saved search selector, update the field selection or histogram resolution, and click on Saved search -> Update search criteria. It is also possible to delete the saved search by selecting Saved search -> Delete saved search.
Histogram

The search page includes a search result histogram, where you can view in a concise way the number of messages received grouped by a certain time period that Graylog will adjust for you.

The histogram also allows you to further narrow down the cause for an issue:

- Delimit the search time range by brushing over the histogram. Just click and drag with your mouse over the chart to select the time range you want to use, and click on the search button to perform that search
- See the time where alerts are triggered in the graph annotations. If you are searching in a stream, you will only see alerts related to that stream
Analysis

Graylog provides several tools to analyze your search results. It is possible to save these analysis into dashboards, so you can check them over time in a more convenient way. To analyze a field from your search results, expand the field in the search sidebar and click on the button of the analysis you want to perform.

Field statistics

Compute different statistics on your fields, to help you better summarize and understand the data in them. The statistical information consist of: total, mean, minimum, maximum, standard deviation, variance, sum, and cardinality. On non-numeric fields, you can only see the total amount of messages containing that field, and the cardinality of the field, i.e. the number of unique values it has.

Quick values

Quick values helps you to find out the distribution of values for a field. Alongside a graphic representation of the common values contained in a field, Graylog will display a table with all different values, allowing you to see the number of times they appear. You can include any value in your search query by clicking on the magnifying glass icon located in the value row.
Field graphs

You can create field graphs for any numeric field, by clicking on the Generate chart button in the search sidebar. Using the options in the Customize menu on top of the field graph, you can change the statistical function used in the graph, the kind of graph to use to represent the values, the graph interpolation, as well as the time resolution.

Once you have customized some field graphs, you can also combine them by dragging them from the hamburger icon on the top corner of the graph, and dropping them into another field graph. You can see the location of the hamburger icon and the end result in the following screenshots:
Field graphs appear every time you perform a search, allowing you to compare data, or combine graphs coming from different streams.

**Decorators**

Decorators allow you to alter message fields during search time automatically, while *preserving the unmodified message on disk*. Decorators are specially useful to make some data in your fields more readable, combine data in some field, or add new fields with more information about the message. As decorators are configured per stream (including the *default stream*), you are also able to present a single message in different streams differently.

As changes made by decorators are not persisted, you cannot search for decorated values or use field analyzers on them. You can still use those features in the original non-decorated fields.

Decorators are applied on a stream-level, and are shared among all users capable of accessing a stream, so all users can share the same results and benefit from the advantages decorators add.

Graylog includes some message decorators out of the box, but you can add new ones from pipelines or by writing your own as plugins.

In order to apply decorators to your search results, click on the *Decorators* tab in your search sidebar, select the decorator you want to apply from the dropdown, and click on *Apply*. Once you save your changes, the search results will already contain the decorated values.
10.6. Decorators

Search result
Found 1,711 messages in 79 ms, searched in 2 indices.

Add count to dashboard ▼  Save search criteria
More actions ▼

Fields  Decorators

Syslog Severity Mapper ▼  Apply

What are message decorators?

✎ Syslog Severity Mapper

• source_field: level
• target_field: level

✎ Format String

• format_string: Request to
  ${controller}#${action} finished in
  ${took_ms}ms with code ${http_response_code}
  require_all_fields: true
  target_field: results
When you apply multiple decorators to the same search results, you can change the order in which they are applied at any time by using drag and drop in the decorator list.

Syslog severity mapper

The syslog severity mapper decorator lets you convert the numeric syslog level of syslog messages, to a human readable string. For example, applying the decorator to the level field in your logs would convert the syslog level 4 to Warning (4).

To apply a syslog severity mapper decorator, you need to provide the following data:

- **Source field**: Field containing the numeric syslog level
- **Target field**: Field to store the human readable string. It can be the same one as the source field, if you wish to replace the numeric value on your search results

Format string

The format string decorator provides a simple way of combining several fields into one. It can also be used to modify the content of a field in, without altering the stored result in Elasticsearch.

To apply a format string decorator you need to provide the following data:

- **Format string**: Pattern used to format the resulting string. You can provide fields in the message by enclosing them in $(). E.g. ${source} will add the contents of the source message field into the resulting string
- **Target field**: Field to store the resulting value
- **Require all fields** (optional): Check this box to only format the string when all other fields are present

For example, using the format string Request to ${controller}#${action} finished in ${took_ms}ms with code ${http_response_code}, could produce the text Request to PostsController#show finished in 57ms with code 200, and make it visible in one of the message fields in your search results.

Pipeline Decorator

The pipeline decorator provides a way to decorate messages by processing them with an existing processing pipeline. In contrast to using a processing pipeline, changes done to the message by the pipeline are not persisted. Instead, the pipeline is used at search time to modify the presentation of the message.

The prerequisite of using the pipeline decorator is that an existing pipeline is required.

**Note**: Please take note, that the pipeline you use for decoration should not be connected to a stream. This would mean that it is run twice (during indexing and search time) for each message, effectively rendering the second run useless.

When you are done creating a pipeline, you can now add a decorator using it on any number of streams. In order to create one, you proceed just like for any other decorator type, by clicking on the Decorator sidebar, selecting the type (“Pipeline Processor Decorator” in this case) and clicking the Apply button next to one.
Upon clicking *Apply*, the pipeline to be used for decorating can be selected.

After selecting a pipeline and clicking *Save*, you are already set creating a new pipeline decorator.

**Debugging decorators**

When a message is not decorated as expected, or you need to know how it looked like originally, you can see all changes that were done during decoration by clicking “Show changes” in the message details.
In this view, deleted content is shown in red, while added content is shown in green. This means that added fields will have a single green entry, removed fields a single red entry and modified fields will have two entries, a red and a green one.

Further functionality

If the existing decorators are not sufficient for your needs, you can either search the Graylog marketplace, or write your own decorator.

Export results as CSV

It is also possible to export the results of your search as a CSV document. To do so, select all fields you want to export in the search sidebar, click on the More actions button, and select Export as CSV.
**Hint:** Some Graylog inputs keep the original message in the the `full_message` field. If you need to export the original message, you can do so by clicking on the List all fields link at the bottom of the sidebar, and then selecting the `full_message` field.

**Warning:** Exporting results to a CSV will **not** preserve sorting because Graylog is using the virtual `_doc` field to “sort” documents for performance reasons. If you need to have the exported data ordered you will need to either make a scroll query to ElasticSearch and process it after, or to download the file and post process it via other means.

### Search result highlighting

Graylog supports search result highlighting since v0.20.2:
Enabling/Disabling search result highlighting

Using search result highlighting will result in slightly higher resource consumption of searches. You can enable and disable it using a configuration parameter in the `graylog.conf` of your Graylog nodes:

```
allow_highlighting = true
```

Search configuration

Graylog allows customizing the options allowed to search queries, like limiting the time range users can select or configuring the list of displayed relative time ranges.
All search configuration settings can be customized using the web interface on the System -> Configurations page in the Search configuration section.

**Query time range limit**

Sometimes the amount of data stored in Graylog is quite big and spans a wide time range (e. g. multiple years). In order to prevent normal users from accidentally running search queries which could use up lots of resources, it is possible to limit the time range that users are allowed to search in.

Using this feature, the time range of a search query exceeding the configured query time range limit will automatically be adapted to the given limit.
The query time range limit is a *duration* formatted according to ISO 8601 following the basic format `P<date>T<time>` with the following rules:

<table>
<thead>
<tr>
<th>Designator</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>P</td>
<td>Duration designator (for period) placed at the start of the duration representation</td>
</tr>
<tr>
<td>Y</td>
<td>Year designator that follows the value for the number of years</td>
</tr>
<tr>
<td>M</td>
<td>Month designator that follows the value for the number of months</td>
</tr>
<tr>
<td>W</td>
<td>Week designator that follows the value for the number of weeks</td>
</tr>
<tr>
<td>D</td>
<td>Day designator that follows the value for the number of days</td>
</tr>
<tr>
<td>T</td>
<td>Time designator that precedes the time components of the representation</td>
</tr>
<tr>
<td>H</td>
<td>Hour designator that follows the value for the number of hours</td>
</tr>
<tr>
<td>M</td>
<td>Minute designator that follows the value for the number of minutes</td>
</tr>
<tr>
<td>S</td>
<td>Second designator that follows the value for the number of seconds</td>
</tr>
</tbody>
</table>

Examples:

<table>
<thead>
<tr>
<th>ISO 8601 duration</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>P30D</td>
<td>30 days</td>
</tr>
<tr>
<td>PT1H</td>
<td>1 hour</td>
</tr>
<tr>
<td>P1DT12H</td>
<td>1 day and 12 hours</td>
</tr>
</tbody>
</table>

More details about the format of ISO 8601 durations can be found on Wikipedia.

**Relative time ranges**

The list of time ranges displayed in the *Relative time frame selector* can be configured, too. It consists of a list of *ISO 8601* durations which the users can select on the search page.
### Update Search Configuration

- **Enable query limit**

**Relative Timerange Options**

Configure the available options for the *relative* time range selector as **ISO8601 duration**

<table>
<thead>
<tr>
<th>Duration</th>
<th>Value</th>
<th>Search Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>PT5M</td>
<td>5</td>
<td>Search in the last 5 minutes</td>
</tr>
<tr>
<td>PT15M</td>
<td>15</td>
<td>Search in the last 15 minutes</td>
</tr>
<tr>
<td>PT30M</td>
<td>30</td>
<td>Search in the last 30 minutes</td>
</tr>
<tr>
<td>PT1H</td>
<td>1h</td>
<td>Search in the last 1 hour</td>
</tr>
<tr>
<td>PT2H</td>
<td>2h</td>
<td>Search in the last 2 hours</td>
</tr>
<tr>
<td>PT8H</td>
<td>8h</td>
<td>Search in the last 8 hours</td>
</tr>
<tr>
<td>P1D</td>
<td>1d</td>
<td>Search in the last 1 day</td>
</tr>
<tr>
<td>P2D</td>
<td>2d</td>
<td>Search in the last 2 days</td>
</tr>
<tr>
<td>P5D</td>
<td>5d</td>
<td>Search in the last 5 days</td>
</tr>
<tr>
<td>P7D</td>
<td>7d</td>
<td>Search in the last 7 days</td>
</tr>
</tbody>
</table>

10.9. Search configuration
CHAPTER 11

Streams

What are streams?

The Graylog streams are a mechanism to route messages into categories in realtime while they are processed. You define rules that instruct Graylog which message to route into which streams. Imagine sending these three messages to Graylog:

```
message: INSERT failed (out of disk space)
level: 3 (error)
source: database-host-1

message: Added user 'foo'.
level: 6 (informational)
source: database-host-2

message: smtp ERR: remote closed the connection
level: 3 (error)
source: application-x
```

One of the many things that you could do with streams is creating a stream called *Database errors* that is catching every error message from one of your database hosts.

Create a new stream with these rules, selecting the option to match all rules:

- **Field** `level` **must be greater than** 4
- **Field** `source` **must match regular expression** `^database-host-\d+$`

This will route every new message with a level higher than *WARN* and a source that matches the database host regular expression into the stream.

A message will be routed into every stream that has all (or any) of its rules matching. This means that a message can be part of many streams and not just one.

The stream is now appearing in the streams list and a click on its title will show you all database errors.

Streams can be used to be alerted in case certain condition happens. We cover more topics related to alerts in *Alerts*.

What’s the difference to saved searches?

The biggest difference is that streams are processed in realtime. This allows realtime alerting and forwarding to other systems. Imagine forwarding your database errors to another system or writing them to a file by regularly reading them from the message storage. Realtime streams do this much better.
Another difference is that searches for complex stream rule sets are always comparably cheap to perform because a message is tagged with stream IDs when processed. A search for Graylog internally always looks like this, no matter how many stream rules you have configured:

```
streams:[STREAM_ID]
```

Building a query with all rules would cause significantly higher load on the message storage.

## How do I create a stream?

1. Navigate to the streams section from the top navigation bar.
2. Click “Create stream”.
3. Save the stream after entering a name and a description. For example All error messages and Catching all error messages from all sources. The stream is now saved but not yet activated.
4. Click on “Edit rules” for the stream you just created. That will open a page where you can manage and test stream rules.
5. Choose how you want to evaluate the stream rules to decide which messages go into the stream:
   - A message must match all of the following rules (logical AND): Messages will only be routed into the stream if all rules in the stream are fulfilled. This is the default behavior
   - A message must match at least one of the following rules (logical OR): Messages will be routed into the stream if one or more rules in the stream are fulfilled
6. Add stream rules, by indicating the field that you want to check, and the condition that should satisfy. Try the rules against some messages by loading them from an input or manually giving a message ID. Once you are satisfied with the results, click on “I’m done”.
7. The stream is still paused, click on the “Start stream” button to activate the stream.

## Index Sets

For starters, you should read Index model for a comprehensive description of the index set functionality in Graylog.

Every stream is assigned to an index set which controls how messages routed into that stream are being stored into Elasticsearch. The stream overview in the web interface shows the assigned index set for each stream.
Index sets can be assigned to a stream when creating the stream and changed later when editing the stream settings.

**Important:** Graylog will not automatically copy messages into new Elasticsearch indices if another index set is being assigned to a stream.

![Graylog creating stream](image)

Graylog routes every message into the **All messages** stream by default, unless the message is removed from this stream with a pipeline rule (see *Processing Pipelines*) or it’s routed into a stream marked with **Remove matches from ‘All messages’ stream**.

The latter is useful if messages should be stored with different settings than the ones in the **Default index set**, for example web server access logs should only be stored for 4 weeks while all other messages should be stored for 1 year.

**Storage requirements**

Graylog writes messages once for each index set into Elasticsearch. This means that if all streams are using the **Default index set**, each message will be written exactly once into Elasticsearch, no matter into how many streams the message has been sent. This can be thought of a kind of de-duplication.

If some streams use other index sets and the **Remove matches from ‘All messages’ stream** setting is not enabled, messages will be written into Elasticsearch at least twice, once for the **Default index set** and once for the assigned index set. This means that the same message will be stored in two or more indices in Elasticsearch with different index settings.

Unless you explicitly want to store messages multiple times in different Elasticsearch indices, either assign the **Default index set** to the respective streams or enable the **Remove matches from ‘All messages’ stream** setting for the respective streams.

**Outputs**

The stream output system allows you to forward every message that is routed into a stream to other destinations.

Outputs are managed globally (like message inputs) and not for single streams. You can create new outputs and activate them for as many streams as you like. This way you can configure a forwarding destination once and select multiple streams to use it.

Graylog ships with default outputs and can be extended with *Plugins*.
Use cases

These are a few example use cases for streams:

- Forward a subset of messages to other data analysis or BI systems to reduce their license costs.
- Monitor exception or error rates in your whole environment and broken down per subsystem.
- Get a list of all failed SSH logins and use quick values to analyze which user names were affected.
- Catch all HTTP POST requests to /login that were answered with a HTTP 302 and route them into a stream called Successful user logins. Now get a chart of when users logged in and use quick values to get a list of users that performed the most logins in the search time frame.

How are streams processed internally?

Every message that comes in is matched against the rules of a stream. For messages satisfying all or at least one of the stream rules (as configured in the stream), the internal ID of that stream is stored in the streams array of the processed message.

All analysis methods and searches that are bound to streams can now easily narrow their operation by searching with a streams:[STREAM_ID] limit. This is done automatically by Graylog and does not have to be provided by the user.
Stream Processing Runtime Limits

An important step during the processing of a message is the stream classification. Every message is matched against the user-configured stream rules. The message is added to the stream if all or any rules of a stream matches, depending on what the user chose. Applying stream rules is done during the indexing of a message only, so the amount of time spent for the classification of a message is crucial for the overall performance and message throughput the system can handle.

There are certain scenarios when a stream rule takes very long to match. When this happens for a number of messages, message processing can stall, messages waiting for processing accumulate in memory and the whole system could become non-responsive. Messages are lost and manual intervention would be necessary. This is the worst case scenario.
To prevent this, the runtime of stream rule matching is limited. When it is taking longer than the configured runtime limit, the process of matching this exact message against the rules of this specific stream is aborted. Message processing in general and for this specific message continues though. As the runtime limit needs to be configured pretty high (usually a magnitude higher as a regular stream rule match takes), any excess of it is considered a fault and is recorded for this stream. If the number of recorded faults for a single stream is higher than a configured threshold, the stream rule set of this stream is considered faulty and the stream is disabled. This is done to protect the overall stability and performance of message processing. Obviously, this is a tradeoff and based on the assumption, that the total loss of one or more messages is worse than a loss of stream classification for these.

There are scenarios where this might not be applicable or even detrimental. If there is a high fluctuation of the message load including situations where the message load is much higher than the system can handle, overall stream matching can take longer than the configured timeout. If this happens repeatedly, all streams get disabled. This is a clear indicator that your system is overutilized and not able to handle the peak message load.

**How to configure the timeout values if the defaults do not match**

There are two configuration variables in the configuration file of the server, which influence the behavior of this functionality.

- `stream_processing_timeout` defines the maximum amount of time the rules of a stream are able to spend. When this is exceeded, stream rule matching for this stream is aborted and a fault is recorded. This setting is defined in milliseconds, the default is 2000 (2 seconds).

- `stream_processing_max_faults` is the maximum number of times a single stream can exceed this runtime limit. When it happens more often, the stream is disabled until it is manually reenabled. The default for this setting is 3.

**What could cause it?**

If a single stream has been disabled and all others are doing well, the chances are high that one or more stream rules are performing bad under certain circumstances. In most cases, this is related to stream rules which are utilizing regular expressions. For most other stream rules types the general runtime is constant, while it varies very much for regular expressions, influenced by the regular expression itself and the input matched against it. In some special cases, the difference between a match and a non-match of a regular expression can be in the order of 100 or even 1000. This is caused by a phenomenon called *catastrophic backtracking*. There are good write-ups about it on the web which will help you understanding it.

**Summary: How do I solve it?**

1. Check the rules of the stream that is disabled for rules that could take very long (especially regular expressions).
2. Modify or delete those stream rules.
3. Re-enable the stream.

**Programmatic access via the REST API**

Many organisations already run monitoring infrastructure that are able to alert operations staff when incidents are detected. These systems are often capable of either polling for information on a regular schedule or being pushed new alerts – this article describes how to use the Graylog Stream Alert API to poll for currently active alerts in order to further process them in third party products.
Checking for currently active alert/triggered conditions

Graylog stream alerts can currently be configured to send emails when one or more of the associated alert conditions evaluate to true. While sending email solves many immediate problems when it comes to alerting, it can be helpful to gain programmatic access to the currently active alerts.

Each stream which has alerts configured also has a list of active alerts, which can potentially be empty if there were no alerts so far. Using the stream’s ID, one can check the current state of the alert conditions associated with the stream using the authenticated API call:

```
GET /streams/<streamid>/alerts/check
```

It returns a description of the configured conditions as well as a count of how many triggered the alert. This data can be used to for example send SNMP traps in other parts of the monitoring system.

Sample JSON return value:

```
{
  "total_triggered": 0,
  "results": [
    {
      "condition": {
        "id": "984d04d5-1791-4500-a17e-cd9621cc2ea7",
        "in_grace": false,
        "created_at": "2014-06-11T12:42:50.312Z",
        "parameters": {
          "field": "one_minute_rate",
          "grace": 1,
          "time": 1,
          "backlog": 0,
          "threshold_type": "lower",
          "type": "mean",
          "threshold": 1
        },
        "creator_user_id": "admin",
        "type": "field_value"
      },
      "triggered": false
    }
  ],
  "calculated_at": "2014-06-12T13:44:20.704Z"
}
```

Note that the result is cached for 30 seconds.

List of already triggered stream alerts

Checking the current state of a stream’s alerts can be useful to trigger alarms in other monitoring systems, but if one wants to send more detailed messages to operations, it can be very helpful to get more information about the current state of the stream, for example the list of all triggered alerts since a certain timestamp.

This information is available per stream using the call:

```
GET /streams/<streamid>/alerts?since=1402460923
```

The since parameter is a unix timestamp value. Its return value could be:
Using this information more detailed messages can be produced, since the response contains more detailed information about the nature of the alert, as well as the number of alerts triggered since the timestamp provided.

Note that currently a maximum of 300 alerts will be returned.

**FAQs**

**Using regular expressions for stream matching**

Stream rules support matching field values using regular expressions. Graylog uses the Java Pattern class to execute regular expressions.

For the individual elements of regular expression syntax, please refer to Oracle’s documentation, however the syntax largely follows the familiar regular expression languages in widespread use today and will be familiar to most.

However, one key question that is often raised is matching a string in case insensitive manner. Java regular expressions are case sensitive by default. Certain flags, such as the one to ignore case sensitivity can either be set in the code, or as an inline flag in the regular expression.

To for example route every message that matches the browser name in the following user agent string:

```
Mozilla/5.0 (Macintosh; Intel Mac OS X 10_9_1) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/32.0.1700.107 Safari/537.36
```

the regular expression `.*applewebkit.*` will not match because it is case sensitive. In order to match the expression using any combination of upper- and lowercase characters use the `(?!)` flag as such:

```
(?!).*applewebkit.*
```

Most of the other flags supported by Java are rarely used in the context of matching stream rules or extractors, but if you need them their use is documented on the same Javadoc page by Oracle.

**Can I add messages to a stream after they were processed and stored?**

No. Currently there is no way to re-process or re-match messages into streams.
Only new messages are routed into the current set of streams.

**Can I write own outputs, alert conditions or notifications?**

Yes. Please refer to the *Plugins* documentation page.
Alerts are always based on streams. You can define conditions that trigger alerts. For example whenever the stream *All production exceptions* has more than 50 messages per minute or when the field *milliseconds* had a too high standard deviation in the last five minutes.

Navigate to the *alerts* section from the top navigation bar to see already configured alerts, alerts that were fired in the past or to configure new alert conditions and notifications.

Graylog ships with default *alert conditions* and *alert notifications*, and both can be extended with *Plugins*.

### Alert states

Graylog alerts are periodical searches that can trigger some notifications when a defined condition is satisfied. Since Graylog 2.2.0, alerts can have two states:

- **Unresolved** Alerts have an unresolved state while the defined condition is satisfied. New alerts are triggered in this state, and they also execute the notifications attached to the stream. These alerts usually require an action on your side.

- **Resolved** Graylog automatically resolves alerts once their alert condition is no longer satisfied. This is the final state of an alert, as Graylog will create a new alert if the alert condition is satisfied again in the future. After an alert is resolved, Graylog will apply the *grace period* you defined in the alert condition, waiting a certain time before creating a new alert for this alert condition.

### Alerts overview

The alerts overview page lets you find out which alerts currently require your attention in an easy way, while also allowing you to check alerts that were triggered in the past and are now resolved.
You can click on an alert name at any time to see more details about it.

**Alert details**

From the alert details page you can quickly check the reason why an alert was triggered, the status and configuration of notifications sent by Graylog, and the search results in the time frame when the alert was unresolved.
Alert timeline

From within the alert details page, you can see a timeline of what occurred since Graylog detected an alert condition was satisfied. This includes the time when Graylog evaluated the condition that triggered the alert, the time when notifications were executed and the results of executing them, and the time when the alert was resolved (if that is the case).

Triggered notifications

Sometimes sending alert notifications may fail for some reason. Graylog includes details of the configured notifications at the time an alert was triggered and the result of executing those notifications, helping you to debug and fix any problems that may arise.

Search results

You can quickly look at messages received while the alert was unresolved from within the alert details page. It is also possible to open that search in the search page, allowing you to further analyse the problem at hand.

Conditions

The first step of managing alerts with Graylog is defining alert conditions.

Alert conditions specify searches that Graylog will execute periodically, and also indicate under which circumstances Graylog should consider those search results as exceptional, triggering an alert in that case.
Click on **Manage conditions** in the **Alerts** section to see your current conditions details, modify them, or add new ones. Clicking on an alert condition’s title will open a detail page where you can also see the notifications that will be executed when the condition is satisfied.

### Alert condition types explained

In this section we explain what the default alert conditions included in Graylog do, and how to configure them. Since Graylog 2.2.0, alert conditions can be extended via **Plugins**, you can find more types in the **Graylog Marketplace** or even create your own.

#### Message count condition

This condition triggers whenever the stream received more than X messages in the last Y minutes. Perfect for example to be alerted when there are many exceptions on your platform. Create a stream that catches every error message and be alerted when that stream exceeds normal throughput levels.

#### Field aggregation condition

Triggers whenever the result of a statistical computation of a numerical message field in the stream is higher or lower than a given threshold. Perfect to monitor performance problems: *Be alerted whenever the standard deviation of the response time of your application was higher than X in the last Y minutes.*

#### Field content condition

This condition triggers whenever the stream received at least one message since the last alert run that has a field set to a given value. *Get an alert when a message with the field ‘type’ set to ‘security’ arrives in the stream.*
**Important:** We do not recommend to run this on analyzed fields like `message` or `full_message` because it is broken down to terms and you might get unexpected alerts. For example a check for `security` would also alert if a message with the field set to `no security` is received because it was broken down to `no` and `security`. This only happens on the analyzed `message` and `full_message` in Graylog.

Please also take note that only a single alert is raised for this condition during the alerting interval, although multiple messages containing the given value may have been received since the last alert.

### Notifications

**Warning:** Starting in Graylog 2.2.0, alert notifications are only triggered once, just when a new alert is created. As long as the alert is unresolved or in grace period, Graylog will not send further notifications. This will help you reducing the noise and annoyance of getting notified way too often when a problem persists for a while. Should your setup require repeated notifications you can enable this during the creation of the alert condition since Graylog 2.2.2.

Notifications (previously known as Alarm Callbacks) enable you to take actions on external systems when an alert is triggered. In this way, you can rely on Graylog to know when something is not right in your logs.

Click on **Manage notifications** in the **Alerts** section to see your current notification details, modify them, test them, or add new ones. Remember that notifications are associated to streams, so **all conditions evaluated in a stream will share the same notifications**.
Alert notifications types explained

In this section we explain what the default alert notifications included in Graylog do, and how to configure them. Alert notifications are meant to be extensible through Plugins, you can find more types in the Graylog Marketplace or even create your own.

**Important:** In previous versions of Graylog (before 2.2.0), the email alarm notification was used, when alert conditions existed for a stream, but no alarm notification had been created before. This has been changed, so that if there is no alarm notification existing for a stream, alerts will be shown in the interface but no other action is performed. To help users coming from earlier version, there is a migration job which is being run once, creating the email alarm notification explicitly for qualifying streams, so the old behavior is preserved.

**Email alert notification**

The email alert notification can be used to send an email to the configured alert receivers when the conditions are triggered.

Three configuration options are available for the alert notification to customize the email that will be sent. The email body and email subject are JMTE templates. JMTE is a minimal template engine that supports variables, loops and conditions. See the JMTE documentation for a language reference.

We expose the following objects to the templates.

- **stream** The stream this alert belongs to.
  - `stream.id` ID of the stream
  - `stream.title` title of the stream
  - `stream.description` stream description

- **stream_url** A string that contains the HTTP URL to the stream.

- **check_result** The check result object for this stream.
  - `check_result.triggeredCondition` string representation of the triggered alert condition
  - `check_result.triggeredAt` date when this condition was triggered
  - `check_result.resultDescription` text that describes the check result

- **backlog** A list of `message` objects. Can be used to iterate over the messages via `foreach`.

- **message** (only available via iteration over the `backlog` object) The message object has several fields with details about the message. When using the `message` object without accessing any fields, the `toString()` method of the underlying Java object is used to display it.
  - `message.id` autogenerated message id
  - `message.message` the actual message text
  - `message.source` the source of the message
  - `message.timestamp` the message timestamp
  - `message.fields` map of key value pairs for all the fields defined in the message

  The `message.fields` fields can be useful to get access to arbitrary fields that are defined in the message. For example `message.fields.full_message` would return the `full_message` of a GELF message.
Editing alert configuration

Title
Notify ops

E-Mail Subject
Graylog alert for stream: ${stream.title}: ${check_result.resultDescription}
The subject of sent out mail alerts

Sender (optional)
graylog@example.org
The sender of sent out mail alerts

E-Mail Body (optional)

```
###########
Alert Description: ${check_result.resultDescription}
Date: ${check_result.triggeredAt}
Stream ID: ${stream.id}
Stream title: ${stream.title}
Stream description: ${stream.description}
Alert Condition Title: ${alertCondition.title}
${if stream_url}Stream URL: ${stream_url}${end}

Triggered condition: ${check_result.triggeredCondition}
```
The template to generate the body from

User Receivers (optional)
Select User Receivers
Graylog usernames that should receive this alert

E-Mail Receivers (optional)

ops@graylog.example.org
E-Mail addresses that should receive this alert

[Save] [Cancel]
HTTP alert notification

The HTTP alert notification lets you configure an endpoint that will be called when the alert is triggered.

Graylog will send a POST request to the notification URL including information about the alert. Here is an example of the payload included in a notification:

```
{
    "check_result": {
        "result_description": "Stream had 2 messages in the last 1 minutes with trigger condition more than 1 messages. (Current grace time: 1 minutes)",
        "triggered_condition": {
            "id": "5e7a9c8d-9bb1-47b6-b8db-4a3a83a25e0c",
            "type": "MESSAGE_COUNT",
            "created_at": "2015-09-10T09:44:10.552Z",
            "creator_user_id": "admin",
            "grace": 1,
            "parameters": {
                "grace": 1,
                "threshold": 1,
                "threshold_type": "more",
                "backlog": 5,
                "time": 1
            },
            "description": "time: 1, threshold_type: more, threshold: 1, grace: 1",
            "type_string": "MESSAGE_COUNT",
            "backlog": 5
        },
        "triggered_at": "2015-09-10T09:45:54.749Z",
        "triggered": true,
        "matching_messages": [
            {
                "index": "graylog2_7",
                "message": "WARN: System is failing",
                "fields": {
                    "gl2_remote_ip": "127.0.0.1",
                    "gl2_remote_port": 56498,
                    "gl2_source_node": "41283fec-36b4-4352-a859-7b3d79846b3c",
                    "gl2_source_input": "55f15092bee8e2841898eb53"
                },
                "id": "b7b08150-57a0-11e5-b2a2-d6b4cd83d1d5",
                "stream_ids": [
                    "55f1509dbee8e2841898eb64"
                ],
                "source": "127.0.0.1",
                "timestamp": "2015-09-10T09:45:49.284Z"
            },
            {
                "index": "graylog2_7",
                "message": "ERROR: This is an example error message",
                "fields": {
                    "gl2_remote_ip": "127.0.0.1",
                    "gl2_remote_port": 56481,
                    "gl2_source_node": "41283fec-36b4-4352-a859-7b3d79846b3c",
                    "gl2_source_input": "55f15092bee8e2841898eb53"
                },
                "id": "afd71342-57a0-11e5-b2a2-d6b4cd83d1d5",
                "stream_ids": [
                    "55f1509dbee8e2841898eb64"
                ],
                "source": "127.0.0.1",
                "timestamp": "2015-09-10T09:45:49.284Z"
            }
        ]
    },
    "triggered_at": "2015-09-10T09:45:54.749Z",
    "triggered": true,
    "matching_messages": []
}
```
"source": "127.0.0.1",
"timestamp": "2015-09-10T09:45:36.116Z"
}
],
"stream": {
  "creator_user_id": "admin",
  "outputs": [],
  "matching_type": "AND",
  "description": "test stream",
  "created_at": "2015-09-10T09:42:53.833Z",
  "disabled": false,
  "rules": [
    {
      "field": "gl2_source_input",
      "stream_id": "55f1509dbee8e2841898eb64",
      "id": "55f150b5bee8e2841898eb7f",
      "type": 1,
      "inverted": false,
      "value": "55f15092bee8e2841898eb53"
    }
  ],
  "alert_conditions": [
    {
      "creator_user_id": "admin",
      "created_at": "2015-09-10T09:44:10.552Z",
      "id": "5e7a9c8d-9bb1-47b6-b8db-4a3a83a25e0c",
      "type": "message_count",
      "parameters": {
        "grace": 1,
        "threshold": 1,
        "threshold_type": "more",
        "backlog": 5,
        "time": 1
      }
    }
  ]
},
"id": "55f1509dbee8e2841898eb64",
"title": "test",
"content_pack": null
]
CHAPTER 13

Dashboards

Why dashboards matter

Using dashboards allows you to build pre-defined views on your data to always have everything important just one click away.

Sometimes it takes domain knowledge to be able to figure out the search queries to get the correct results for your specific applications. People with the required domain knowledge can define the search query once and then display the results on a dashboard to share them with co-workers, managers, or even sales and marketing departments.

This guide will take you through the process of creating dashboards and storing information on them. At the end you will have a dashboard with automatically updating information that you can share with anybody or just a subset of people based on permissions.
How to use dashboards

Creating an empty dashboard

Navigate to the Dashboards section using the link in the top menu bar of your Graylog web interface. The page is listing all dashboards that you are allowed to view. (More on permissions later.) Hit the Create dashboard button to create a new empty dashboard.

The only required information is a title and a description of the new dashboard. Use a specific but not too long title so people can easily see what to expect on the dashboard. The description can be a bit longer and could contain more detailed information about the displayed data or how it is collected.

Hit the Create button to create the dashboard. You should now see your new dashboard on the dashboards overview page. Click on the title of your new dashboard to see it. Next, we will be adding widgets to the dashboard we have just created.

Adding widgets

You should have your empty dashboard in front of you. Let’s add some widgets! You can add search result information to dashboards with a couple of clicks. The following search result types can be added to dashboards:

- Search result counts
- Search result histogram charts
- Statistical values
- Field value charts
• Stacked charts
• Quick values results

You can learn more about the different widget types in Widget types explained.

Once you can see the results of your search, you will see buttons with the “Add to dashboard” text, that will allow you to select the dashboard where the widget will be displayed, and configure the widget.
Examples

It is strongly recommended to read the getting started guide on basic searches and analysis first. This will make the following examples more obvious for you.

• **Top log sources today**
  – Example search: `*`, timeframe: Last 24 hours
  – Expand the `source` field in the the sidebar and hit *Quick values*
  – Add quick values to dashboard

• **Number of exceptions in a given app today**
  – Example search: `source:myapp AND Exception`, timeframe: Last 24 hours
  – Add search result count to dashboard

• **Response time chart of a given app**
  – Example search: `source:myapp2`, any timeframe you want
  – Expand a field representing the response time of requests in the sidebar and hit *Generate chart*
  – Add chart to dashboard
Widgets from streams

You can of course also add widgets from stream search results. Every widget added this way will always be bound to streams. If you have a stream that contains every SSH login you can just search for everything (*) in that stream and store the result count as *SSH logins* on a dashboard.

Result

You should now see widgets on your dashboard. You will learn how to modify the dashboard, and edit widgets in the next chapter.

Widget types explained

Graylog supports a wide variety of widgets that allow you to quickly visualize data from your logs. This section intends to give you some information to better understand each widget type, and how they can help you to see relevant details from the many logs you receive.

Search result counts

This kind of widget includes a count of the number of search results for a given search. It can help you to quickly visualize things like the number of exceptions an application logs, or the number of requests your site receives.
All search result counts created with a relative time frame can additionally display trend information. The trend is calculated by comparing the count for the given time frame, with the one resulting from going further back the same amount of time. For example, to calculate the trend in a search result count with a relative search of 5 minutes ago, Graylog will count the messages in the last 5 minutes, and compare that with the count of the previous 5 minutes.

**Search result histogram charts**

The search result histogram displays a chart using the time frame of your search, graphing the number of search result counts over time. It may help you to visualize how the number of request to your site change over time, or to see how many downloads a file has over time.

Changing the graph resolution, you can decide how much time each bar of the graph represents.

**Statistical values**

You can add to your dashboard any statistical value calculated for a field. This may help you to see the mean time response for your application, or how many unique servers are handling requests to your application, by using the cardinality value of that field. Please refer to *Field statistics* for more information on the available statistical functions and how to display them in your searches.

As with search result counts, you can also add trend information to statistical value widgets created with a relative time frame.

**Field value charts**

To draw an statistical value over time, you can use a field value chart. They could help you to see the evolution of the number of unique users visiting your site in the last week. In the *Field graphs* section we explain how to create these charts and ways you can customize them.

**Stacked charts**

Stacked charts group several field value charts under the same axes. They let you compare different values in a compact way, like the number of visits to two different websites. As explained in *Field graphs*, stacked charts are basically field value charts represented in the same axes.

**Quick values results**

In order to show a list of values a certain field contains and their distribution, you can use a quick value widget. This may help you to see the percentage of failed requests in your application, or which parts of your application experience more problems. Please refer to *Quick values* to see how to request this information in your search result page.

The quick values information can be represented as a pie chart and/or as a table, so you can choose what is the best fit for your needs.

**Modifying dashboards**

You need to unlock dashboards to make any changes to them. Hit the “Unlock/Edit” button in the top right corner of a dashboard to unlock it. You should now see different icons at the bottom of each widget, that allow you to perform more actions.
Unlocked dashboard widgets explained

Unlocked dashboard widgets have two buttons that should be pretty self-explanatory.

- Delete widget
- Edit widget configuration
- Change widget size (when you hover over the widget)

Widget cache times

Widget values are cached in graylog-server by default. This means that the cost of value computation does not grow with every new device or even browser tab displaying a dashboard. Some widgets might need to show real-time information (set cache time to 1 second) and some widgets might be updated way less often (like Top SSH users this month, cache time 10 minutes) to save expensive computation resources.

Repositioning widgets

Just grab a widget with your mouse in unlocked dashboard mode and move it around. Other widgets should adopt and re-position intelligently to make place for the widget you are moving. The positions are automatically saved when dropping a widget.
Resizing widgets

When hovering over a widget, you will see that a gray arrow appears in its bottom-right corner. You can use that icon to resize widgets. Their contents will adapt to the new size automatically!

Dashboard permissions

Graylog users in the Admin role are always allowed to view and edit all dashboards. Users in the Reader role are by default not allowed to view or edit any dashboard.
Navigate to System -> Roles and create a new role that grant the permissions you wish. You can then assign that new role to any users you wish to give dashboard permissions in the System -> Users page.

You can read more about user permissions and roles.

That's it!

Congratulations, you have just gone through the basic principles of Graylog dashboards. Now think about which dashboards to create. We suggest:

- Create dashboards for yourself and your team members
- Create dashboards to share with your manager
- Create dashboards to share with the CIO of your company

Think about which information you need access to frequently. What information could your manager or CIO be interested in? Maybe they want to see how the number of exceptions went down or how your team utilized existing hardware better. The sales team could be interested to see signup rates in realtime and the marketing team will love you for providing insights into low level KPIs that is just a click away.
CHAPTER 14

Extractors

The problem explained

Syslog (RFC3164, RFC5424) is the de facto standard logging protocol since the 1980s and was originally developed as part of the sendmail project. It comes with some annoying shortcomings that we tried to improve in GELF for application logging.

Because syslog has a clear specification in its RFCs it should be possible to parse it relatively easy. Unfortunately there are a lot of devices (especially routers and firewalls) out there that send logs looking like syslog but actually breaking several rules stated in the RFCs. We tried to write a parser that reads all of them as good as possible and failed. Such a loosely defined text message usually breaks the compatibility in the first date field already. Some devices leave out hostnames completely, some use localized time zone names (e. g. “MESZ” instead of “CEST”), and some just omit the current year in the timestamp field.

Then there are devices out there that at least do not claim to send syslog when they don’t but have another completely separate log format that needs to be parsed specifically.

We decided not to write custom message inputs and parsers for all those thousands of devices, formats, firmwares and configuration parameters out there but came up with the concept of Extractors introduced the v0.20.0 series of Graylog.

Graylog extractors explained

The extractors allow you to instruct Graylog nodes about how to extract data from any text in the received message (no matter from which format or if an already extracted field) to message fields. You may already know why structuring data into fields is important if you are using Graylog: There are a lot of analysis possibilities with full text searches but the real power of log analytics unveils when you can run queries like http_response_code:>=500 AND user_id:9001 to get all internal server errors that were triggered by a specific user.

Wouldn’t it be nice to be able to search for all blocked packages of a given source IP or to get a quickterms analysis of recently failed SSH login usernames? Hard to do when all you have is just a single long text message.

Attention: Graylog extractors only work on text fields but won’t be executed for numeric fields or anything other than a string.

Creating extractors is possible via either Graylog REST API calls or from the web interface using a wizard. Select a message input on the System -> Inputs page and hit Manage extractors in the actions menu. The wizard allows you to load a message to test your extractor configuration against. You can extract data using for example regular expressions,
Grok patterns, substrings, or even by splitting the message into tokens by separator characters. The wizard looks like this and should be pretty intuitive:

You can also choose to apply so called converters on the extracted value to for example convert a string consisting of numbers to an integer or double value (important for range searches later), anonymize IP addresses, lower-/uppercase a string, build a hash value, and much more.
Import extractors

The recommended way of importing extractors in Graylog is using *Content packs*. The *Graylog Marketplace* provides access to many content packs that you can easily download and import into your Graylog setup.

You can still import extractors from JSON if you want to. Just copy the JSON extractor export into the import dialog of a message input of the fitting type (every extractor set entry in the directory tells you what type of input to spawn, e.g. syslog, GELF, or Raw/plaintext) and you are good to go. The next messages coming in should already include the extracted fields with possibly converted values.

A message sent by Heroku and received by Graylog with the imported *Heroku* extractor set on a plaintext TCP input looks like this: (look at the extracted fields in the message detail view)
email: a700ec50-0bf-de13-b51a-12313c013090

Received by: Heroku test 2 on: 3c1749a2 / 54.247.
Index: graylog2_1

bytes: 31094
client_ip: 164.177.
connect_time_ms: 0


drain_id: d.158ca49c-4116-451a-9e6e-8158d609eb53
dyno: web.1
facility: local3
heroku_component: router
heroku_source_type: heroku
http_method: get
http_status: 200
level: Info [6]
message: at=info method=GET path=/ host=www.graylog2.org request id=f3ff87e-
Using regular expressions to extract data

Extractors support matching field values using regular expressions. Graylog uses the Java Pattern class to evaluate regular expressions.

For the individual elements of regular expression syntax, please refer to Oracle’s documentation, however the syntax largely follows the familiar regular expression languages in widespread use today and will be familiar to most.

However, one key question that is often raised is matching a string in case insensitive manner. Java regular expressions are case sensitive by default. Certain flags, such as the one to ignore case sensitivity can either be set in the code, or as an inline flag in the regular expression.

For example, to create an extractor that matches the browser name in the following user agent string:

Mozilla/5.0 (Macintosh; Intel Mac OS X 10_9_1) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/32.0.1700.107 Safari/537.36

the regular expression (applewebkit) will not match because it is case sensitive. In order to match the expression using any combination of upper- and lowercase characters use the (?i) flag as such:

(?i)(applewebkit)

Most of the other flags supported by Java are rarely used in the context of matching stream rules or extractors, but if you need them their use is documented on the same Javadoc page by Oracle. One common reason to use regular expression flags in your regular expression is to make use of what is called non-capturing groups. Those are parentheses which only group alternatives, but do not make Graylog extract the data they match and are indicated by (?).

Using Grok patterns to extract data

Graylog also supports the extracting data using the popular Grok language to allow you to make use of your existing patterns.

Grok is a set of regular expressions that can be combined to more complex patterns, allowing to name different parts of the matched groups.

By using Grok patterns, you can extract multiple fields from a message field in a single extractor, which often simplifies specifying extractors.

Simple regular expressions are often sufficient to extract a single word or number from a log line, but if you know the entire structure of a line beforehand, for example for an access log, or the format of a firewall log, using Grok is advantageous.

For example a firewall log line could contain:

len=50824 src=172.17.22.108 sport=829 dst=192.168.70.66 dport=513

We can now create the following patterns on the System/Grok Patterns page in the web interface:

BASE10NUM (?<![0-9.+-])\d+(?:\.[0-9]+)\d+|(?<![0-9.+-])\d+(?:\.[0-9]+)\d+
NUMBER (?<!BASE10NUM)
IPV6 (((0-9A-Fa-f)\1,4)\6)|((0-9A-Fa-f)\1,4)\6)|((0-9A-Fa-f)\1,4)\6)|((0-9A-Fa-f)\1,4)\6)|((0-9A-Fa-f)\1,4)\6)|((0-9A-Fa-f)\1,4)\6)
IPV4 (?<!IPV6)\6)|((0-9A-Fa-f)\1,4)\6)|((0-9A-Fa-f)\1,4)\6)
IP (?!IPV6)\6)|((0-9A-Fa-f)\1,4)\6)
DATA .*?

Then, in the extractor configuration, we can use these patterns to extract the relevant fields from the line:

len=%{NUMBER:length} src=%{IP:srcip} sport=%{NUMBER:srcport} dst=%{IP:dstip} dport=%{NUMBER:dstport}
This will add the relevant extracted fields to our log message, allowing Graylog to search on those individual fields, which can lead to more effective search queries by allowing to specifically look for packets that came from a specific source IP instead of also matching destination IPs if one would only search for the IP across all fields.

If the Grok pattern creates many fields, which can happen if you make use of heavily nested patterns, you can tell Graylog to skip certain fields (and the output of their subpatterns) by naming a field with the special keyword `UNWANTED`.

Let’s say you want to parse a line like:

```
type:44 bytes:34 errors:122
```

but you are only interested in the second number `bytes`. You could use a pattern like:

```
type:%{BASE10NUM:type} bytes:%{BASE10NUM:bytes} errors:%{BASE10NUM:errors}
```

However, this would create three fields named `type`, `bytes`, and `errors`. Even not naming the first and last patterns would still create a field names `BASE10NUM`. In order to ignore fields, but still require matching them use `UNWANTED`:

```
type:%{BASE10NUM:UNWANTED} bytes:%{BASE10NUM:bytes} errors:%{BASE10NUM:UNWANTED}
```

This now creates only a single field called `bytes` while making sure the entire pattern must match.

If you already know the data type of the extracted fields, you can make use of the type conversion feature built into the Graylog Grok library. Going back to the earlier example:

```
len=50824 src=172.17.22.108 sport=829 dst=192.168.70.66 dport=513
```

We know that the content of the field `len` is an integer and would like to make sure it is stored with that data type, so we can later create field graphs with it or access the field’s statistical values, like average etc.

Grok directly supports converting field values by adding `;datatype` at the end of the pattern, like:

```
len=%{NUMBER:length;int} src=%{IP:srcip} sport=%{NUMBER:srcport} dst=%{IP:dstip} dport=%{NUMBER:dstport}
```

The currently supported data types, and their corresponding ranges and values, are:

<table>
<thead>
<tr>
<th>Type</th>
<th>Range</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>byte</td>
<td>-128 ... 127</td>
<td>%{NUMBER:fieldname;byte}</td>
</tr>
<tr>
<td>short</td>
<td>-32768 ... 32767</td>
<td>%{NUMBER:fieldname;short}</td>
</tr>
<tr>
<td>int</td>
<td>-2^31 ... 2^31-1</td>
<td>%{NUMBER:fieldname;int}</td>
</tr>
<tr>
<td>long</td>
<td>-2^63 ... 2^63-1</td>
<td>%{NUMBER:fieldname;long}</td>
</tr>
<tr>
<td>float</td>
<td>32-bit IEEE 754</td>
<td>%{NUMBER:fieldname;float}</td>
</tr>
<tr>
<td>double</td>
<td>64-bit IEEE 754</td>
<td>%{NUMBER:fieldname;double}</td>
</tr>
<tr>
<td>boolean</td>
<td>true, false</td>
<td>%{DATA:fieldname:boolean}</td>
</tr>
<tr>
<td>string</td>
<td>Any UTF-8 string</td>
<td>%{DATA:fieldname:string}</td>
</tr>
<tr>
<td>date</td>
<td>See SimpleDateFormat</td>
<td>%{DATA:timestamp;date;dd/MM/yyyy:HH:mm:ss Z}</td>
</tr>
<tr>
<td>datet ime</td>
<td>Alias for date</td>
<td></td>
</tr>
</tbody>
</table>

There are many resources are the web with useful patterns, and one very helpful tool is the Grok Debugger, which allows you to test your patterns while you develop them.

Graylog uses Java Grok to parse and run Grok patterns.

**Using the JSON extractor**

Since version 1.2, Graylog also supports extracting data from messages sent in JSON format.
Using the JSON extractor is easy: once a Graylog input receives messages in JSON format, you can create an extractor by going to System -> Inputs and clicking on the Manage extractors button for that input. Next, you need to load a message to extract data from, and select the field containing the JSON document. The following page let you add some extra information to tell Graylog how it should extract the information. Let’s illustrate how a message would be extracted with an example message:

```json
{"level": "ERROR", "details": {"message": "This is an example error message", "controller": "IndexController"}}
```

Using the default settings, that message would be extracted into these fields:

- `details.tags`: one, two, three
- `level`: ERROR
- `details.controller`: IndexController
- `details.message`: This is an example error message

In the create extractor page, you can also customize how to separate list of elements, keys, and key/values. It is also possible to flatten JSON structures or expand them into multiple fields, as shown in the example above.

### Automatically extract all key=value pairs

Sometimes you will receive messages like this:

```text
This is a test message with some key/value pairs. key1=value1 some_other_key=foo
```

You might want to extract all key=value pairs into Graylog message fields without having to specify all possible key names or even their order. This is how you can easily do this:

Create a new extractor of type “Copy Input” and select to read from the field `message`. (Or any other string field that contains key=value pairs.) Configure the extractor to store the (copied) field value to the same field. In this case `message`. The trick is to add the “Key=Value pairs to fields” converter as last step. Because we use the “Copy Input” extractor, the converter will run over the complete field you selected and convert all key=value pairs it can find.

This is a screenshot of the complete extractor configuration:
... and this is the resulting message:

```
This is a test message with some key=value pairs. key1=value1 some_other_key=foo
```

This is a test message with some key=value pairs. key1=value1 some_other_key=foo
Normalization

Many log formats are similar to each other, but not quite the same. In particular they often only differ in the names attached to pieces of information.

For example, consider different hardware firewall vendors, whose models log the destination IP in different fields of the message, some use dstip, some dst and yet others use destination-address:

- 2004-10-13 10:37:17 PDT Packet Length=50824, Source address=172.17.22.108, Source port=829, Destination address=192.168.70.66, Destination port=513
- 2004-10-13 10:37:17 PDT len=50824 src=172.17.22.108 sport=829 dst=192.168.70.66 dport=513
- 2004-10-13 10:37:17 PDT length="50824" srcip="172.17.22.108" srcport="829" dstip="192.168.70.66" dstport="513"

You can use one or more non-capturing groups to specify the alternatives of the field names, but still be able to extract the a parentheses group in the regular expression. Remember that Graylog will extract data from the first matched group of the regular expression. An example of a regular expression matching the destination IP field of all those log messages from above is:

```regex
(?:dst|dstip|destination\saddress)="(?<d{1,3}\.d{1,3}\.d{1,3}\.d{1,3})"?
```

This will only extract the IP address without caring about which of the three naming schemes was used in the original log message. This way you don’t have to set up three different extractors.

The standard date converter

Date parser converters for extractors allow you to convert extracted data into timestamps - Usually used to set the timestamp of a message based on some date it contains. Let’s assume we have this message from a network device:

```text
<131>: foo-bar-dc3-org-de01: Mar 12 00:45:38: %LINK-3-UPDOWN: Interface GigabitEthernet3/31, changed
```

Extracting most of the data is not a problem and can be done easily. Using the date in the message (Mar 12 00:45:38) as Graylog message timestamp however needs to be done with a date parser converter.

Use a copy input extractor rule to select the timestamp and apply the Date converter with a format string:

```regex
MMM dd HH:mm:ss
```

(format string table at the end of this page)
Store as field:

```
timestamp
```

Choose a field name. The extracted value will be stored in it. Call it `http_response_code` for example if you are extracting a HTTP response code.

**Copy**  **Cut**

Do you want to copy or cut from source?

Extractor title:

```
Timestamp
```

A descriptive name of this extractor.

Add converter:

```
Date
```

**Add**

**Convert to date type**

Format string:

```
MMM dd HH:mm:ss
```

Please note that you cannot use the cutting feature on standard fields like `message` and `source`.

Create extractor
4765e370-aa42-11e3-a7dd-4c8d79f2b596
Received by Cisco System Messages on fb66b27e / 10.226.163.44
Timestamp: 2014-03-12 00:45:38.000
Index: graylog2_356

facility
  local0
level
  Error [3]
local_facility
  link
local_level
  3
message
  Interface GigabitEthernet0/31, changed state to down
source
  foo-bar-dc3-org-de01
type
  updown
# Standard date converter format string table

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
<th>Presentation</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>era</td>
<td>text</td>
<td>AD</td>
</tr>
<tr>
<td>C</td>
<td>century of era (&gt;=0)</td>
<td>number</td>
<td>20</td>
</tr>
<tr>
<td>Y</td>
<td>year of era (&gt;=0)</td>
<td>year</td>
<td>1996</td>
</tr>
<tr>
<td>x</td>
<td>weekyear</td>
<td>year</td>
<td>1996</td>
</tr>
<tr>
<td>w</td>
<td>week of weekyear</td>
<td>number</td>
<td>27</td>
</tr>
<tr>
<td>e</td>
<td>day of week</td>
<td>number</td>
<td>2</td>
</tr>
<tr>
<td>E</td>
<td>day of week</td>
<td>text</td>
<td>Tuesday; Tue</td>
</tr>
<tr>
<td>y</td>
<td>year</td>
<td>year</td>
<td>1996</td>
</tr>
<tr>
<td>D</td>
<td>day of year</td>
<td>number</td>
<td>189</td>
</tr>
<tr>
<td>M</td>
<td>month of year</td>
<td>month</td>
<td>July; Jul; 07</td>
</tr>
<tr>
<td>d</td>
<td>day of month</td>
<td>number</td>
<td>10</td>
</tr>
<tr>
<td>a</td>
<td>halfday of day</td>
<td>text</td>
<td>PM</td>
</tr>
<tr>
<td>K</td>
<td>hour of halfday (0~11)</td>
<td>number</td>
<td>0</td>
</tr>
<tr>
<td>h</td>
<td>clockhour of halfday (1~12)</td>
<td>number</td>
<td>12</td>
</tr>
<tr>
<td>H</td>
<td>hour of day (0~23)</td>
<td>number</td>
<td>0</td>
</tr>
<tr>
<td>k</td>
<td>clockhour of day (1~24)</td>
<td>number</td>
<td>24</td>
</tr>
<tr>
<td>m</td>
<td>minute of hour</td>
<td>number</td>
<td>30</td>
</tr>
<tr>
<td>s</td>
<td>second of minute</td>
<td>number</td>
<td>55</td>
</tr>
<tr>
<td>S</td>
<td>fraction of second</td>
<td>millis</td>
<td>978</td>
</tr>
<tr>
<td>z</td>
<td>time zone</td>
<td>text</td>
<td>Pacific Standard Time; PST</td>
</tr>
<tr>
<td>Z</td>
<td>time zone offset/id</td>
<td>zone</td>
<td>-0800; -08:00; America/Los_Angeles</td>
</tr>
<tr>
<td>'</td>
<td>escape for text</td>
<td>delimitier</td>
<td>.</td>
</tr>
<tr>
<td>''</td>
<td>single quote</td>
<td>literal</td>
<td>.</td>
</tr>
</tbody>
</table>

# The flexible date converter

Now imagine you had one of those devices that send messages that are not so easy to parse because they do not follow a strict timestamp format. Some network devices for example like to send days of the month without adding a padding 0 for the first 9 days. You’ll have dates like Mar 9 and Mar 10 and end up having problems defining a parser string for that. Or maybe you have something else that is really exotic like just last wednesday as timestamp. The flexible date converter is accepting any text data and tries to build a date from that as good as it can.

Examples:

- **Mar 12**, converted at 12:27:00 UTC in the year 2014: 2014-03-12T12:27:00.000
- **2014-3-12 12:27**: 2014-03-12T12:27:00.000
- **Mar 12 2pm**: 2014-03-12T14:00:00.000

Note that the flexible date converter is using UTC as time zone by default unless you have time zone information in the parsed text or have configured another time zone when adding the flexible date converter to an extractor (see this comprehensive list of time zones available for the flexible date converter).
Graylog’s new processing pipelines plugin allows greater flexibility in routing, blacklisting, modifying and enriching messages as they flow through Graylog.

Pipelines and rules are no longer configuration for pre-build code (like extractors and stream rules are) but are represented as code, much like Drools rules are. This gives them their great flexibility and extensibility.

The language itself is very simple and can be extended by functions, which are fully pluggable.

The following pages introduce the concepts of pipelines, rules, stream connections and the built in functions.

**Pipelines**

**Overview**

Pipelines are the central concept which tie together the processing steps Graylog applies to your messages.

They contain rules and can be connected to streams, allowing fine grained control which processing is done on which kinds of messages.

Because processing rules are simply conditions followed by a list of actions and do not have control flow by themselves, pipelines have one additional concept: stages.

Think of stages as groups of conditions and actions which need to run in order. All stages with the same priority run at the same time across all connected pipelines. They provide the necessary control flow to decide whether or not to run the rest of a pipeline.

**Pipeline structure**

Internally pipelines are represented as code. Let’s have a look at a simple example and understand what each part does:

```plaintext
pipeline "My new pipeline"
stage 1 match all
   rule "has firewall fields"
   rule "from firewall subnet"
stage 2 match either
   rule "geocode IPs"
   rule "anonymize source IPs"
end
```
This code snippet declares a new pipeline with the name `My new pipeline` which declares two stages. Stages are run in the order of their given `priority` and aren’t otherwise named. Stage priorities can be any integer, even negative ones. This allows flexible ordering to run certain rules before or after others, even without modifying existing pipelines, which can come in handy when dealing with changing data formats.

For example, if there was a second pipeline declared, which contained a stage with the priority 0, that would run before either of the ones from the example. The order in which stages are declared is irrelevant, they are sorted according to their priority.

Stages then list the `rule references` they want to be executed as well as declaring whether any or all rules’ conditions need to be satisfied to continue to evaluate the pipeline.

In our example, imagine rule “`has firewall fields`” checks for the presence of two message fields: `src_ip` and `dst_ip` but does not have any actions to run. Then for messages that do not have both fields, the condition would be false, and the pipeline would be aborted after stage 1, because the stage requires that all rules need to be satisfied. Stage 2 would not begin to run in this case. `match either` acts as an OR operator, only requiring a single (or more) rules to match. Actions are still being run for all matching rules, even if the pipeline stops after the stage.

Rules are referenced by their names and can thus be shared among many different pipelines. The intention is to create building blocks which then make it easier to process the data specific to your organization or use case.

Read more about `Rules` in the next section.

### Rules

#### Overview

Rules are the cornerstone of the processing pipelines. They contain the logic about how to change, enrich, route, and drop messages.

To avoid the complexities of a complete programming language, Graylog supports a small rule language to express the processing logic. The rule language is limited on purpose to allow for easier understanding, better runtime optimization and fast learning.

The real work of rules is done in `functions` which are completely pluggable. Graylog already ships with a great number of built-in functions that range from converting data types over string processing, like `substring, regex` etc, to JSON parsing.

We expect that special purpose functions will be written and shared by the community, allow for faster innovation and problem solving than previously possible.

#### Rule structure

Picking up from the previous example in the `Pipelines` section, let’s look at examples of some of the rules we’ve referenced:

```java
rule "has firewall fields"
when
    has_field("src_ip") && has_field("dst_ip")
then
end

rule "from firewall subnet"
when
    cidr_match("10.10.10.0/24", to_ip($message.gl2_remote_ip))
```
Firstly, apart from naming the rule, their structure follows a simple `when, then` pattern. In the `when` clause we specify a boolean expression which is evaluated in the context of the current message in the pipeline. These are the conditions that are being used by the pipeline processor to determine whether to run a rule and collectively whether to continue in a pipeline.

Note that we are already calling the built-in function `has_field` with a field name. In the rule `has firewall fields` we make sure the message contains both `src_ip` as well as `dst_ip` as we want to use them in a later stage.

The rule has no actions to run, because we are only interested in using it as a condition at this point.

The second rule uses another built-in function `cidr_match`. That function takes a CIDR pattern and an IP address. In this case we reference a field from the currently processed message using the message reference syntax `$message`.

The field `gl2_remote_ip` is always set by Graylog upon receiving a message, so we do not check whether that field exists, otherwise we would have used another `has_field` function call to make sure it is there.

However, note the call to `to_ip` around the field reference. This is necessary because the field is stored as a `string` internally. In order to successfully match the CIDR pattern, we need to convert it to an IP address.

This is an important feature of Graylog's rule language, it enforces type safety to ensure that you end up with the data in the correct format. All too often everything is treated as a string, which wastes enormous amounts of cycles to convert data all the time as well as preventing to do proper analysis over the data.

Again we have no actions to immediately run, so the `then` block is empty.

### Data Types

As we have seen in the previous section, we need to make sure to use the proper data types when calling functions.

Graylog's rule language parser rejects invalid use of types, making it safe to write rules.

The six built-in types in Graylog are `string` (a UTF-8 string), `double` (corresponds to Java's `Double`), `long` (Java's `Long`), `boolean` (Java's `Boolean`), `void` (indicating a function has no return value to prevent it being used in a condition) and `ip` (a subset of `InetAddress`), but plugins are free to add additional types as they see fit. The rule processor takes care of ensuring that values and functions agree on the types being used.

Conventionally functions that convert types start with the prefix `to`, please refer to the [Functions index](#) for a list.

### Conditions

In Graylog's rules the `when` clause is a boolean expression, which is evaluated against the processed message.

Expressions support the common boolean operators `AND` (or `& &`), `OR` (or `||`), `NOT` (!) and comparison operators (`<`, `<=`, `>`, `>=`, `==`, `!=`).

Additionally any function that returns a value can be called (e.g. `route_to_stream` does not return a value) but the resulting expression must eventually be a boolean.

The condition must not be empty, but can instead simply use the boolean literal `true` in case you always want to execute the actions inside the rule.

If a condition calls a function which is not present, e.g. due to a missing plugin, the call evaluates to false instead.
Actions

A rule’s then clause contains a list of actions which are evaluated in the order they appear.

There are two different types of actions:

# Function calls # Variable assignments

Function calls look exactly like they do in conditions and all of the functions in the system can be used, including the functions that do not return a value.

Variable assignments have the following form:

```
let name = value;
```

They are useful to avoid recomputing expensive parsing of data, holding on to temporary values or making rules more readable.

Variables need to be defined before they can used and can be accessed using the `name.field` notation in any place where a value is required.

The list of actions can also be empty, turning the rule into a pure condition which can be useful in combination with stages to guide the processing flow.

Stream connections

Overview

Pipelines by themselves do not process any messages. Instead they can be connected to streams, allowing fine grained control over which message types are being processed by which pipelines.

Using the built-in function `route_to_stream` a rule action can cause a message to be routed to the given stream. The pipeline engine will then look up the pipelines connected to that stream and start evaluating the additional pipelines.

To begin processing pipelines the incoming messages must therefore be directed to the initial set of pipelines.

The All messages stream

All messages received by Graylog are initially routed into the All messages stream. You can use this stream as the entry to pipeline processing, allowing incoming messages to be routed to more streams and being processed subsequently.

However, if you prefer to use the original stream matching functionality, you can configure the pipeline processor to run after the message filter chain and connect pipelines to existing streams. This gives you fine grained control over the extraction, conversion and enrichment process.

Functions

Overview

Functions are the means of how to interact with the messages Graylog processes.

They are written in Java and are pluggable, allowing extending the capabilities of Graylog in a simple manner.
Conceptually a function receives parameters, the current message context and returns a value. The data types of its return value and parameters determines where in the rules it can be used. Graylog makes sure the rules are sound from a data type perspective.

A function’s parameters can be passed as named pairs or by position, as long as optional parameters are declared as coming last. The function’s documentation below indicates which parameters are optional.

Let’s look at a small example to illustrate these properties:

```plaintext
rule "function howto"
when
  has_field("transaction_date")
then
  // the following date format assumes there's no time zone in the string
  let new_date = parse_date(to_string($message.transaction_date), "yyyy-MM-dd HH:mm:ss");
  set_field("transaction_year", new_date.year);
end
```

In this example, we check if the current message contains the field `some_date` and then, after converting it to a string, try to parse it according to the format string `yyyy-MM-dd HH:mm:ss`, so for example the string `2016-03-05 14:45:02` would match. Parse date returns a `DateTime` object from the Java Joda-Time library, allowing easier access to the date’s components.

We then add the transaction’s year as a new field to the message.

You’ll note that we haven’t said in which time zone the timestamp is in, but still Graylog had to pick one (Graylog never relies on the local time of your server as that makes it nearly impossible to figure out why date handling came up with its result).

The reason Graylog knows which timezone to pick is because `parse_date` actually takes three parameters rather than the two we’ve given it in this example. The third one is a `String` called `timezone` and is optional. It has a default value of "UTC".

Let’s assume we have another field in the message, called `transaction_timezone`. It is send by the application and contains the time zone ID the transaction was done in (hopefully no application in the world sends its data like this, though):

```plaintext
rule "function howto"
when
  has_field("transaction_date") && has_field("transaction_timezone")
then
  // the following date format assumes there's no time zone in the string
  let new_date = parse_date(
    to_string($message.transaction_date),
    "yyyy-MM-dd HH:mm:ss",
    to_string($message.transaction_timezone)
  );
  set_field("transaction_year", new_date.year);
end
```

Now, if the time zone is specified in the other field, we pass it to `parse_date`.

In this case we only have a single optional parameter, which makes it easy to simply omit it from the end of the function call. However, if there are multiple ones, or if there are many parameters and it gets difficult to keep track of which positions correspond to which parameters, you can also use the named parameter variant of function calls. In this mode the order of the parameters does not matter, but all required ones still need to be there.

In our case the alternative version of calling `parse_date` would look like this:

```plaintext
rule "function howto"
when
end
```

15.4. Functions
has_field("transaction_date") && has_field("transaction_timezone")
then
  // the following date format assumes there's no time zone in the string
  let new_date = parse_date(
    value: to_string($message.transaction_date),
    pattern: "yyyy-MM-dd HH:mm:ss",
    timezone: to_string($message.transaction_timezone)
  );
  set_field("transaction_year", new_date.year);
end

All parameters in Graylog’s processing functions are named, please refer to the function index.

**Function Index**

The following list describes the built in functions that ship with Graylog. Additional third party functions are available via other plugins in the marketplace.

**Table 15.1: Built-in Functions**

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>to_bool</td>
<td>Converts the single parameter to a boolean value using its string value.</td>
</tr>
<tr>
<td>to_double</td>
<td>Converts the first parameter to a double floating point value.</td>
</tr>
<tr>
<td>to_long</td>
<td>Converts the first parameter to a long integer value.</td>
</tr>
<tr>
<td>to_string</td>
<td>Converts the first parameter to its string representation.</td>
</tr>
<tr>
<td>to_url</td>
<td>Converts a value to a valid URL using its string representation.</td>
</tr>
<tr>
<td>is_null</td>
<td>Checks whether a value is ‘null’.</td>
</tr>
<tr>
<td>is_not_null</td>
<td>Checks whether a value is not ‘null’.</td>
</tr>
<tr>
<td>abbreviate</td>
<td>Abbreviates a String using ellipses.</td>
</tr>
<tr>
<td>capitalize</td>
<td>Capitalizes a String changing the first letter to title case.</td>
</tr>
<tr>
<td>uncapitalize</td>
<td>Uncapitalizes a String changing the first letter to lower case.</td>
</tr>
<tr>
<td>uppercase</td>
<td>Converts a String to upper case.</td>
</tr>
<tr>
<td>lowercase</td>
<td>Converts a String to lower case.</td>
</tr>
<tr>
<td>swapcase</td>
<td>Swaps the case of a String.</td>
</tr>
<tr>
<td>contains</td>
<td>Checks if a string contains another string.</td>
</tr>
<tr>
<td>substring</td>
<td>Returns a substring of value with the given start and end offsets.</td>
</tr>
<tr>
<td>concat</td>
<td>Concatenates two strings.</td>
</tr>
<tr>
<td>split</td>
<td>Split a string around matches of this pattern (Java syntax).</td>
</tr>
<tr>
<td>regex</td>
<td>Match a regular expression against a string, with matcher groups.</td>
</tr>
<tr>
<td>grok</td>
<td>Applies a Grok pattern to a string.</td>
</tr>
<tr>
<td>key_value</td>
<td>Extracts key/value pairs from a string.</td>
</tr>
<tr>
<td>crc32</td>
<td>Returns the hex encoded CRC32 digest of the given string.</td>
</tr>
<tr>
<td>crc32c</td>
<td>Returns the hex encoded CRC32C (RFC 3720, Section 12.1) digest of the given string.</td>
</tr>
<tr>
<td>md5</td>
<td>Returns the hex encoded MD5 digest of the given string.</td>
</tr>
<tr>
<td>murmur3_32</td>
<td>Returns the hex encoded MurmurHash3 (32-bit) digest of the given string.</td>
</tr>
<tr>
<td>murmur3_128</td>
<td>Returns the hex encoded MurmurHash3 (128-bit) digest of the given string.</td>
</tr>
<tr>
<td>sha1</td>
<td>Returns the hex encoded SHA1 digest of the given string.</td>
</tr>
<tr>
<td>sha256</td>
<td>Returns the hex encoded SHA256 digest of the given string.</td>
</tr>
<tr>
<td>sha512</td>
<td>Returns the hex encoded SHA512 digest of the given string.</td>
</tr>
<tr>
<td>parse_json</td>
<td>Parse a string into a JSON tree.</td>
</tr>
<tr>
<td>select_jsonpath</td>
<td>Selects one or more named JSON Path expressions from a JSON tree.</td>
</tr>
<tr>
<td>to_ip</td>
<td>Converts the given string to an IP object.</td>
</tr>
</tbody>
</table>

Continued on next page
<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>cidr_match</td>
<td>Checks whether the given IP matches a CIDR pattern.</td>
</tr>
<tr>
<td>from_input</td>
<td>Checks whether the current message was received by the given input.</td>
</tr>
<tr>
<td>route_to_stream</td>
<td>Assigns the current message to the specified stream.</td>
</tr>
<tr>
<td>create_message</td>
<td>Currently incomplete Creates a new message which will be evaluated by the entire processing pipeline.</td>
</tr>
<tr>
<td>clone_message</td>
<td>Clones a message.</td>
</tr>
<tr>
<td>drop_message</td>
<td>This currently processed message will be removed from the processing pipeline after the rule finishes.</td>
</tr>
<tr>
<td>has_field</td>
<td>Checks whether the currently processed message contains the named field.</td>
</tr>
<tr>
<td>remove_field</td>
<td>Removes the named field from the currently processed message.</td>
</tr>
<tr>
<td>set_field</td>
<td>Sets the name field to the given value in the currently processed message.</td>
</tr>
<tr>
<td>set_fields</td>
<td>Sets multiple fields to the given values in the currently processed message.</td>
</tr>
<tr>
<td>rename_field</td>
<td>Rename a message field.</td>
</tr>
<tr>
<td>syslog_facility</td>
<td>Converts a syslog facility number to its string representation.</td>
</tr>
<tr>
<td>syslog_level</td>
<td>Converts a syslog level number to its string representation.</td>
</tr>
<tr>
<td>expand_syslog_priority</td>
<td>Converts a syslog priority number to its level and facility.</td>
</tr>
<tr>
<td>expand_syslog_priority_as_string</td>
<td>Converts a syslog priority number to its level and facility string representations.</td>
</tr>
<tr>
<td>now</td>
<td>Returns the current date and time.</td>
</tr>
<tr>
<td>parse_date</td>
<td>Parses a date and time from the given string, according to a strict pattern.</td>
</tr>
<tr>
<td>flex_parse_date</td>
<td>Attempts to parse a date and time using the Natty date parser.</td>
</tr>
<tr>
<td>format_date</td>
<td>Formats a date and time according to a given formatter pattern.</td>
</tr>
<tr>
<td>to_date</td>
<td>Converts a type to a date.</td>
</tr>
<tr>
<td>years</td>
<td>Create a period with a specified number of years.</td>
</tr>
<tr>
<td>months</td>
<td>Create a period with a specified number of months.</td>
</tr>
<tr>
<td>weeks</td>
<td>Create a period with a specified number of weeks.</td>
</tr>
<tr>
<td>days</td>
<td>Create a period with a specified number of days.</td>
</tr>
<tr>
<td>hours</td>
<td>Create a period with a specified number of hours.</td>
</tr>
<tr>
<td>minutes</td>
<td>Create a period with a specified number of minutes.</td>
</tr>
<tr>
<td>seconds</td>
<td>Create a period with a specified number of seconds.</td>
</tr>
<tr>
<td>millis</td>
<td>Create a period with a specified number of millis.</td>
</tr>
<tr>
<td>period</td>
<td>Parses an ISO 8601 period from the specified string.</td>
</tr>
</tbody>
</table>

**to_bool**

**to_bool**(value: any)

Converts the single parameter to a boolean value using its string value.

**to_double**

**to_double**(value: any, [default: double])

Converts the first parameter to a double floating point value.

**to_long**

**to_long**(value: any, [default: long])

Converts the first parameter to a long integer value.
to_string

to_string(value: any, [default: string])
Converts the first parameter to its string representation.

to_url

to_url(url: any, [default: string])
Converts the given url to a valid URL.

is_null

is_null(value: any)
Checks if the given value is null.

is_not_null

is_not_null(value: any)
Checks if the given value is not null.

abbreviate

abbreviate(value: string, width: long)
Abbreviates a String using ellipses, the width defines the maximum length of the resulting string.

capitalize

capitalize(value: string)
Capitalizes a String changing the first letter to title case.

uncapitalize

uncapitalize(value: string)
Uncapitalizes a String changing the first letter to lower case.

uppercase

uppercase(value: string, [locale: string])
Converts a String to upper case. The locale (IETF BCP 47 language tag) defaults to “en”.

lowercase

lowercase(value: string, [locale: string])
Converts a String to lower case. The locale (IETF BCP 47 language tag) defaults to “en”.
swapcase

swapcase(value: string)

Swaps the case of a String changing upper and title case to lower case, and lower case to upper case.

contains

contains(value: string, search: string, [ignore_case: boolean])

Checks if value contains search, optionally ignoring the case of the search pattern.

substring

substring(value: string, start: long, [end: long])

Returns a substring of value starting at the start offset (zero based indices), optionally ending at the end offset. Both offsets can be negative, indicating positions relative to the end of value.

concat

concat(first: string, second: string)

Returns a new string combining the text of first and second.

split

split(pattern: string, value: string, [limit: int])

Split a value around matches of pattern. Use limit to indicate the number of times the pattern should be applied.

Note: Patterns have to be valid Java String literals, please ensure you escape any backslashes in your regular expressions!

regex

regex(pattern: string, value: string, [group_names: array[string]])

Match the regular expression in pattern against value. Returns a match object, with the boolean property matches to indicate whether the regular expression matched and, if requested, the matching groups as groups. The groups can optionally be named using the group_names array. If not named, the groups names are strings starting with "0".

Note: Patterns have to be valid Java String literals, please ensure you escape any backslashes in your regular expressions!

grok

grok(pattern: string, value: string, [only_named_captures: boolean])

Applies the grok pattern grok to value. Returns a match object, containing a Map of field names and values. You can set only_named_captures to true to only return matches using named captures.
Tip: The result of executing the grok function can be passed as argument for set_fields to set the extracted fields into a message.

**key_value**

```javascript
key_value(
  value: string,
  [delimiters: string],
  [kv_delimiters: string],
  [ignore_empty_values: boolean],
  [allow_dup_keys: boolean],
  [handle_dup_keys: string],
  [trim_key_chars: string],
  [trim_value_chars: string]
)
```

Extracts key-value pairs from the given value and returns them as a Map of field names and values. You can optionally specify:

- **delimiters** Characters used to separate pairs. We will use each character in the string, so you do not need to separate them. Default value: `<whitespace>`.
- **kv_delimiters** Characters used to separate keys from values. Again, there is no need to separate each character. Default value: `=`.
- **ignore_empty_values** Ignores keys containing empty values. Default value: `true`.
- **allow_dup_keys** Indicates if duplicated keys are allowed. Default value: `true`.
- **handle_dup_keys** How to handle duplicated keys (if allow_dup_keys is set). It can take the values `take_first`, which will only use the first value for the key; or `take_last`, which will only use the last value for the key. Setting this option to any other value will change the handling to concatenate, which will combine all values given to the key, separating them with the value set in this option. For example, setting handle_dup_keys: `"",` would combine all values given to a key a, separating them with a comma, such as `1,2,foo`. Default value: `take_first`.
- **trim_key_chars** Characters to trim (remove from the beginning and end) from keys. Default value: no trim.
- **trim_value_chars** Characters to trim (remove from the beginning and end) from values. Default value: no trim.

Tip: The result of executing the key_value function can be passed as argument for set_fields to set the extracted fields into a message.

**crc32**

```javascript
crc32(value: string)
```

Creates the hex encoded CRC32 digest of the value.

**crc32c**

```javascript
crc32c(value: string)
```

Creates the hex encoded CRC32C (RFC 3720, Section 12.1) digest of the value.
md5
md5(value: string)
Creates the hex encoded MD5 digest of the value.

murmur3_32
murmur3_32(value: string)
Creates the hex encoded MurmurHash3 (32-bit) digest of the value.

murmur3_128
murmur3_128(value: string)
Creates the hex encoded MurmurHash3 (128-bit) digest of the value.

sha1
sha1(value: string)
Creates the hex encoded SHA1 digest of the value.

sha256
sha256(value: string)
Creates the hex encoded SHA256 digest of the value.

sha512
sha512(value: string)
Creates the hex encoded SHA512 digest of the value.

parse_json
parse_json(value: string)
 Parses the value string as JSON, returning the resulting JSON tree.

select_jsonpath
select_jsonpath(json: JsonNode, paths: Map<string, string>)
Evaluates the given paths against the json tree and returns the map of the resulting values.

to_ip
to_ip(ip: string)
Converts the given ip string to an IpAddress object.
cidr_match

cidr_match(cidr: string, ip: IpAddress)
Checks whether the given ip address object matches the cidr pattern.

from_input

from_input(id: string | name: string)
Checks whether the currently processed message was received on the given input. The input can be looked up by either specifying its name (the comparison ignores the case) or the id.

route_to_stream

route_to_stream(id: string | name: string, [message: Message])
Routes the message to the given stream. The stream can be looked up by either specifying its name or the id.
If message is omitted, this function uses the currently processed message.
This causes the message to be evaluated on the pipelines connected to that stream, unless the stream has already been processed for this message.

create_message

create_message([message: string], [source: string], [timestamp: DateTime])
Creates a new message with from the given parameters. If any of them is omitted, its value is taken from the corresponding fields of the currently processed message. If timestamp is omitted, the timestamp of the created message will be the timestamp at that moment.

clone_message

clone_message([message: Message])
Clones a message. If message is omitted, this function uses the currently processed message.

drop_message

drop_message(message: Message)
The processing pipeline will remove the given message after the rule is finished executing.
If message is omitted, this function uses the currently processed message.
This can be used to implement flexible blacklisting based on various conditions.

has_field

has_field(field: string, [message: Message])
Checks whether the given message contains a field with the name field.
If message is omitted, this function uses the currently processed message.
remove_field

remove_field(field: string, [message: Message])
Removes the given field with the name `field` from the given `message`, unless the field is reserved.
If `message` is omitted, this function uses the currently processed message.

set_field

set_field(field: string, value: any, [prefix: string], [suffix: string], [message: Message])
Sets the given field named `field` to the new `value`. The `field` name must be valid, and specifically cannot include a . character. It is trimmed of leading and trailing whitespace. String values are trimmed of whitespace as well.
The optional `prefix` and `suffix` parameters specify which prefix or suffix should be added to the inserted field name.
If `message` is omitted, this function uses the currently processed message.

set_fields

set_fields(fields: Map<string, any>, [prefix: string], [suffix: string], [message: Message])
Sets all of the given name-value pairs in `field` in the given message. This is a convenience function acting like `set_field`. It can be helpful for using the result of a function like `select_jsonpath` or `regex` in the currently processed message especially when the key names are the result of a regular expression.
The optional `prefix` and `suffix` parameters specify which prefix or suffix should be added to the inserted field names.
If `message` is omitted, this function uses the currently processed message.

rename_field

rename_field(old_field: string, new_field: string, [message: Message])
Modifies the field name `old_field` to `new_field` in the given message, keeping the field value unchanged.

syslog_facility

syslog_facility(value: any)
Converts the syslog facility number in `value` to its string representation.

syslog_level

syslog_level(value: any)
Converts the syslog severity number in `value` to its string representation.
**expand_syslog_priority**

```python
expand_syslog_priority(value: any)
```

Converts the syslog priority number in `value` to its numeric severity and facility values.

**expand_syslog_priority_as_string**

```python
expand_syslog_priority_as_string(value: any)
```

Converts the syslog priority number in `value` to its severity and facility string representations.

**now**

```python
now([timezone: string])
```

Returns the current date and time. Uses the default time zone **UTC**.

**parse_date**

```python
parse_date(value: string, pattern: string, [timezone: string])
```

 Parses the `value` into a date and time object, using the `pattern`. If no timezone is detected in the pattern, the optional timezone parameter is used as the assumed timezone. If omitted the timezone defaults to **UTC**.

The format used for the `pattern` parameter is identical to the pattern of the Joda-Time DateTimeFormat.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
<th>Presentation</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>era</td>
<td>text</td>
<td>AD</td>
</tr>
<tr>
<td>C</td>
<td>century of era (&gt;=0)</td>
<td>number</td>
<td>20</td>
</tr>
<tr>
<td>Y</td>
<td>year of era (&gt;=0)</td>
<td>year</td>
<td>1996</td>
</tr>
<tr>
<td>x</td>
<td>weekyear</td>
<td>year</td>
<td>1996</td>
</tr>
<tr>
<td>w</td>
<td>week of weekyear</td>
<td>number</td>
<td>27</td>
</tr>
<tr>
<td>e</td>
<td>day of week</td>
<td>number</td>
<td>2</td>
</tr>
<tr>
<td>E</td>
<td>day of week</td>
<td>text</td>
<td>Tuesday; Tue</td>
</tr>
<tr>
<td>y</td>
<td>year</td>
<td>year</td>
<td>1996</td>
</tr>
<tr>
<td>D</td>
<td>day of year</td>
<td>number</td>
<td>189</td>
</tr>
<tr>
<td>M</td>
<td>month of year</td>
<td>month</td>
<td>July; Jul; 07</td>
</tr>
<tr>
<td>d</td>
<td>day of month</td>
<td>number</td>
<td>10</td>
</tr>
<tr>
<td>a</td>
<td>halfday of day</td>
<td>text</td>
<td>PM</td>
</tr>
<tr>
<td>K</td>
<td>hour of halfday (0~11)</td>
<td>number</td>
<td>0</td>
</tr>
<tr>
<td>h</td>
<td>clockhour of halfday (1~12)</td>
<td>number</td>
<td>12</td>
</tr>
<tr>
<td>H</td>
<td>hour of day (0~23)</td>
<td>number</td>
<td>0</td>
</tr>
<tr>
<td>k</td>
<td>clockhour of day (1~24)</td>
<td>number</td>
<td>24</td>
</tr>
<tr>
<td>m</td>
<td>minute of hour</td>
<td>number</td>
<td>30</td>
</tr>
<tr>
<td>s</td>
<td>second of minute</td>
<td>number</td>
<td>55</td>
</tr>
<tr>
<td>S</td>
<td>fraction of second</td>
<td>millis</td>
<td>978</td>
</tr>
<tr>
<td>z</td>
<td>time zone</td>
<td>text</td>
<td>Pacific Standard Time; PST</td>
</tr>
<tr>
<td>Z</td>
<td>time zone offset/id</td>
<td>zone</td>
<td>-0800; -08:00; America/Los_Angeles</td>
</tr>
<tr>
<td>’</td>
<td>escape for text</td>
<td>delimiter</td>
<td></td>
</tr>
<tr>
<td>”</td>
<td>single quote</td>
<td>literal</td>
<td>.</td>
</tr>
</tbody>
</table>
**flex_parse_date**

`flex_parse_date(value: string, [default: DateTime], [timezone: string])`

Uses the Natty date parser to parse a date and time `value`. If no timezone is detected in the pattern, the optional timezone parameter is used as the assumed timezone. If omitted the timezone defaults to UTC.

In case the parser fails to detect a valid date and time the default date and time is being returned, otherwise the expression fails to evaluate and will be aborted.

**format_date**

`format_date(value: DateTime, format: string, [timezone: string])`

Returns the given date and time `value` formatted according to the `format` string. If no timezone is given, it defaults to UTC.

**to_date**

`to_date(value: any, [timezone: string])`

Converts `value` to a date. If no timezone is given, it defaults to UTC.

**years**

`years(value: long)`

Create a period with `value` number of years.

**months**

`months(value: long)`

Create a period with `value` number of months.

**weeks**

`weeks(value: long)`

Create a period with `value` number of weeks.

**days**

`days(value: long)`

Create a period with `value` number of days.

**hours**

`hours(value: long)`

Create a period with `value` number of hours.
minutes

minutes(value: long)
Create a period with value number of minutes.

seconds

seconds(value: long)
Create a period with value number of seconds.

millis

millis(value: long)
Create a period with value number of milliseconds.

period

period(value: string)
Parses an ISO 8601 period from value.

Usage

Overview

Once you have understood the concepts explained in Pipelines, Rules, and Stream connections, you are ready to start creating your own processing pipelines. This page gives you the information you need to get started with the user interface.

Configuration

Configure the message processor

Before start using the processing pipelines you need to ensure the message processor is enabled and correctly configured. You can do so by going to the System -> Configurations page, and check the configuration in the Message Processors Configuration section.
In that screen, you need to **enable the Pipeline Processor** message processor, and you must also **set the Pipeline Processor after the Message Filter Chain**, if you want to be able to use pipelines in your streams, and get access to fields coming from extractors.

### Manage rules

You can create, edit, and delete your pipeline rules in the **Manage rules** page, under **System -> Pipelines**.

Clicking on **Create Rule** or **Edit** in one of the rules will open a page where you can write your own rule, alongside with some documentation to make the task a bit more manageable.
Once there are some rules in Graylog, you can create pipelines that use them to modify and enrich your messages.

To manage your pipelines, access Manage pipelines page under System -> Pipelines. From that page you can create new pipelines, edit existing ones, and delete pipelines you don’t need any more.

In order to create or edit pipelines, and as explained in Pipelines, you need to add your rules into a stage with a certain priority. The Web Interface will let you add rules to the default stage (0), and to create new stages with different priorities.
A pipeline can have more than one stage, and when you create or edit a stage you need to select how to proceed to the next stage in the pipeline:

**All rules on this stage match the message**  This option will only consider further stages in the pipeline when all conditions in rules evaluated in this stage are true. This equals to match all in the **Pipelines** section.

**At least one of the rules on this stage matches the message**  Selecting this option will continue to further stages in the pipeline when one or more of the conditions in rules evaluated in this stage are true. This equals to match either in the **Pipelines** section.

### Connect pipelines to streams

You can decide which streams are connected to a pipeline from the pipeline details page. Under **System -> Pipelines**, click on the title of the pipeline you want to connect to a stream, and then click on the **Edit connections** button.

You can assign many pipelines to the same stream, so they are all evaluated for all messages routed into such stream.

Remember from the **Stream connections** documentation, that the **All messages** stream is the one where all messages are routed, and probably a good place to apply pipelines to all your messages.
Simulate your changes

After performing some changes in a processing pipeline, you most likely want to see how they are applied to incoming messages. This is what the pipeline simulator is for.

Click the **Simulate processing** button under System -> Pipelines or in the pipeline details page to access the pipeline simulator.

In order to test the message processing, you need to feed a raw message that will be routed into the stream you want to simulate. The raw message should be using the same format as Graylog will receive, for instance, you can type a GELF message in the **Raw message** field, just as your GELF library would send. Don’t forget to select the right codec for the message you provide.

Once you provide the message and codec to use as example, clicking on **Load message** will start the simulation, and show the results when they are ready.

The simulation provides the following results:

**Changes summary**  Provides a summary of modified fields in the original message, and a list of added and dropped messages.

**Results preview**  Shows all fields in the processed message.

**Simulation trace**  Displays a trace of the processing, indicating which rules were evaluated, which were executed. It also includes a timeline in microseconds to allow you to see which rules and pipelines are taking most of the time during message processing.
Message rewriting with Drools

Note: Since Graylog 2.0 you can use the processing pipelines for more flexible message rewriting.

Graylog can optionally use Drools Expert to evaluate all incoming messages against a user defined rules file. Each message will be evaluated prior to being written to the outputs.

The rule file location is defined in the Graylog configuration file:

```
# Drools Rule File (Use to rewrite incoming log messages)
rules_file = /etc/graylog.d/rules/graylog.drl
```

The rules file is located on the file system with a .drl file extension. The rules file can contain multiple rules, queries and functions, as well as some resource declarations like imports, globals, and attributes that are assigned and used by your rules and queries.

For more information on the DRL rules syntax please read the Drools User Guide.

**Getting Started**

1. Uncomment the `rules_file` line in the Graylog configuration file.
2. Copy the sample rules file to the location specified in your Graylog configuration file.
3. Modify the rules file to parse/rewrite/filter messages as needed.

**Example rules file**

This is an example rules file:

```
rule "Overwrite localhost"
when
  m : Message( source == "localhost" )
then
  m.addField("source", "localhost.example.com");
  log.info("[Overwrite localhost] rule fired: {"", m);
end

rule "Drop UDP and ICMP Traffic from firewall"
when
```
m : Message( getField("full_message") matches "(?i).*(ICMP|UDP) Packet(.|\n|\r)\s+source == "firewall" ) then
    m.setFilterOut(true);
    log.info("[Drop UDP and ICMP Traffic from firewall] rule fired: {}", m);
end

The log object being used to write log messages from within Drools rules is an instance of the SLF4J Logger interface.

Parsing Message and adding fields

In the following script we turn the PID and the source IP into additional fields:

```java
import org.graylog2.plugin.Message
import java.util.regex.Matcher
import java.util.regex.Pattern

// Raw Syslog
// Example:
// Apr 18 15:34:58 server01 smtp-glass[3371]: NEW (1/0) on=1.1.1.1:9100, src=2.2.2.2:38776, ident=, dst=3.3.3.3:25, id=1303151698.3371
rule "SMTP Glass Logging to GELF"
   when
      m : Message( message matches "^smtp-glass.*" )
   then
      Matcher matcher = Pattern.compile("smtp-glass\[(\d+)\].* src (\d+.(\d+.(\d+.(\d+))\s+ident=, dst=3.3.3.3:25, id=1303151698.3371")
       .matcher(m.getMessage());
      if (matcher.find()) {
         m.addField("_pid", Long.valueOf(matcher.group(1)));
         m.addField("_src", matcher.group(2));
      }
end
```

Another example: Adding additional fields and changing the message itself

We send Squid access logs to Graylog using Syslog. The problem is that the host field of the message was set to the IP address of the Squid proxy, which not very useful. This rule overwrites the source and adds other fields:

```java
import java.util.regex.Matcher
import java.util.regex.Pattern
import java.net.InetAddress;

/*
Raw Syslog: squid[2099]: 1339551529.881 55647 1.2.3.4 TCP_MISS/200 22 GET http://www.google.com/
squid\[(d+)], (d+).d+) *(d+.*(d+.d+.d+.d+) *(w+)/(w+) \d+ (w+) \d+)
matched: 13:1339551529.881
matched: 29:55647
matched: 35:1.2.3.4
matched: 47:TCP_MISS/200
matched: 60:22
matched: 64:GET
matched: 68:http://www.google.com/
*/

rule "Squid Logging to GELF"
   when
      m : Message( getField("facility") == "local5" )
```
then
Matcher matcher = Pattern.compile("squid\/[\d+/]: (\d+\.\d+) *(\d+:\d+.\d+) *(\d+) *(\d+\.\d+\.\d+) *(\d+) *(\d+)\/(\d+)\s+(.*)").matcher(m.getMessage());

if (matcher.find()) {
    m.addField("facility", "squid");
    InetAddress addr = InetAddress.getByName(matcher.group(3));
    String host = addr.getHostName();
    m.addField("source", host);
    m.addField("message", matcher.group(6) + " " + matcher.group(7));
    m.addField("_status", matcher.group(4));
    m.addField("_size", matcher.group(5));
}

Blacklisting messages

You can also use Drools rules to blacklist messages.
Note: Since Graylog 2.0 you can use the processing pipelines for blacklisting.

If you have messages coming into Graylog that should be discarded before being written to Elasticsearch or forwarded to another system you can use Drools rules to perform custom filtering.

The rule file location is defined in the Graylog configuration file:

```
# Drools Rule File (Use to rewrite incoming log messages)
rules_file = /etc/graylog.d/rules/graylog.drl
```

The rules file is located on the file system with a .drl file extension. The rules file can contain multiple rules, queries and functions, as well as some resource declarations like imports, globals, and attributes that are assigned and used by your rules and queries.

For more information on the DRL rules syntax please read the Drools User Guide.

**How to**

The general idea is simple: Any Message marked with setFilterOut(true) will be discarded when processed in the Graylog filter chain. You can either write and load your own filter plugin that can execute any Java code to mark messages or just use the Drools rules. The following example shows how to do this.

**Based on regular expressions**

Put this into your rules_file:

```java
import org.graylog2.plugin.Message
import java.util.regex.Matcher
import java.util.regex.Pattern

rule "Blacklist all messages that start with 'firewall'"
when
    m : Message( message matches "^firewall.*" )
then
    System.out.println("DEBUG: Blacklisting message."); // Don't do this in production.
    m.setFilterOut(true);
end
```
This rule will blacklist any message that starts with the string “firewall” (matches "^firewall.*").
Geolocation

Graylog lets you extract and visualize geolocation information from IP addresses in your logs. Here we will explain how to install and configure the geolocation resolution, and how to create a map with the extracted geo-information.

Setup

The Graylog Map Widget is the plugin providing geolocation capabilities to Graylog. The plugin is compatible with Graylog 2.0.0 and higher, and it is installed by default, although some configuration is still required on your side. This section explains how to configure the plugin in detail.

In case you need to reinstall the plugin for some reason, you can find it inside the Graylog tarball in our downloads page. Follow the instructions in Installing and loading plugins to install it.

Configure the database

In first place, you need to download a geolocation database. We currently support MaxMind City databases in the MaxMind DB format, as the GeoIP2 City Database or GeoLite2 City Database that MaxMind provides.

The next step is to store the geolocation database in all servers running Graylog. As an example, if you were using the Graylog OVA, you could save the database in the /var/opt/graylog/data folder, along with other data used by Graylog. Make sure you grant the right permissions so the user running Graylog can read the file.

Then you need to configure Graylog to start using the geolocation database to resolve IPs in your logs. To do that, open Graylog web interface in your favourite browser, and go to System -> Configurations. You can find the geolocation configuration under the Plugins / Geo-Location Processor section, as seen in the screenshot.
In the configuration modal, you need to check the *Enable geolocation processor*, and enter the path to the geolocation database you use. Once you are all set, click on save to store the configuration changes.
Configure the message processor

The last step before being able to resolve locations from IPs in your logs, is to activate the GeoIP Resolver processor. In the same System -> Configurations page, update the configuration in the Message Processors Configuration section.
In that screen, you need to **enable the GeoIP Resolver**, and you must also **set the GeoIP Resolver as the last message processor to run**, if you want to be able to resolve geolocation from fields coming from extractors.
That's it, at this point Graylog will start looking for fields **containing exclusively** an IPv4 or IPv6 address, and extracting their geolocation into a `<field>_geolocation` field.

**Note:** In case you are not sending structured logs to Graylog, you can use extractors to store the IP addresses in your messages into their own fields. Check out the **Extractors** documentation for more information.

**Important:** The GeoIP Resolver processor will **not process** any internal message fields, i.e. any field starting with `gl2_` such as `gl2_remote_ip`.

**Verify the geolocation configuration (Optional)**

To ensure the geolocation resolution is working as expected, you can do the following:

1. Create a TCP Raw/Plaintext input:
2. Send a message only containing an IP to the newly created input. As an example, we will be using the `nc` command:

   `nc -w0 <graylog_host> 5555 <<< '8.8.8.8'

3. Verify that the message contains a `message_geolocation` field:
4. Delete the input if you don’t need it any more

In case the message does not contain a `message_geolocation` field, please check your Graylog server logs, and ensure you followed the steps in the Configure the database section.

**Visualize geolocations in a map**

Graylog can display maps from geolocation stored in any field, as long as the geo-points are using the `latitude,longitude` format.

**Display a map in the search results page**

On any search result page, you can expand the field you want to use to draw a map in the search sidebar, and click on the *World Map* link. That will show a map with all different points stored in that field.
Add map to a dashboard

You can add the map visualization into any dashboards as you do with other widgets. Once you displayed a map in the search result page, click on Add to dashboard, and select the dashboard where you want to add the map.
18.2. Visualize geolocations in a map
FAQs

Will Graylog extract IPs from all fields?

Yes, as long as they contain exclusively an IP address.

What geo-information is extracted from IPs?

Since version 2.2.0, Graylog extracts the IP coordinates, country ISO code, and the city name if available.

Where is the extracted geo-information stored?

Extracted geo-information is stored in new message fields, named as the original field, and appended suffix describing the stored information. That is, if the original field was called `ip_address`, the extracted geo-information will be stored as follows:

- `ip_address_geolocation` will contain the geo-coordinates
- `ip_address_country_code` will contain the country ISO code
- `ip_address_city_name` will contain the city name (if available) or N/A in other case
Which geo-points format does Graylog use to store geolocation information?

Graylog stores the geolocation information in the `latitude,longitude` format.

I have a field in my messages with geolocation information already, can I use it in Graylog?

Yes, as long as it contains geolocation information in the `latitude,longitude` format.

Not all fields containing IP addresses are resolved. Why does this happen?

Most likely it is a misconfiguration issue. Please ensure that the IPs you want to get geolocation information from are in their own fields, and also ensure that the GeoIP Resolver is enabled, and in the right order in the Message Processors Configuration, as explained in Configure the message processor.
Every Graylog node is constantly keeping track about every indexing operation it performs. This is important for making sure that you are not silently losing any messages. The web interface can show you a number of write operations that failed and also a list of failed operations. Like any other information in the web interface this is also available via the REST APIs so you can hook it into your own monitoring systems.

Information about the indexing failure is stored in a capped MongoDB collection that is limited in size. A lot (many tens of thousands) of failure messages should fit in there but it should not be considered a complete collection of all errors ever thrown.

**Common indexer failure reasons**

There are some common failures that can occur under certain circumstances. Those are explained here:

**MapperParsingException**

An error message would look like this:
You tried to write a string into a numeric field of the index. The indexer tried to convert it to a number, but failed because the string did contain characters that could not be converted.

This can be triggered by for example sending GELF messages with different field types or extractors trying to write strings without converting them to numeric values first. The recommended solution is to actively decide on field types. If you sent in a field like http_response_code with a numeric value then you should never change that type in the future.

The same can happen with all other field types like for example booleans.

Note that index cycling is something to keep in mind here. The first type written to a field per index wins. If the Graylog index cycles then the field types are starting from scratch for that index. If the first message written to that index has the http_response_code set as string then it will be a string until the index cycles the next time. Take a look at Index model for more information.
Graylog has a granular permission system which secures the access to its features. Each interaction which can look at data or change configuration in Graylog must be performed as an authenticated user.

Each user can have varying levels of access to Graylog’s features, which can be controlled with assigning roles to users.

The following sections describe the capabilities of users and roles and also how to use LDAP for authentication.

**Users**

It is recommended to create an account for each individual user accessing Graylog.

User accounts have the usual properties such as a login name, email address, full name, password etc. In addition to these fields, you can also configure the session timeout, roles and timezone.
Sessions

Each login for a user creates a session, which is bound to the browser the user is currently using. Whenever the user interacts with Graylog this session is extended.

For security reasons you will want to have Graylog expire sessions after a certain period of inactivity. Once the interval specified by `timeout` expires the user will be logged out of the system. Requests like displaying throughput statistics do not extend the session, which means that if the user keeps Graylog open in a browser tab, but does not interact with it, their session will expire as if the browser was closed.

Logging out explicitly terminates the session.

Timezone

Since Graylog internally processes and stores messages in the UTC timezone, it is important to set the correct timezone for each user.

Even though the system defaults are often enough to display correct times, in case your team is spread across different timezones, each user can be assigned and change their respective timezone setting. You can find the current timezone settings for the various components on the `System / Overview` page of your Graylog web interface.

Initial Roles

Each user needs to be assigned at least one role, which governs the basic set of permissions this user has in Graylog.

Normal users, which do not need to create inputs, outputs or perform administrative tasks like managing access control etc, should be assigned the built in `Reader` role in addition to the custom roles which grant access to streams and dashboards.

Roles

In Graylog, roles are named collections of individual permissions which can be assigned to users. Previous Graylog versions could only assign individual permissions to each user in the system, making updating stream or dashboard permissions for a large group of users difficult to deal with.

Starting with Graylog 1.2 you can create roles which bundle permissions to streams and dashboards. These roles can then be assigned to any number of users and later be updated to include new streams and dashboards.
The two roles Admin and Reader are built in and cannot be changed. The Admin role grants all permissions and should only be assigned to users operating Graylog. The Reader role grants the basic permissions every user needs to be able to use Graylog. The interface will ensure that every user at least has the Reader role in addition to more business specific roles you create.

Roles cannot be deleted as long as users are still assigned to them to prevent accidentally locking users out.

Creating a role

In order to create a new role, choose the green Add new role button on the System / Authentication / Roles page. This will display a dialog allowing you to describe the new role and select the permissions it grants.
After naming the role, select the permissions you want to grant using the buttons to the right of the respective stream or dashboard names. For each stream or dashboard you can select whether to grant edit or read permissions, but note that edit permissions always imply read permissions as well.

In case you have many streams or dashboards you can use the filter to narrow the list down, and use the checkboxes on the left hand side of the table to select multiple items. You can then use the bulk action buttons on the right hand side to toggle the permissions for all of the selected items at once.
Once you are done, be sure to save your changes. The save button is disabled until you select at least one permission.

**Editing a role**

Administrators can edit roles to add or remove access to new streams and dashboards in the system. The two built in Admin and Reader roles cannot be edited or deleted because they are vital for Graylog’s permission system.

Simply choose the Edit button on the System / Authentication / Roles page and change the settings of the role in the following page:
You can safely rename the role as well as updating its description, the existing role assignment for users will be kept.

Deleting a role

Deleting roles checks whether a role still has users assigned to it, to avoid accidentally locking users out. If you want to remove a role, please remove it from all users first.

Permission system

The Graylog permission system is extremely flexible and allows you to create users that are only allowed to perform certain REST calls. The Roles UI allows you to create roles based on stream or dashboard access but does not expose permissions on a REST call level yet. This guide describes how to create those roles using the Graylog REST API.

Imagine we want to create a role that is only allowed to start or stop message processing on graylog-server nodes.

REST call permissions

Almost every REST call in Graylog has to be authenticated or it will return an HTTP 403 (Forbidden). In addition to that, the requesting user also has to have the permissions to execute the REST call. A Graylog admin user can always
execute all calls and roles based on the standard stream or dashboard permissions can execute calls related to those entities.

If you want to create a user that can only execute calls to start or stop message processing you have to find the name of the required permission first.

You can learn about available permissions by querying the /system/permissions endpoint:

```
```

The server responds with a list such as this:

```json
{
   "permissions": {
      "outputs": [ "create", "edit", "terminate", "read" ],
      "users": [ "tokencreate", "rolesedit", "edit", "permissionedit", "list", "tokenlist", "create", "passwordchange", "tokenremove" ],
      "processing": [ "changestate" ],
      ...  
   }
}
```

Starting and stopping message processing corresponds to the changestate permission in the processing category. We combine both pieces to the permission key `processing:changestate`.

### Creating the role

You can create a new role using the REST API like this:

```
curl -v -XPOST -u ADMIN:PASSWORD -H 'Content-Type: application/json' 'http://graylog.example.org:9000/api/roles' -d ... "Change processing state","description": "Permission to start or stop processing on graylog-server nodes"}''
```

Notice the `processing:changestate` permission that we assigned. Every user with this role will be able to start and stop processing on graylog-server nodes. Graylog’s standard reader permissions do not provide any access to data or maintenance functionalities.

This is the POST body in an easier to read formatting:

```
{
   "name": "Change processing state",
   "description": "Permission to start or stop processing on graylog-server nodes",
   "permissions": [
      "processing:changestate"
   ],
   "read_only": false
}
```

### Assigning the role to a user

Create a new user in the Graylog web interface and assign the new role to it:
Every user needs to at least have the standard “Reader” permissions but those do not provide any access to data or maintenance functionalities.

Now request the user information to see what permissions have been assigned:

```bash
```

```
{
  "id" : "563d1024d4c63709999c4ac2",
  "username" : "maintenanceuser",
  "email" : "it-ops@example.org",
  "full_name" : "Rock Solid",
  "permissions" : [
    "indexercluster:read",
    "messagecount:read",
    "journal:read",
    "inputs:read",
    "metrics:read",
    "processing:changestate",
    "savedsearches:edit",
    "fieldnames:read",
    "buffers:read",
    "system:read",
    "users:edit:maintenanceuser",
    "users:passwordchange:maintenanceuser",
    "savedsearches:create",
    "jvmstats:read",
    "throughput:read",
    "savedsearches:read",
    "messages:read"
  ],
  "preferences" : {
    "updateUnfocussed" : false,
    "enableSmartSearch" : true
  }
}
```
Now you can use this user in your maintenance scripts or automated tasks.

## External authentication

**LDAP / Active Directory**

It is possible to use an external LDAP or Active Directory server to perform user authentication in Graylog. Since Graylog 1.2.0, you can also use LDAP groups to perform authorization by mapping them to Graylog roles.

### Configuration

To set up your LDAP or Active Directory server, go to **System / Authentication / LDAP/Active Directory**. Once LDAP is enabled, you need to provide some details about the directory server.
Please test the server connection before continuing to the next steps.

### User mapping

In order to be able to look for users in the LDAP server you configured, Graylog needs to know some more details about it: the base tree to limit user search queries, the pattern used to look for users, and the field containing the full name of the user. You can test the configuration any time by using the login test form that you can find at the bottom of that page.
The login test information will indicate if Graylog was able to load the given user (and perform authentication, if a password was provided), and it will display all LDAP attributes belonging to the user, as you can see in the screenshot.

That’s it for the basic LDAP configuration. Don’t forget to save your settings at this point!

**Group mapping**

You can additionally control the default permissions for users logging in with LDAP or Active Directory by mapping LDAP groups into Graylog roles. That is extremely helpful if you already use LDAP groups to authorize users in your organization, as you can control the default permissions members of LDAP groups will have.

Once you configure group mapping, Graylog will rely on your LDAP groups to assign roles into users. That means that each time an LDAP user logs into Graylog, their roles will be assigned based on the LDAP groups their belong to.

In first place, you need to fill in the details in the **Group Mapping** section under **System / Authentication / LDAP/Active Directory**, by giving the base where to limit group searches, a pattern used to look for groups, and the group name attribute.

Then you need to select which default user role will be assigned to any users authenticated with the LDAP server should have. It is also possible to assign additional roles to any users logging in with LDAP. Please refer to **Roles** for more details about user roles.

**Note:** Graylog only synchronizes with LDAP when users log in. After changing the default and additional roles for LDAP users, you may need to modify existing users manually or delete them in order to force them to log in again.

You can test the group mapping information by using the login test form, as it will display LDAP groups that the test user belongs to. Save the LDAP settings once you are satisfied with the results.
Finally, in order to map LDAP groups into roles, you need to go to **System / Authentication / LDAP/Active Directory -> LDAP group mapping**. This page will load all available LDAP groups using the configuration you previously provided, and will allow you to select a Graylog role which defines the permissions that group will have inside Graylog.

**Note:** Loading LDAP groups may take some time in certain configurations, specially if you have many groups. In those cases, creating a better filter for groups may help with the loading times.

**Note:** Remember that Graylog only synchronizes with LDAP when users log in, so you may need to modify existing users manually after changing the LDAP group mapping.

**Troubleshooting**

LDAP referrals for groups can be a problem during group mapping. Referral issues are most likely to come up with larger AD setups. The Active Directory servers literally refer to other servers in search results, and it is the client’s responsibility to follow all referrals. Support for that is currently not implemented in Graylog.

Referral issues can be detected by warnings in the server logs about group mapping failing, for example:

```
2016-04-11T15:52:06.045Z WARN [LdapConnector] Unable to iterate over user's groups, unable to perform group mapping. Graylog does not support LDAP referrals at the moment. Please see http://docs.graylog.org/en/2.2/pages/users_and_roles/external_auth.html#troubleshooting
```

These issues may be resolved by either managing the groups manually, or configuring the LDAP connection to work against the **global catalog**. The first solution means simply that the LDAP group settings must not be set, and the groups are managed locally. The global catalog solution requires using the 3268/TCP, or 3269/TCP (TLS) port of eligible Active Directory server. The downside is that using the global catalog service consumes slightly more server resources.
Single Sign-On

The SSO Authentication Plugin for Graylog allows to use arbitrary HTTP request headers for authenticating Graylog users.

Once the plugin has been downloaded and installed on all Graylog nodes, it can be configured on the System / Authentication / Single Sign-On (SSO) page.

The HTTP request header containing the Graylog username can be configured in the Username Header field and should contain exactly one HTTP header name. Most HTTP request header based single sign-on solutions are using the Remote-User or X-Forwarded-User HTTP request header.

In order to only allow trusted proxy servers to provide the Graylog username, the Request must come from a trusted proxy checkbox must be checked. The list of trusted proxy servers can be edited on each Graylog node in the configuration file using the trusted_proxies configuration setting.

If user accounts not existing in the Graylog user database should automatically be created on the first login, the Automatically create user checkbox must be checked. The automatically created users can also be customized to retrieve their full name or email address from another HTTP request header, otherwise the defaults are being used.

Authentication providers

Graylog 2.1.0 and later supports pluggable authentication providers. This means, that Graylog cannot only use the builtin authentication mechanisms like its internal user database, LDAP/Active Directory, or access tokens, but can also be extended by plugins to support other authentication mechanisms, for example Single Sign-On or Two Factor Authentication.
Configuration

The order in which the authentication providers will be queried can be configured in the Graylog web interface on the System / Authentication / Configure Provider Order page.

If a user tries to log into Graylog, the authentication providers will be queried in the configured order until a successful authentication attempt has been made (in which case the user will be logged in) or all authentication providers have denied authentication (in which case the user will not be logged in and get an error message).

By clicking on the Update button on the System / Authentication / Configure Provider Order page, the order of authentication providers can be customized.
Update Authentication Provider Configuration

Order
Use drag and drop to change the execution order of the authentication providers.

- Sessions
- API Tokens
- Single Sign-On (SSO)
- LDAP/Active Directory
- Passwords
- Admin user

Status
Change the checkboxes to change the status of an authentication provider.

<table>
<thead>
<tr>
<th>Provider</th>
<th>Enabled</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sessions</td>
<td>✔️</td>
</tr>
<tr>
<td>API Tokens</td>
<td>✔️</td>
</tr>
<tr>
<td>Single Sign-On (SSO)</td>
<td>✔️</td>
</tr>
<tr>
<td>LDAP/Active Directory</td>
<td>✔️</td>
</tr>
<tr>
<td>Passwords</td>
<td>✔️</td>
</tr>
<tr>
<td>Admin user</td>
<td>✔️</td>
</tr>
</tbody>
</table>

Cancel  Save
About Plugins

Graylog offers various extension points to customize and extend its functionality through writing Java code. The first step for writing a plugin is creating a skeleton that is the same for each type of plugin. The next chapter is explaining how to do this and will then go over to chapters explaining plugin types in detail.

Plugin Types

Graylog comes with a stable plugin API for the following plugin types:

- **Inputs**: Accept/write any messages into Graylog
- **Outputs**: Forward ingested messages to other systems as they are processed
- **Services**: Run at startup and able to implement any functionality
- **Alert Conditions**: Decide whether an alert will be triggered depending on a condition
- **Alert Notifications**: Called when a stream alert condition has been triggered
- **Processors**: Transform/drop incoming messages (can create multiple new messages)
- **Filters**: (Deprecated) Transform/drop incoming messages during processing
- **REST API Resources**: An HTTP resource exposed as part of the Graylog REST API
- **Periodical**: Called at periodical intervals during server runtime
- **Decorators**: Used during search time to modify the presentation of messages
- **Authentication Realms**: Allowing to implement different authentication mechanisms (like single sign-on or 2FA)

API concepts

Graylog uses certain patterns in its code bases to make it easier to write extensions. It is important to know about these to be successful in writing custom for it.

You can browse the Graylog Javadoc documentation for details on each class and method mentioned here.
Factory Class

Many newer Graylog extension points split the common aspects of custom code into three different classes:

- instance creation - an, usually inner, interface commonly called Factory
- configuration - the factory returns a ConfigurationRequest instance (or a wrapped instance of it), commonly called Config
- descriptor - the factory returns a display descriptor instance, commonly called Descriptor

Say Graylog exposes an extension point interface called ExtensionPoint, which contains inner interfaces called Factory, Config and Descriptor. An implementation of ExtensionPoint then looks as following:

```java
public AwesomeExtension implements ExtensionPoint {
    public interface Factory extends ExtensionPoint.Factory {
        @Override
        AwesomeExtension create(Decorator decorator);

        @Override
        AwesomeExtension.Config getConfig();

        @Override
        AwesomeExtension.Descriptor getDescriptor();
    }

    public static class Config implements ExtensionPoint.Config {
        @Override
        public ConfigurationRequest getRequestedConfiguration() {
            return new ConfigurationRequest();
        }
    }

    public static class Descriptor extends ExtensionPoint.Descriptor {
        public Descriptor() {
            super("awesome", "http://docs.graylog.org/", "Awesome Extension");
        }
    }
}
```

This pattern is used to prevent instantiation of extensions just to get their descriptor or configuration information, because some extensions might be expensive to set up or require some external service and configuration to work.

The factory itself is built using Guice’s assisted injection for auto-wired factories. This allows plugin authors (and Graylog’s internals as well) to cleanly describe their extension as well as taking advantage of dependency injection.

To register such an extension, Graylog typically offers a convenience method via its Guice modules (GraylogModule or PluginModule). For example alert conditions follow the same pattern and are registered as such:

```java
public class SampleModule extends PluginModule {
    // other methods omitted for clarity
    @Override
    protected void configure() {
        addAlertCondition(SampleAlertCondition.class.getCanonicalName(),
                         SampleAlertCondition.class,
                         SampleAlertCondition.Factory.class);
    }
}
```
Alert Conditions

An alert condition determines whether an alert is triggered. The result of a condition is sent to an alert notification for sending to remote systems.

In Graylog alerting is based on searches and typically includes a list of messages that lead to the alert. However nothing prevents user code to query other systems than Elasticsearch to produce alerts.

Class Overview

The central interface is `org.graylog2.plugin.alarms.AlertCondition` which is also the type that a plugin module must register using `org.graylog2.plugin.PluginModule#addAlertCondition`.

Alert conditions are configurable at runtime and thus need a corresponding `org.graylog2.plugin.configuration.ConfigurationRequest`.

Like many other types they also require a `org.graylog2.plugin.alarms.AlertCondition.Descriptor` for displaying information about the alert condition.

Typically you will not implement `AlertCondition` directly, but instead use `org.graylog2.alerts.AbstractAlertCondition` which handles the configuration persistence for you automatically and implements two helper to provide the result of a condition check.

Example

Please refer to the sample plugin implementation for the full code.

Bindings

Compare with the code in the sample plugin.

```java
public class SampleModule extends PluginModule {

    @Override
    public Set<? extends PluginConfigBean> getConfigBeans() {
        return Collections.emptySet();
    }

    @Override
    protected void configure() {
        addAlertCondition(SampleAlertCondition.class.getCanonicalName(),
                         SampleAlertCondition.class,
                         SampleAlertCondition.Factory.class);
    }
}
```

User Interface

Alert conditions have no special user interface elements.
Alert Notifications

Alert Notifications are responsible for sending information about alerts to external systems, such as sending an email, push notifications, opening tickets, writing to chat systems etc.

They receive the stream they were bound to as well as the result of the configured Alert Conditions.

Note: Alert Notifications were called Alarm Callbacks in previous versions of Graylog.
The old name is still used in the code and REST API endpoints for backwards compatibility, so you will see it when implementing your plugins.

Class Overview

The interface to implement is org.graylog2.plugin.alarms.callbacks.AlarmCallback which is also the type that a plugin module must register using org.graylog2.plugin.PluginModule#addAlarmCallback.

Example Alert Notification

You can find a minimal implementation in the sample plugin.

To create an alert notification plugin implement the AlarmCallback interface interface:

```java
public class SampleAlertNotification implements AlarmCallback
```

Your IDE should offer you to create the methods you need to implement:

```java
public void initialize(Configuration configuration) throws AlarmCallbackConfigurationException

This is called once at the very beginning of the lifecycle of this plugin. It is common practice to store the Configuration as a private member for later access.

public void call(Stream stream, AlertCondition.CheckResult checkResult) throws AlarmCallbackException

This is the actual alert notification being triggered. Implement your login that interacts with a remote system here, for example sending a push notification, posts into a chat system etc.

public ConfigurationRequest getRequestedConfiguration()

Plugins can request configurations. The UI in the Graylog web interface is generated from this information and the filled out configuration values are passed back to the plugin in initialize(Configuration configuration).

The return value must not be null.

This is an example configuration request:

```java
final ConfigurationRequest configurationRequest = new ConfigurationRequest();
configurationRequest.addField(new TextField("service_key", "Service key", "", "JIRA API token. You can find this token in your account settings.
ConfigurationField.Optional.NOT_OPTIONAL)); // required, must be filled out
class SampleAlertNotification implements AlarmCallback
configurationRequest.addField(new BooleanField("use_https", "HTTPs", true,"Use HTTP for API communication?");
```

public String getName()

Return a human readable name of this plugin.
public Map<String, Object> getAttributes()

Return attributes that might be interesting to be shown under the alert notification in the Graylog web interface. It is
common practice to at least return the used configuration here.

public void checkConfiguration() throws ConfigurationException

Throw a ConfigurationException if the user should have entered missing or invalid configuration parameters.

Caution: The alert notification may be created multiple times, so be sure to not perform business logic in the
constructor.
You should however inject custom dependencies, such as a specific client library or other objects in the constructor.

Bindings

Compare with the code in the sample plugin.

```java
public class SampleModule extends PluginModule {
    @Override
    public Set<? extends PluginConfigBean> getConfigBeans() {
        return Collections.emptySet();
    }

    @Override
    protected void configure() {
        addAlarmCallback(SampleAlertNotification.class);
    }
}
```

User Interface

Alert notifications have no custom user interface elements.

Decorators

Decorators can be used to transform a message field at display time. Multiple decorators can be applied at the same
time, but you cannot make any assumptions about their order, as that is user defined. Stacked decorators receive the
value of the previous decorator results.

They are typically used to map between the stored value and a human readable form of that value, for example like the
Syslog severity mapper (compare its code) maps between numeric values and their textual representation.

Other uses include looking up user names based on a user’s ID in a remote database, triggering a whois request on a
domain name etc.

Class Overview

You need to implement the org.graylog2.plugin.decorators.SearchResponseDecorator interface. This class must declare a Factory Class.

Beyond the factory, configuration and descriptor classes, the only thing that a decorator needs to implement is the
apply function:
SearchResponse apply(SearchResponse searchResponse);

The `org.graylog2.rest.resources.search.responses.SearchResponse` class represents the result that is being returned to the web interface (or other callers of the REST API).

You are free to modify any field, create new fields or remove fields. However, the web interface makes certain assumptions regarding fields that start with `gl2_` and requires at least the `timestamp`, `source` and `message` fields to be present.

Thrown exceptions are being logged as errors and lead to returning the original search response, without any modifications.

**Example**

Please refer to the sample plugin implementation for the full code.

**Bindings**

Compare with the code in the sample plugin.

```java
public class SampleModule extends PluginModule {
    
    @Override
    public Set<? extends PluginConfigBean> getConfigBeans() {
        return Collections.emptySet();
    }

    @Override
    protected void configure() {
        installSearchResponseDecorator(searchResponseDecoratorBinder(),
                                        PipelineProcessorMessageDecorator.class,
                                        PipelineProcessorMessageDecorator.Factory.class);
    }
}
```

**User Interface**

Decorators have no custom user interface elements.

**Writing Plugins**

What you need in your development environment before starting is:

- `git`
- `maven`

There are lots of different ways to get those on your local machine, unfortunately we cannot list all of them, so please refer to your operating system-specific documentation.

Graylog uses a couple of conventions and techniques in its code, so be sure to read about the API concepts for an overview.
Sample Plugin

To go along with this documentation, there is a sample plugin on Github. This documentation will link to specific parts for your reference. It is fully functional, even though it does not implement any useful functionality. Its purpose to provide a reference for helping to implement your own plugins.

Creating a plugin skeleton

The easiest way to get started is to use our Graylog meta project, which will create a complete plugin project infrastructure will all required classes, build definitions, and configurations. Using the meta project allows you to have the Graylog server project and your own plugins (or 3rd party plugins) in the same project, which means that you can run and debug everything in your favorite IDE or navigate seamlessly in the code base.

Note: We are working on a replacement tool for the graylog-project meta project, but for the time being it still works.

Maven is a widely used build tool for Java, that comes pre-installed on many operating systems or can be installed using most package managers. Make sure that you have at least version 3 before you go on.

Use it like this:

```bash
$ git clone git@github.com:Graylog2/graylog-project.git
```

This will create a checkout of the meta project in your current work dir. Now change to the graylog-project directory and execute the step which to download the necessary base modules:

```bash
$ scripts/bootstrap
```

Now you can bootstrap the plugin you want to write from here, by doing:

```bash
$ scripts/bootstrap-plugin jira-alarmcallback
```

It will ask you a few questions about the plugin you are planning to build. Let’s say you work for a company called ACMECorp and want to build an alarm callback plugin that creates a JIRA ticket for each alarm that is triggered:

```
groupId: com.acmecorp
classifier: 1.0.0
package: com.acmecorp
pluginClassName: JiraAlarmCallback
```

Note that you do not have to tell the archetype wizard what kind of plugin you want to build because it is creating the generic plugin skeleton for you but nothing that is related to the actual implementation. More on this in the example plugin chapters later.

You now have a new folder called graylog-plugin-jira-alarmcallback that includes a complete plugin skeleton including Maven build files. Every Java IDE out there can now import the project automatically without any required further configuration.

In IntelliJ IDEA for example you can just use the File -> Open dialog to open the graylog-project directory as a fully configured Java project, which should include the Graylog server and your plugin as submodules.

Please pay close attention to the README file of the Graylog meta project and follow any further instructions listed there to set up your IDE properly.

If you want to continue working on the command line, you can do the following to compile the server and your plugin:
$ mvn package

The anatomy of a plugin

Each plugin contains information to describe itself and register the extensions it contains.

Note: A single plugin can contain multiple extensions to Graylog.

For example a hypothetical plugin might contribute an input, an output and alert notifications to communicate with systems. For convenience this would be bundled in a single plugin registering multiple extensions.

Required classes

At the very minimum you need to implement two interfaces:

- org.graylog2.plugin.Plugin - which is the entry to your plugin code
- org.graylog2.plugin.PluginMetaData - which describes your plugin

The bootstrap-plugin script generates these implementations for you, and you simply need to fill out the details.

Graylog uses Java’s ServiceLoader mechanism to find your plugin’s main class, so if you rename your Plugin implementation, you need to also adjust the service file. Please also see Google Guava’s AutoService which Graylog uses in conjunction with the plain ServiceLoader.

In addition to the service, Graylog needs an additional resource file called graylog-plugin.properties in a special location. This file contains information about the plugin, specifically which classloader the plugin needs to be in, so it needs to be read before the plugin is actually loaded. Typically you can simply take the default that has been generated for you.

Registering your extension

So far the plugin itself does not do anything, because it neither implements any of the available extensions, nor could Graylog know which ones are available from your code.

Graylog uses dependency injection to wire up its internal components as well as the plugins. Thus the extensions a plugin provides need to be exposed as a PluginModule which provides you with a lot of helper methods to register the various available extensions to cut down the boiler plate code you have to write.

An empty module is created for you.

Caution: The PluginModule exposes a lot of extension points, but not all of them are considered stable API for external use.
If in doubt, please reach out to us on our community support channels.

Please refer to the available Plugin Types for detailed information what you can implement. The Sample Plugin contains stub implementations for each of the supported extensions.

Web Plugin creation

Sometimes your plugin is not only supposed to work under the hoods inside a Graylog server as an input, output, alarm callback, etc. but you also want to contribute previously nonexisting functionality to Graylog’s web interface. Since
version 2.0 this is now possible. When using the most recent Graylog meta project to bootstrap the plugin skeleton, you are already good to go for this. Otherwise please see our chapter about Creating a plugin skeleton.

The Graylog web interface is written in JavaScript, based on React. It is built using webpack, which is bundling all JavaScript code (and other files you use, like stylesheets, fonts, images, even audio or video files if you need them) into chunks digestible by your browser and npm, which is managing our external (and own) dependencies. During the build process all of this will be bundled and included in the jar file of your plugin.

This might be overwhelming at first if you are not accustomed to JS-development, but fortunately we have set up a lot to make writing plugins easier for you!

If you use our proposed way for Creating a plugin skeleton, and followed the part about the Writing Plugins, you are already good to go for building a plugin with a web part. All you need is a running Graylog server on your machine. Everything else is fetched at build time!

Getting up and running with a web development environment is as easy as this:

```
$ scripts/start-web-dev
[...]
$ open http://localhost:8080
```

This starts the development web server. It even tries to open a browser window going to it (probably working on Mac OS X only).

If your Graylog server is not running on http://localhost:9000/api/, then you need to edit graylog2-server/graylog2-web-interface/config.js (in your graylog-project directory) and adapt the gl2ServerUrl parameter.

### Web Plugin structure

These are the relevant files and directories in your plugin directory for the web part of it:

- **webpack.config.js** This is the configuration file for the webpack module bundler. Most of it is already preconfigured by our PluginWebpackConfig class, so the file is very small. You can override/extend every configuration option by passing a webpack snippet though.
- **build.config.js.sample** In this file you can customize some of the parameters of the build. There is one mandatory parameter named `web_src_path` which defines the absolute or relative location to a checkout of the Graylog source repository.
- **package.json** This is a standard npm JSON file describing the web part of your plugin, especially its dependencies. You can read more about its format.
- **src/web** This is where the actual code for the web part of your plugin goes to. For the start there is a simple `index.jsx` file, which shows you how to register your plugin and the parts it provides with the Graylog web interface. We will get to this in detail later.

### Required conventions for web plugins

#### Plugin Entrypoint

There is a single file which is the entry point of your plugin, which means that the execution of your plugin starts there. By convention this is `src/web/index.jsx`. You can rename/move this file, you just have to adapt your webpack configuration to reflect this change, but it is not recommended.

In any case, this file needs to contain the following code at the very top:
This part is responsible to include and execute the `webpack-entry` file, which is responsible to set up webpack to use the correct URL format when loading assets for this plugin. If you leave this out, erratic behavior will be the result.

**Linking to other pages from your plugin**

If you want to generate links from the web frontend to other pages of your plugin or the main web interface, you need to use the `Routes.pluginRoute()` helper method to generate the URLs properly. See this file for more information.

**Best practices for web plugin development**

**Using ESLint**

ESLint is an awesome tool for linting JavaScript code. It makes sure that any written code is in line with general best practices and the project-specific coding style/guideline. We at Graylog are striving to make the best use of this tools as possible, to help our developers and you to generate top quality code with little bugs. Therefore we highly recommend to enable it for a Graylog plugin you are writing.

**Code Splitting**

Both the web interface and plugins for it depend on a number of libraries like React, RefluxJS and others. To prevent those getting bundled into both the web interface and plugin assets, therefore wasting space or causing problems (especially React does not like to be present more than once), we extract those into a commons chunk which is reused by the web interface and plugins.

This has no consequences for you as a plugin author, because the configuration to make use of this is already generated for you when using the meta project or the maven archetype. But here are some details about it:

Common libraries are built into a separate vendor bundle using an own configuration file named `webpack.vendor.js`. Using the DLLPlugin a manifest is extracted which allow us to reuse the generated bundle. This is then imported in our main web interface webpack configuration file and the corresponding generated webpack config file for plugins.

**Building plugins**

Building the plugin is easy because the meta project has created all necessary files and settings for you. Just run `mvn package` either from the meta project’s directory (to build the server and the plugin) or from the plugin directory (to build the plugin only):

```
$ mvn package
```

This will generate a .jar file in target/ that is the complete plugin file:

```
$ ls target/jira-alarmcallback-1.0.0-SNAPSHOT.jar
target/jira-alarmcallback-1.0.0-SNAPSHOT.jar
```
Installing and loading plugins

The only thing you need to do to run the plugin in Graylog is to copy the .jar file to your plugins folder that is configured in your graylog.conf. The default is just plugins/ relative from your graylog-server directory.

This is a list of default plugin locations for the different installation methods.

<table>
<thead>
<tr>
<th>Installation Method</th>
<th>Directory</th>
</tr>
</thead>
<tbody>
<tr>
<td>Virtual Machine Appliances</td>
<td>/opt/graylog/plugins/</td>
</tr>
<tr>
<td>Operating System Packages</td>
<td>/usr/share/graylog-server/plugin/</td>
</tr>
<tr>
<td>Manual Setup</td>
<td>/&lt;extracted-graylog-tarball-path&gt;/plugin/</td>
</tr>
</tbody>
</table>

Restart graylog-server and the plugin should be available to use from the web interface immediately.
CHAPTER 22

External dashboards

There are other frontends that are connecting to the Graylog REST API and display data or information in a special way.

**CLI stream dashboard**

This official Graylog dashboard which is developed by us is showing live information of a specific stream in your terminal. For example it is the perfect companion during a deployment of your platform: Run it next to the deployment output and show information of a stream that is catching all errors or exceptions on your systems.

The CLI stream dashboard documentation is available on GitHub.
Graylog Marketplace

The Graylog Marketplace is the central directory of add-ons for Graylog. It contains plugins, content packs, GELF libraries and more content built by Graylog developers and community members.

GitHub integration

The Marketplace is deeply integrated with GitHub. You sign-in with your GitHub account if you want to submit content and only have to select an existing repository to list on the Marketplace.

From there on you manage your releases and code changes in GitHub. The Marketplace will automatically update your content.

There is no need to sign-in if you only want to browse or download content.
General best practices

README content

We kindly ask you to provide an as descriptive as possible README file with your submission. This file will be displayed on the Marketplace detail page and should provide the following information:

- What is it.
- Why would you want to use it? (Use cases)
- Do you have to register somewhere to get for example an API token?
- How to install and configure it.
- How to use it in a Graylog context.

Take a look at the Splunk plug-in as an example.

The README supports Markdown for formatting. You cannot submit content that does not contain a README file.

License

You cannot submit content that does not contain a LICENSE or COPYING file. We recommend to consult ChooseALicense.com if you are unsure which license to use.

4 Types of Add-Ons

Plug-Ins: Code that extends Graylog to support a specific use case that it doesn’t support out of the box.

Content Pack: A file that can be uploaded into your Graylog system that sets up streams, inputs, extractors, dashboards, etc. to support a given log source or use case.

GELF Library: A library for a programming language or logging framework that supports sending log messages in GELF format for easy integration and pre-structured messages.

Other Solutions: Any other content or guide that helps you integrate Graylog with an external system or device. For example, how to configure a specific device to support a format Graylog understands out of the box.

Contributing plug-ins

You created a Graylog plugin and want to list it in the Marketplace? This is great. Here are the simple steps to follow:

1. Create a GitHub repository for your plugin
2. Include a README and a LICENSE file in the repository.
3. Push all your code to the repository.
4. Create a GitHub release and give it the name of the plugin version. For example 0.1. The Marketplace will always show and link the latest version. You can upload as many release artifacts as you want here. For example the .jar file together with DEB and RPM files. The Marketplace will link to the detail page of a release for downloads.
5. Submit the repository to the Marketplace
Contributing content packs

Graylog content packs can be shared on the Marketplace by following these steps:

1. Export a Graylog content pack from the Graylog Web Interface and save the generated JSON in a file called `content_pack.json`.
2. Create a GitHub repository for your content pack
3. Include a `README` and a `LICENSE` file in the repository.
4. Include the `content_pack.json` file in the root of your GitHub repository.
5. Submit the repository to the Marketplace

Contributing GELF libraries

A GELF library can be added like this:

1. Create a GitHub repository for your GELF library.
2. Include a `README` and a `LICENSE` file in the repository.
3. Describe where to download and how to use the GELF library in the `README`.

Contributing other content

You want to contribute content that does not really fit into the other categories but describes how to integrate a certain system or make it send messages to Graylog?

This is how you can do it:

1. Create a GitHub repository for your content
2. Include a `README` and a `LICENSE` file in the repository.
3. All content goes into the `README`. 
General

Do I need to buy a license to use Graylog?

We believe software should be open and accessible to all. You should not have to pay to analyze your own data, no matter how much you have.

Graylog is licensed under the GNU General Public License. We do not require license fees for production or non-production use.

How long do you support older versions of the Graylog product?

For our commercial support customers, we support older versions of Graylog up to 12 months after the next major release is available. So if you’re using 1.X, you will continue to receive 1.X support up to a full year after 2.0 has been released.

Architecture

What is MongoDB used for?

Graylog uses MongoDB to store your configuration data, not your log data. Only metadata is stored, such as user information or stream configurations. None of your log messages are ever stored in MongoDB. This is why MongoDB does not have a big system impact, and you won’t have to worry too much about scaling it. With our recommended setup architecture, MongoDB will simply run alongside your graylog-server processes and use almost no resources.

Can you guide me on how to replicate MongoDB for High Availability?

MongoDB actually supplies this information as part of their documentation. Check out:

- About MongoDB Replica Sets.
- How to convert a standalone MongoDB node to a replica set.

After you’ve done this, add all MongoDB nodes into the replica_set configuration in all graylog-server.conf files.
I have datacenters across the world and do not want logs forwarding from everywhere to a central location due to bandwidth, etc. How do I handle this?

You can have multiple graylog-server instances in a federated structure, and forward select messages to a centralized GL server.

**Which load balancers do you recommend we use with Graylog?**

You can use any. We have clients running AWS ELB, HAProxy, F5 BIG-IP, and KEMP.

**Isn’t Java slow? Does it need a lot of memory?**

This is a concern that we hear from time to time. We understand Java has a bad reputation from slow and laggy desktop/GUI applications that eat a lot of memory. However, we are usually able to prove this assumption wrong. Well written Java code for server systems is very efficient and does not need a lot of memory resources.

Give it a try, you might be surprised!

**Does Graylog encrypt log data?**

All log data is stored in Elasticsearch. Elastic recommends you use *dm-crypt* at the file system level.

**Where are the log files Graylog produces?**

You can find the log data for Graylog under the below directory with timestamps and levels and exception messages. This is useful for debugging or when the server won’t start.

/var/log/graylog-server/server.log

If you use the pre-build appliances, take a look into

/var/log/graylog/<servicename>/current

**Installation / Setup**

**Should I download the OVA appliances or the separate packages?**

If you are downloading Graylog for the first time to evaluate it, go for the appliance. It is really easy, and can be quickly setup so you can understand if Graylog is right for you. If you are wanting to use Graylog at some scale in production, and do things like high availability (Mongo replication) we recommend you go for the separate packages.

**How do I find out if a specific log source is supported?**

We support many log sources – and more are coming everyday. For a complete list, check out Graylog Marketplace, the central repository of Graylog extensions. There are 4 types of content on the Marketplace:

- Plug-Ins: Code that extends Graylog to support a specific use case that it doesn’t support out of the box.
- Content Pack: A file that can be uploaded into your Graylog system that sets up streams, inputs, extractors, dashboards, etc. to support a given log source or use case.
• GELF Library: A library for a programming language or logging framework that supports sending log messages in GELF format for easy integration and pre-structured messages.

• Other Solutions: Any other content or guide that helps you integrate Graylog with an external system or device. For example, how to configure a specific device to support a format Graylog understands out of the box.

Can I install the Graylog Server on Windows?

Even though our engineers say it is “technically possible”, don’t do it. The Graylog server is built using Java, so technically it can run anywhere. But we currently have it optimized to run better on other operating systems. If you don’t feel comfortable running your own Linux system, we recommend you use our Linux virtual appliance which will run under VMWare.

Can I run Graylog on Azure?

You can create a Linux VM and use our step-by-step to install your customized Graylog. As a second option you can use this guide to convert our Appliance into some Azure compatible virtual machine.

Functionality

Can Graylog automatically clean old data?

Absolutely we have data retention features.

Does Graylog support LDAP / AD and its groups?

Yup, we’re all over this too with read/write roles and group permissions. To start, see this. If you want to get very granular, you can go through the Graylog REST API.

Do we have a user audit log for compliance?

Graylog Enterprise includes an audit log plugin. You can explore the documentation for more details.

It seems like Graylog has no reporting functionality?

That’s correct. We currently don’t have built-in reporting functionality that sends automated reports. However, you can use our REST API to generate and send you own reports. A cron job and the scripting language of your choice should do the trick.

Can I filter inbound messages before they are processed by the Graylog server?

Yes, check out our page on how to use blacklisting.

Dedicated Partition for the Journal

If you create a dedicated Partition for your Kafka Journal, you need to watch that this is a clean directory. Even lost+found can break it, for your reference.
Raise the Java Heap

If you need to raise the Java Heap of the Graylog Server or Elasticsearch in a System that runs as virtual appliances you can use the advanced settings.

On Systems that are installed with DEB / APT this setting can be made in /etc/default/graylog-server. Systems that are installed with RPM / YUM / DNF the file is found in /etc/sysconfig/graylog-server.

How can I start an input on a port below 1024?

If you try to start an input on one of the privileged ports, it will only work for the “root” user. To be able to use a privileged port, you can use authbind on Debian-based systems, or you redirect the traffic with an iptables rule like this:

```
iptables -t nat -A PREROUTING -p tcp --dport 514 -j REDIRECT --to 1514
iptables -t nat -A PREROUTING -p udp --dport 514 -j REDIRECT --to 1514
```

The input needs to be started on port 1514 in this case and will be made available on port 514 to the outside. The clients can then send data to port 514.

Graylog & Integrations

What is the best way to integrate my applications to Graylog?

We recommend that you use GELF. It’s easy for your application developers and eliminates the need to store the messages locally. Also, GELF can just send what app person wants so you don’t have to build extractors or do any extra processing in Graylog.

I have a log source that creates dynamic syslog messages based on events and subtypes and grok patterns are difficult to use - what is the best way to handle this?

Not a problem! Use our key=value extractor.

I want to archive my log data. Can I write to another database, for example HDFS / Hadoop, from Graylog?

Yes, you can output data from Graylog to a different database. We currently have an HDFS output plug-in in the Marketplace - thank you sivasamyk!

It’s also easy and fun to write your own, which you can then add to Graylog Marketplace for others to use.

I don’t want to use Elasticsearch as my backend storage system – can I use another database, like MySQL, Oracle, etc?

You can, but we don’t suggest you do. You will not be able to use our query functionality or our analytic engine on the dataset outside the system. We only recommend another database if you want it for secondary storage.
How can I create a restricted user to check internal Graylog metrics in my monitoring system?

You can create a restricted user which only has access to the `/system/metrics` resource on the Graylog REST API. This way it will be possible to integrate the internal metrics of Graylog into your monitoring system. Giving the user only restricted access will minimize the impact of these credentials getting compromised.

Send a POST request via the Graylog API Browser or curl to the `/roles` resource of the Graylog REST API:

```json
{
  "name": "Metrics Access",
  "description": "Provides read access to all system metrics",
  "permissions": ["metrics:*"],
  "read_only": false
}
```

The following curl command will create the required role (modify the URL of the Graylog REST API, here `http://127.0.0.1:9000/api/`, and the user credentials, here `admin/admin`, according to your setup):

```
$ curl -u admin:admin -H "Content-Type: application/json" -X POST -d '{"name": "Metrics Access", "description": "Provides read access to all system metrics", "permissions": ["metrics:*"], "read_only": false}' 'http://127.0.0.1:9000/api/roles'
```

Troubleshooting

I’m sending in messages, and I can see they are being accepted by Graylog, but I can’t see them in the search. What is going wrong?

A common reason for this issue is that the timestamp in the message is wrong. First, confirm that the message was received by selecting ‘all messages’ as the time range for your search. Then identify and fix the source that is sending the wrong timestamp.

I have configured an SMTP server or an output with TLS connection and receive handshake errors. What should I do?

Outbound TLS connections have CA (certification authority) certificate verification enabled by default. In case the target server’s certificate is not signed by a CA found from trust store, the connection will fail. A typical symptom for this is the following error message in the server logs:

```
Caused by: javax.mail.MessagingException: Could not convert socket to TLS; nested exception is
javax.net.ssl.SSLHandshakeException: sun.security.validator.ValidatorException: PKIX path building failed: sun.security.provider.certpath.SunCertPathBuilderException: unable to find valid certification path to requested target
```

This should be corrected by either adding the missing CA certificates to the Java default trust store (typically found at `$JAVA_HOME/jre/lib/security/cacerts`), or a custom store that is configured (by using `-Djavax.net.ssl.trustStore`) for the Graylog server process. The same procedure applies for both missing valid CAs and self-signed certificates.

For Debian/Ubuntu-based systems using OpenJDK JRE, CA certificates may be added to the systemwide trust store. After installing the JRE (including `ca-certificates-java`, ergo `ca-certificates` packages), place `name-of-certificate-dot-crt` (in PEM format) into `/usr/local/share/ca-certificates/` and run `/usr/sbin/update-ca-certificates`. The hook script in `/etc/ssl/certs/java/cacerts` should automatically generate `/etc/ssl/certs/java/cacerts`

Fedora/RHEL-based systems may refer to Shared System Certificates in the Fedora Project Wiki.
Suddenly parts of Graylog did not work as expected

If you notice multiple different non working parts in Graylog and found something like `java.lang.OutOfMemoryError: unable to create new native thread` in your Graylog Server logfile, you need to raise the process/thread limit of the graylog user. The limit can be checked with `ulimit -u` and you need to check how you can raise `nproc` in your OS.

I cannot go past page 66 in search results

Elasticsearch limits the number of messages per search result to 10000 by default. Graylog displays 150 messages per page, which means that the last full page with default settings will be page 66.

You can increase the maximum result window by adjusting the parameter `index.max_result_window` as described in the Elasticsearch index modules dynamic settings, but be careful as this requires more memory in your Elasticsearch nodes for deep pagination.

This setting can be dynamically updated in Elasticsearch, so that it does not require a cluster restart to be effective.

My field names contain dots and stream alerts do not match anymore

Due to restrictions in certain Elasticsearch versions, Graylog needs to convert field names that contain . characters with another character, by default the replacement character is `_`.

This replacement is done just prior to writing messages to Elasticsearch, which causes a mismatch between what stream rules and alert conditions see as field names when they are evaluated.

Stream rules, the conditions that determine whether or not a message is routed to a stream, are being run as data is being processed by Graylog. These see the field names as containing the dots.

However, alert conditions, which are also attached to streams, are converted to searches and run in the background. They operate on stored data in Elasticsearch and thus see the replacement character for the dots. Thus alert conditions need to use the `_` instead of . when referring to fields. There is currently no way to maintain backwards compatibility and transparently fixing this issue, so you need to take action.

The best option, apart from not sending fields with dots, is to remember to write alert conditions using the replacement character, and never use . in the field names. In general Graylog will use the version with _ in searches etc.

For example, if an incoming message contains the field `docker.container` stream rules use that name, whereas alert conditions need to use `docker_container`. You will notice that the search results also use the latter name.

Have another troubleshooting question?

See below for some additional support options where you can ask your question.

Support

I think I’ve found a bug, how do I report it?

Think you spotted a bug? Oh no! Please report it in our issue trackers so we can take a look at it. All issue trackers are hosted on GitHub, tightly coupled to our code and milestones. Don’t hesitate to open issues – we’ll just close them if there is nothing to do. Most issues will be in the Graylog server repository, but you should choose others if you have found a bug in one of the plugins.
I'm having issues installing or configuring Graylog, where can I go for support?

Check out the Graylog Community Forums – you can search for your problem which may already have an answer, or post a new question.

Another source is the Graylog channel on Matrix.org or the #graylog IRC chat channel on freenode (both are bridged, so you’ll see messages from either channels). Our developers and a lot of community members hang out here. Just join the channel and add any questions, suggestions or general topics you have.

If you’re looking for professional commercial support from the Graylog team, we do that too. Please get in touch here for more details.
Structured events from anywhere. Compressed and chunked.

The Graylog Extended Log Format (GELF) is a log format that avoids the shortcomings of classic plain syslog:

- Limited to length of 1024 bytes – Not much space for payloads like backtraces
- No data types in structured syslog. You don’t know what is a number and what is a string.
- The RFCs are strict enough but there are so many syslog dialects out there that you cannot possibly parse all of them.
- No compression

Syslog is okay for logging system messages of your machines or network gear. GELF is a great choice for logging from within applications. There are libraries and appenders for many programming languages and logging frameworks so it is easy to implement. You could use GELF to send every exception as a log message to your Graylog cluster. You don’t have to care about timeouts, connection problems or anything that might break your application from within your logging class because GELF can be sent via UDP.

GELF via UDP

Chunking

UDP datagrams are usually limited to a size of 8192 bytes. A lot of compressed information fits in there but you sometimes might just have more information to send. This is why Graylog supports chunked GELF.

You can define chunks of messages by prepending a byte header to a GELF message including a message ID and sequence number to reassemble the message later.

Most GELF libraries support chunking transparently and will detect if a message is too big to be sent in one datagram.

Of course TCP would solve this problem on a transport layer but it brings other problems that are even harder to tackle: You would have to care about slow connections, timeouts and other nasty network problems.

With UDP you may just lose a message while with TCP it could bring your whole application down when not designed with care.

Of course TCP makes sense in some (especially high volume environments) so it is your decision. Many GELF libraries support both TCP and UDP as transport. Some do even support HTTP.

Prepend the following structure to your GELF message to make it chunked:
• **Chunked GELF magic bytes - 2 bytes**: 0x1e 0x0f

• **Message ID - 8 bytes**: Must be the same for every chunk of this message. Identifying the whole message and is used to reassemble the chunks later. Generate from millisecond timestamp + hostname for example.

• **Sequence number - 1 byte**: The sequence number of this chunk. Starting at 0 and always less than the sequence count.

• **Sequence count - 1 byte**: Total number of chunks this message has.

All chunks **MUST** arrive within 5 seconds or the server will discard all already arrived and still arriving chunks. A message **MUST NOT** consist of more than 128 chunks.

**Compression**

When using UDP as transport layer, GELF messages can be sent uncompressed or compressed with either GZIP or ZLIB.

Graylog nodes detect the compression type in the GELF magic byte header automatically.

Decide if you want to trade a bit more CPU load for saving a lot of network bandwidth. GZIP is the protocol default.

**GELF via TCP**

At the current time, GELF TCP only supports uncompressed and non-chunked payloads. Each message needs to be delimited with a null byte (\0) when sent in the same TCP connection.

**Attention:** GELF TCP does not support compression due to the use of the null byte (\0) as frame delimiter.

**GELF Payload Specification**

Version 1.1 (11/2013)

A GELF message is a JSON string with the following fields:

• **version string** *(UTF-8)*
  – GELF spec version – “1.1”; **MUST** be set by client library.

• **host string** *(UTF-8)*
  – the name of the host, source or application that sent this message; **MUST** be set by client library.

• **short_message string** *(UTF-8)*
  – a short descriptive message; **MUST** be set by client library.

• **full_message string** *(UTF-8)*
  – a long message that can i.e. contain a backtrace; optional.

• **timestamp number**
  – Seconds since UNIX epoch with optional decimal places for milliseconds; **SHOULD** be set by client library. Will be set to the current timestamp (now) by the server if absent.

• **level number**
– the level equal to the standard syslog levels; optional, default is 1 (ALERT).

• **facility string** (UTF-8)
  – optional, deprecated. Send as additional field instead.

• **line number**
  – the line in a file that caused the error (decimal); optional, deprecated. Send as additional field instead.

• **file string** (UTF-8)
  – the file (with path if you want) that caused the error (string); optional, deprecated. Send as additional field instead.

• **[additional field] string** (UTF-8) or **number**
  – every field you send and prefix with an underscore (_;) will be treated as an additional field. Allowed characters in field names are any word character (letter, number, underscore), dashes and dots. The verifying regular expression is: `^[\w\.-]*$`. Libraries SHOULD not allow to send id as additional field (_id). Graylog server nodes omit this field automatically.

### Example payload

This is an example GELF message payload. Any graylog-server node accepts and stores this as a message when GZIP/ZLIB compressed or even when sent uncompressed over a plain socket (without newlines):

```json
{
  "version": "1.1",
  "host": "example.org",
  "short_message": "A short message that helps you identify what is going on",
  "full_message": "Backtrace here\n\nmore stuff",
  "timestamp": 1385053862.3072,
  "level": 1,
  "_user_id": 9001,
  "_some_info": "foo",
  "_some_env_var": "bar"
}
```

### Sending GELF messages via UDP using netcat

Sending an example message to a GELF UDP input (running on host graylog.example.com on port 12201):

```
echo -n '{ "version": "1.1", "host": "example.org", "short_message": "A short message", "level": 5, "_some_info": "foo", "_some_env_var": "bar" }'
```

### Sending GELF messages via TCP using netcat

Sending an example message to a GELF TCP input (running on host graylog.example.com on port 12201):

```
echo -n -e '{ "version": "1.1", "host": "example.org", "short_message": "A short message", "level": 5, "_some_info": "foo", "_some_env_var": "bar" }'
```

### Sending GELF messages via TCP using curl

Sending an example message to a GELF HTTP input (running on http://graylog.example.com:12201/gelf):

```bash
```
curl -X POST -H 'Content-Type: application/json' -d '{ "version": "1.1", "host": "example.org", "short_message": "A short message", "_some_info": "foo" }' 'http://graylog.example.com:12201/gelf'
The thinking behind the Graylog architecture and why it matters to you

A short history of Graylog

The Graylog project was started by Lennart Koopmann some time around 2009. Back then the most prominent log management software vendor issued a quote for a one year license of their product that was so expensive that he decided to write a log management system himself. Now you might call this a bit over optimistic (I’ll build this in two weeks, end of quote) but the situation was hopeless: there was basically no other product on the market and especially no open source alternatives.

The log management market today

Things have changed a bit since 2009. Now there are viable open source projects with serious products and a growing list of SaaS offerings for log management.

Architectural considerations

Graylog has been successful in providing log management software because it was built for log management from the beginning. Software that stores and analyzes log data must have a very specific architecture to do it efficiently. It is more than just a database or a full text search engine because it has to deal with both text data and metrics data on a time axis. Searches are always bound to a time frame (relative or absolute) and only going back into the past because future log data has not been written yet. A general purpose database or full text search engine that could also store and index the private messages of your online platform for search will never be able to effectively manage your log data. Adding a specialized frontend on top of it makes it look like it could do the job in a good way but is basically just putting lipstick on the wrong stack.

A log management system has to be constructed of several services that take care of processing, indexing, and data access. The most important reason is that you need to scale parts of it horizontally with your changing use cases and usually the different parts of the system have different hardware requirements. All services must be tightly integrated to allow efficient management and configuration of the system as a whole. A data ingestion or forwarder tool is hard to tedious to manage if the configuration has to be stored on the client machines and is not possible via for example REST APIs controlled by a simple interface. A system administrator needs to be able to log into the web interface of a log management product and select log files of a remote host (that has a forwarder running) for ingestion into the tool.

You also want to be able to see the health and configuration of all forwarders, data processors and indexers in a central place because the whole log management stack can easily involve thousands of machines if you include the log
emitting clients into this calculation. You need to be able to see which clients are forwarding log data and which are not to make sure that you are not missing any important data.

**Graylog is coming the closest to the Splunk architecture:**

- **Graylog was solely built as a log management system from the first line of code.** This makes it very efficient and easy to use.

- The `graylog-server` component sits in the middle and works around shortcomings of Elasticsearch (a full text search engine, not a log management system) for log management. It also builds an abstraction layer on top of it to make data access as easy as possible without having to select indices and write tedious time range selection filters, etc. - Just submit the search query and Graylog will take care of the rest for you.

- All parts of the system are tightly integrated and many parts speak to each other to make your job easier.

- Like WordPress makes MySQL a good solution for blogging, Graylog makes Elasticsearch a good solution for logging. You should never have a system or frontend query Elasticsearch directly for log management so we are putting `graylog-server` in front of it.

**Unlimited data collection**

Volume based license models are making your job unnecessarily hard. Price is a big factor here but it is even worse that volume based license models make you (or your manager makes you) try to save volume. This means that you will
be finding yourself thinking about which data really needs to be ingested. The big problem is that you do not know
what you might need the data for in the moment you are sending (or not sending) it. We have seen operations teams
during a downtime wishing that they had collected the data of a certain log file that was now not searchable. This is
counter-productive and dangerous. You can be limited by disk space or other resources but never by the license
that somebody bought.

It is also a law of the market that you have to build your volume pricing model on the amount of data that is usually
collected today. The amount of generated data has increased dramatically and vendors are nailed to their pricing
model from 2008. This is why you get quotes that fill you with sadness in today’s world.

**Blackboxes**

Closed source systems tend to become black boxes that you cannot extend or adapt to fit the needs of your use case.
This is an important thing to consider especially for log management software. The use cases can range from simple
going centralization to ultra flexible data bus requirements. A closed source system will always make you depending
on the vendor because there is no way to adapt. As your setup reaches a certain point of flexibility you might hit a wall
earlier than expected.

Consider spending a part of the money you would spend for the wrong license model for developing your own plugins
or integrations.

**The future**

Graylog is the only open source log management system that will be able to deliver functionality and scaling in a way
that Splunk does. It will be possible to replace Elasticsearch with something that is really suited for log data analysis
without even changing the public facing APIs.
Graylog 2.2.3

Released: 2017-04-04

https://www.graylog.org/blog/92-announcing-graylog-v2-2-3

Core

- Prevent unwanted session extension. Graylog2/graylog2-server#3583
- Properly escape username/roles in web interface. Graylog2/graylog2-server#3588
- Allow “-” in the path pattern for the index range rebuild endpoint. Graylog2/graylog2-server#3600
- Copy Query Button Fix. Graylog2/graylog2-server#3491
- Fixing slicing of alert notifications in pagination. Graylog2/graylog2-server#3619
- Fix cloning alert conditions with identical id when cloning stream. Graylog2/graylog2-server#3616
- Use UTC for embedded Date objects as well. Graylog2/graylog2-server#3626
- Prevent session extension for polling system messages. Graylog2/graylog2-server#3638
- Support replacing whitespace in nested keys for JSON extractor. Graylog2/graylog2-server#3623
- Prevent NPE due to race between rotation and retention threads. Graylog2/graylog2-server#3640
- Prevent session extension in SystemJobsStore. Graylog2/graylog2-server#3625
- Render loading indicator on no results page. Graylog2/graylog2-server#3667
- Using consistent collection of non displayable fields to filter against. Graylog2/graylog2-server#3668
- Ensure that plugin RPMs will be built for Linux. Graylog2/graylog2-server#3659
- Fix navigation in LDAP components. Graylog2/graylog2-server#3670
- Publish GrokPatternsChangedEvent when using content packs. Graylog2/graylog2-server#3621
- Add support for Cisco and FortiGate syslog messages. Graylog2/graylog2-server#3599
- Effectively change user permissions when listing inputs. Graylog2/graylog2-server#3682
- Restore removal of role permissions upon roles update. Graylog2/graylog2-server#3684
- Comply with grace condition when repeat alert notifications is enabled Graylog2/graylog2-server#3676

Pipeline Processor
• Use uppercase timezone in TimezoneAwareFunction and fix default value.  Graylog2/graylog-plugin-pipeline-processor#169

Graylog 2.2.2

Released: 2017-03-03
https://www.graylog.org/blog/90-announcing-graylog-v2-2-2

Core
• Give an option to repeat alert notifications.  Graylog2/graylog2-server#3511
• Wait until notification types are loaded.  Graylog2/graylog2-server#3534
• Fixing slicing of alert conditions in pagination.  Graylog2/graylog2-server#3528
• Fix command line help of the server jar.  Graylog2/graylog2-server#3527

Graylog 2.2.1

Released: 2017-02-20
https://www.graylog.org/blog/89-announcing-graylog-v2-2-1

Core
• Allow version ‘0’ for structured syslog messages.  Graylog2/graylog2-server#3502
• Ignore Content-Type in HttpTransport.  Graylog2/graylog2-server#3477
• Ensure that index_prefix is lower case.  Graylog2/graylog2-server#3476
• Add missing whitespace in SystemOutputsPage.  Graylog2/graylog2-server#3505

Graylog 2.2.0

Released: 2017-02-14
https://www.graylog.org/blog/88-announcing-graylog-v2-2-0

Core
• Warn about pipeline stream routing  Graylog2/graylog2-server#3472
• Add npm shrinkwrap for 2.2.0  Graylog2/graylog2-server#3468
• Use consistent case in old message count conditions  Graylog2/graylog2-server#3454
• Fix stream filter view.  Graylog2/graylog2-server#3390
• Use the default index set by default in stream create form.  Graylog2/graylog2-server#3397
• Fix broken decorator with duplicate messages.  Graylog2/graylog2-server#3400
• Updating index sets store when stream form is opened.  Graylog2/graylog2-server#3410
• Only show extractor actions on string fields.  Graylog2/graylog2-server#3404
• Use correct format when adding timestamp to search.  Graylog2/graylog2-server#3412
- Properly escape strings sent to /messages/{index}/analyze. Graylog2.graylog2-server#3418
- Retain input and stream IDs in content packs. Graylog2.graylog2-server#3416
- Use “order”: -1 in default index template to allow override. Graylog2.graylog2-server#3422
- Improve base-line memory usage. Graylog2.graylog2-server#3425
- Use condition and notification placeholders. Graylog2.graylog2-server#3432
- Get field stats for indices only, which contain field. Graylog2.graylog2-server#3424
- Redirect to overview after editing index set configuration. Graylog2.graylog2-server#3438
- Send “stream” property when updating a decorator. Graylog2.graylog2-server#3439
- Adding simple cache for index sets in MongoIndexSetRegistry. Graylog2.graylog2-server#3440
- Restrict search in RecentMessageLoader to 1 hour. Graylog2.graylog2-server#3367
- Remove stray whitespace in MongoIndexSet. Graylog2.graylog2-server#3371
- Add more detail to index range system notification. Graylog2.graylog2-server#3372
- Suppress error notification when loading a potentially missing input. Graylog2.graylog2-server#3373
- Ensure resolved at is non-null on resolvedSecondsAgo. Graylog2.graylog2-server#3376
- Do not allow testing a message against the default stream. Graylog2.graylog2-server#3377
- Propagate shard failure in multi-index search to global search failure. Graylog2.graylog2-server#3379
- Add support for arrays to “contains” stream rule. Graylog2.graylog2-server#3380
- Automatically unsubscribe from DecoratorsStore in SearchPage. Graylog2.graylog2-server#3363
- Keep modified search bar params when opening modal. Graylog2.graylog2-server#3384
- Reset keyword content after changing range type. Graylog2.graylog2-server#3386
- Disable the “set as default” action for the default index set. Graylog2.graylog2-server#3385
- Unregistering component callbacks from Stream(Rules)Store. Graylog2.graylog2-server#3378
- Check for stream existence when displaying replay link. Graylog2.graylog2-server#3387
- Reuse Generator instance in DocumentationResource. Graylog2.graylog2-server#3293
- Fix: Refreshing saved searches store after deleting one. Graylog2.graylog2-server#3309
- Escape curly braces in Swagger resource URIs for challenged browsers. Graylog2.graylog2-server#3312
- Unbreak Firefox by requesting JSON when pinging the API. Graylog2.graylog2-server#3314
- Escape search query when using surrounding search. Graylog2.graylog2-server#3314
- Close idle GELF HTTP connections after a timeout. Graylog2.graylog2-server#3315
- Ensure that index ranges are deleted when an index set is removed. Graylog2.graylog2-server#3310
- Ignore reopened indices for count-based retention strategies. Graylog2.graylog2-server#3321
- Ignore deprecated “default” field in IndexSetConfig. Graylog2.graylog2-server#3329
- Use last executed search data on auto-refresh. Graylog2.graylog2-server#3330
- Show stats for each index set on the index sets overview page. Graylog2.graylog2-server#3332
- Not fetching/checking unnecessary streams in AlertScannerThread. Graylog2.graylog2-server#3336
- Add more metrics for extractors. Graylog2.graylog2-server#3332
• Saved search improvements. Graylog2/graylog2-server#3338
• Warn when neither reader nor admin roles are selected for a user. Graylog2/graylog2-server#3337
• Prevent setting default index set readonly. Graylog2/graylog2-server#3339
• Add migration to fix parameter value types for alert conditions. Graylog2/graylog2-server#3340
• Fix unlock dashboard from link. Graylog2/graylog2-server#3344
• Allow re-configuration of shards and replicas in the UI. Graylog2/graylog2-server#3349
• Adapt grace period to latest changes in alerts. Graylog2/graylog2-server#3346
• Go back in history instead to users page when canceling user form. Graylog2/graylog2-server#3350
• Improve wrapping of entity title and description. Graylog2/graylog2-server#3351
• Keep stream filter after editing stream. Graylog2/graylog2-server#3352
• Guard against duplication key error from MongoDB. Graylog2/graylog2-server#3358
• Keep calling convention of SearchPage._refreshData consistent. Graylog2/graylog2-server#3357
• Creating MongoDB indices in services running conditional queries. Graylog2/graylog2-server#3345
• Fix NPE in MongoDbSessionDAO#doDelete(). Graylog2/graylog2-server#3227
• Support syslog messages with ISO-8601 timestamps. Graylog2/graylog2-server#3228
• Use local copies of Droid Sans font in Swagger UI. Graylog2/graylog2-server#3229
• Remove empty row if input description is empty. Graylog2/graylog2-server#3237
• Make “sender” optional in EmailAlarmCallback. Graylog2/graylog2-server#3224
• Fix URL for throbber image. Graylog2/graylog2-server#3242
• Remove special handling in SplitAndCountConverter. Graylog2/graylog2-server#3230
• Add missing AuditBindings to journal commands. Graylog2/graylog2-server#3226
• Don’t check private key/certificate if REST API and web interface on same port. Graylog2/graylog2-server#3231
• Add configuration settings for timeout and concurrency of OptimizeIndexJob. Graylog2/graylog2-server#3225
• Change way of exporting CSV search results. Graylog2/graylog2-server#3238
• Add space in processing limit notification. Graylog2/graylog2-server#3256
• Only recalculate index set ranges in index set maintenance menu. Graylog2/graylog2-server#3252
• Fix alert condition validations. Graylog2/graylog2-server#3257
• Validate alarm callbacks before saving them. Graylog2/graylog2-server#3262
• Only update index ranges for managed indices. Graylog2/graylog2-server#3259
• Resolve alerts from deleted alert conditions. Graylog2/graylog2-server#3265
• Alert UI changes. Graylog2/graylog2-server#3266
• Properly track stream throughput for the default stream. Graylog2/graylog2-server#3278
• Add support for OPTIONS request to HttpTransport. Graylog2/graylog2-server#3234
• Add list of stream IDs to Message#toElasticSearchObject(). Graylog2/graylog2-server#3277
• Fix document counts with an empty index set. Graylog2/graylog2-server#3291
- Handle ElasticsearchException in Counts class. Graylog2/graylog2-server#3288
- Move client-side split/import of Grok pattern files to server. Graylog2/graylog2-server#3284
- Showing index set in stream listing only if user is permitted to. Graylog2/graylog2-server#3300
- Fix reloading after notification changes. Graylog2/graylog2-server#3264
- Add “messages:analyze” permission to reader permission set. Graylog2/graylog2-server#3305
- Disable alarm notification controls if user is not permitted to edit. Graylog2/graylog2-server#3303
- Changing conditional to check for presence of index set definition. Graylog2/graylog2-server#3304
- Allow to get a thread dump as plain text. Graylog2/graylog2-server#3289
- Add missing authentication to ClusterResource and ClusterStatsResource. Graylog2/graylog2-server#3279
- Save anchor and last rotation in TimeBasedRotationStrategy depending on IndexSet. Graylog2/graylog2-server#3306
- Fix loading of plugins in development mode. Graylog2/graylog2-server#3185
- Add contains string tester. Graylog2/graylog2-server#3186
- Index set fixes. Graylog2/graylog2-server#3193
- Add synthetic JavaBean getters to AutoValue classes. Graylog2/graylog2-server#3188
- Improve IndexSetValidator. [Graylog2/graylog2-server#3197](https://github.com/Graylog2/graylog2-server/issues/3197)
- Add custom Jackson (de-) serializer for ZonedDateTime and DateTime. Graylog2/graylog2-server#3198
- Improved alarm callback testing. Graylog2/graylog2-server#3196
- Fix DateTime serialization. Graylog2/graylog2-server#3202
- Fix labels on field graphs. Graylog2/graylog2-server#3204
- Alerts cleanup. Graylog2/graylog2-server#3205
- Index set UI fixes. Graylog2/graylog2-server#3203
- Fix quickvalues and field statistics refresh. Graylog2/graylog2-server#3206
- Allow fetching streams by index set ID. Graylog2/graylog2-server#3207
- UI improvements. Graylog2/graylog2-server#3213
- IndexSet default setting. Graylog2/graylog2-server#3209
- Alerts UI improvements. Graylog2/graylog2-server#3214
- Create DefaultIndexSetConfig when creating the default index set. Graylog2/graylog2-server#3215
- ClusterEventPeriodical must use MongoJackObjectMapper. Graylog2/graylog2-server#3217
- Fix V20161130141500_DefaultStreamRecalcIndexRanges. Graylog2/graylog2-server#3222
- Migration improvements. Graylog2/graylog2-server#3211

**Beats plugin**

- Add support for Metricbeat
- Extract “fields” for every type of beat

**Pipeline processor plugin**

- Allow duplicate stream titles in route_to_stream. Graylog2/graylog-plugin-pipeline-processor#154
Graylog Documentation, Release 2.2.1

- Do not use lambdas with gauge metrics. Graylog2/graylog-plugin-pipeline-processor#152
- Add clone_message() function. Graylog2/graylog-plugin-pipeline-processor#153
- Track total pipeline interpreter executionTime as a single metric. Graylog2/graylog-plugin-pipeline-processor#155

Collector sidecar plugin

- Fix: Reload on Beats configuration change.
- Update Beats to version 5.1.1
- Fix race conditions in start/stop/restart code for the exec runner Graylog2/collector-sidecar#123
- Add debug switch Graylog2/collector-sidecar#124
- Using Modern UI in a standard way (thanks to @nicozanf) Graylog2/collector-sidecar#125
- Extract etag cache into its own service. Graylog2/graylog-plugin-collector#43

Graylog 2.1.3

Released: 2017-01-26
https://www.graylog.org/blog/84-announcing-graylog-2-1-3

Core

- Use “order”: -1 in default index template to allow override. Graylog2/graylog2-server#3426
- Add missing authentication to ClusterResource and ClusterStatsResource. Graylog2/graylog2-server#3427
- Unbreak Firefox by requesting JSON when pinging the API. Graylog2/graylog2-server#3430
- Use custom Grizzly error page to prevent XSS. Graylog2/graylog2-server#3428

Beats plugin

- Add support for Metricbeat. Graylog2/graylog-plugin-beats#19
- Extract “fields” for every type of beat. Graylog2/graylog-plugin-beats#18

Graylog 2.1.2

Released: 2016-11-04
https://www.graylog.org/blog/75-announcing-graylog-v2-1-2

Core

- Improve logging in DecodingProcessor. Graylog2/graylog2-server#3025, Graylog2/graylog2-server#3034
- Support all ZLIB compression levels for GELF messages. Graylog2/graylog2-server#3022, Graylog2/graylog2-server#3036
- Implement “contains” stream rule. Graylog2/graylog2-server#3020, Graylog2/graylog2-server#3037
- Make ValidatorProvider a Singleton. Graylog2/graylog2-server#3019, Graylog2/graylog2-server#3038
- Fix NPE in MongoProbe if MongoDB doesn’t run with MMAPv1. Graylog2/graylog2-server#3018, Graylog2/graylog2-server#3039
• Fix NPE in Indices#numberOfMessages(String). Graylog2/graylog2-server#3016, Graylog2/graylog2-server#3041
• Only create new LdapConnectionConfig if LDAP is enabled. Graylog2/graylog2-server#3017, Graylog2/graylog2-server#3040
• Properly track replace-all flag and pass through to API. Graylog2/graylog2-server#3023, Graylog2/graylog2-server#3043
• Replace Jersey GZipEncoder with Grizzly’s GZipFilter. Graylog2/graylog2-server#3021, Graylog2/graylog2-server#3044
• Prevent n+1 query loading for Stream Rules. Graylog2/graylog2-server#3024, Graylog2/graylog2-server#3035. Thank you @bjoernhaeuser!
• Handle search execution errors. Graylog2/graylog2-server#3027, Graylog2/graylog2-server#3045
• Calculate cardinality on field graphs. Graylog2/graylog2-server#3028, Graylog2/graylog2-server#3046
• Update stats function in field graph description. Graylog2/graylog2-server#3029, Graylog2/graylog2-server#3047
• Use response status 500 if search failed but wasn’t syntax error. Graylog2/graylog2-server#3026, Graylog2/graylog2-server#3042
• Improved search indicators. Graylog2/graylog2-server#3031, Graylog2/graylog2-server#3050
• Fix field analyzers loading when search changes. Graylog2/graylog2-server#3030, Graylog2/graylog2-server#3049
• Close search query autocompletion on enter. Graylog2/graylog2-server#3032, Graylog2/graylog2-server#3051
• Refresh stream search when stream changes. Graylog2/graylog2-server#3033, Graylog2/graylog2-server#3052
• Update Joda-Time and moment-timezone. Graylog2/graylog2-server#3059, Graylog2/graylog2-server#3060
• Search button does not always trigger a new search. Graylog2/graylog2-server#3062, Graylog2/graylog2-server#3063

Beats plugin

• Fix frame decoding in case of lost connection. Graylog2/graylog-plugin-beats#14, Graylog2/graylog-plugin-beats#15, Graylog2/graylog-plugin-beats#17. Thank you @hc4!
• Support messages >1024 bytes in BeatsFrameDecoder. Graylog2/graylog-plugin-beats#10, Graylog2/graylog-plugin-beats#12

Pipeline processor plugin

• Don’t doubly negate the value of the expression. Graylog2/graylog-plugin-pipeline-processor#126, Graylog2/graylog-plugin-pipeline-processor#127

Graylog 2.1.1

Released: 2016-09-14
https://www.graylog.org/blog/69-announcing-graylog-v2-1-1

Core

• Proxied requests query other nodes in parallel. Graylog2/graylog2-server#2764, Graylog2/graylog2-server#2779
• Fix 404s on IE 11 using compatibility view. Graylog2/graylog2-server#2768, Graylog2/graylog2-server#2782
• Modify actions in search page triggering a page reload. Graylog2/graylog2-server#2488, Graylog2/graylog2-server#2798
• Do not display login form while loading. Graylog2/graylog2-server#2770, Graylog2/graylog2-server#2802
• Check in SearchPage if search is in progress, reuse promise then. Graylog2/graylog2-server#2799, Graylog2/graylog2-server#2803
• Use index and message_id as message identifier. Graylog2/graylog2-server#2801, Graylog2/graylog2-server#2804
• Fix: file handle leak in KeyUtil (SSL). Graylog2/graylog2-server#2808. Thank you @gbu-censhare!
• Use current search time configuration for CSV export. Graylog2/graylog2-server#2795, Graylog2/graylog2-server#2809
• Explicitly close okhttp response body, avoiding leak connection warning. Graylog2/graylog2-server#2811. Thank you @chainkite!
• Properly close OkHttp Response objects to avoid resource leaks. Graylog2/graylog2-server#2812
• Remove ldap settings check from authenticators. Graylog2/graylog2-server#2817, Graylog2/graylog2-server#2820

Map plugin

• Ignore internal message fields (starting with “gl2_”). Graylog2/graylog-plugin-map-widget#17

Pipeline processor plugin

• Display boolean values in pipeline simulator. Graylog2/graylog-plugin-pipeline-processor#54, Graylog2/graylog-plugin-pipeline-processor#99
• Use case insensitive lookup for timezone IDs. Graylog2/graylog-plugin-pipeline-processor#100, Graylog2/graylog-plugin-pipeline-processor#102

Graylog 2.1.0

Released: 2016-09-01
https://www.graylog.org/blog/68-announcing-graylog-v-2-1-0-ga

Core

• Refactoring of audit events. Graylog2/graylog2-server#2687
• Add a prop to display/hide the page selector. Graylog2/graylog2-server#2711
• Ensure that rest_transport_uri can override the URI scheme of rest_listen_uri. Graylog2/graylog2-server#2680, Graylog2/graylog2-server#2704
• Handle indexer cluster down in web interface. Graylog2/graylog2-server#2623, Graylog2/graylog2-server#2713
• Prevent NPE and verbose logging if converter returns null. Graylog2/graylog2-server#2717, Graylog2/graylog2-server#2729
• Let widget replay search button open in a new tab or window. Graylog2/graylog2-server#2725, Graylog2/graylog2-server#2726
• Return "id" instead of "_id" for message decorators. Graylog2/graylog2-server#2734, Graylog2/graylog2-server#2735
• Make id field consistent for alarm callback histories. Graylog2/graylog2-server#2737
• Audit event changes. Graylog2/graylog2-server#2718
• Let specific stores reuse promises if request is in progress. Graylog2/graylog2-server#2625, Graylog2/graylog2-server#2712
• Disable editing controls for decorator if user lacks permissions. Graylog2/graylog2-server#2730, Graylog2/graylog2-server#2736
• Styling of decorator list. Graylog2/graylog2-server#2743, Graylog2/graylog2-server#2744
• Do not load plugins for journal commands. Graylog2/graylog2-server#2667
• Use proper other count for pie chart slices. Graylog2/graylog2-server#2639, Graylog2/graylog2-server#2671
• Removing unused prop type in StreamRuleList component. Graylog2/graylog2-server#2673
• Add a generic search form component. Graylog2/graylog2-server#2678
• Decorator improvements. Graylog2/graylog2-server#2519, Graylog2/graylog2-server#2666, Graylog2/graylog2-server#2674
• Only show notification link when there are notifications. Graylog2/graylog2-server#2677
• Enable gzip per default for REST API listener. Graylog2/graylog2-server#2670, Graylog2/graylog2-server#2672
• Improvements in raw message loader. Graylog2/graylog2-server#2684
• Allow users of MessageFields to disable field actions. Graylog2/graylog2-server#2685
• Generating a relative redirect URL for web interface in root resource. Graylog2/graylog2-server#2593, Graylog2/graylog2-server#2675
• Add help text for session’s client address. Graylog2/graylog2-server#2656, Graylog2/graylog2-server#2692
• Fix content pack extractor validation. Graylog2/graylog2-server#2663, Graylog2/graylog2-server#2697
• Reset users’ startpages if referenced stream/dashboard is deleted. Graylog2/graylog2-server#2400, Graylog2/graylog2-server#2695, Graylog2/graylog2-server#2702
• Fix token creation via API browser. Graylog2/graylog2-server#2668, Graylog2/graylog2-server#2698
• Allow surrounding search to be opened in new tab. Graylog2/graylog2-server#2531, Graylog2/graylog2-server#2699
• Reformatting component, adding error handler for fetching dashboard. Graylog2/graylog2-server#2576, Graylog2/graylog2-server#2703
• Add format string message decorator. Graylog2/graylog2-server#2660
• Reloading CurrentUserStore when updated user is the current user. Graylog2/graylog2-server#2705, Graylog2/graylog2-server#2706
• General UI improvements Graylog2/graylog2-server#2700
• Add Syslog severity mapper decorator. Graylog2/graylog2-server#2590
• Improvements in message decorators. Graylog2/graylog2-server#2592, Graylog2/graylog2-server#2591, Graylog2/graylog2-server#2598, Graylog2/graylog2-server#2654
• Revert “Move link to API Browser into System menu”. Graylog2/graylog2-server#2586, Graylog2/graylog2-server#2587
• Print - instead of null when client did not provide user agent header. Graylog2/graylog2-server#2601. Thank you @mikkolehtisalo!
• Change logging in normalizeDn() to debug to avoid noisy warnings. Graylog2/graylog2-server#2599
• Ensure that `{rest,web}_{listen,transport,endpoint}_uri` settings are absolute URIs. Graylog2/graylog2-server#2589, Graylog2/graylog2-server#2596, Graylog2/graylog2-server#2600

• Use HTTP and HTTPS default ports for network settings. Graylog2/graylog2-server#2595, Graylog2/graylog2-server#2605

• Dashboard improvements. Graylog2/graylog2-server#2084, Graylog2/graylog2-server#2281, Graylog2/graylog2-server#2626

• Ensure that `password_secret` is at least 16 characters long. Graylog2/graylog2-server#2619, Graylog2/graylog2-server#2622

• Reduce production .js files sizes by 51%. Graylog2/graylog2-server#2617

• Allow `web_endpoint_uri` to be a relative URI. Graylog2/graylog2-server#2600, Graylog2/graylog2-server#2614

• Use default session attribute for principal. Graylog2/graylog2-server#2620, Graylog2/graylog2-server#2621

• Compile regex pattern for MetricFilter only once. Graylog2/graylog2-server#2637. Thank you again @mikkolehtisalo!

• Fix NPE in Indices#checkForReopened(IndexMetaData). Graylog2/graylog2-server#2628, Graylog2/graylog2-server#2635

• Mark message offset as committed in case of a decoding error. Graylog2/graylog2-server#2627, Graylog2/graylog2-server#2643

• Fix cloning streams and editing legacy stream rules. Graylog2/graylog2-server#2244, Graylog2/graylog2-server#2346, Graylog2/graylog2-server#2646

• Add back storing of index failures in MongoDB. Graylog2/graylog2-server#2633, Graylog2/graylog2-server#2644

• Enable running Graylog REST API on different context path. Graylog2/graylog2-server#2603, Graylog2/graylog2-server#2397, Graylog2/graylog2-server#2634

• Add support for ECDSA private keys to KeyUtil. Graylog2/graylog2-server#2454, Graylog2/graylog2-server#2641

• Check for conflict of `rest_listen_uri` and `web_listen_uri`. Graylog2/graylog2-server#2634, Graylog2/graylog2-server#2652

• Remove uppercase example decorator before 2.1 final. Graylog2/graylog2-server#2588, Graylog2/graylog2-server#2598, Graylog2/graylog-plugin-pipeline-processor#73

• Make sure to include charset in getBytes and other relevant code sections. Graylog2/graylog2-server#2567, Graylog2/graylog2-server#2574

• Landing page greets with 2.0 in 2.1. Graylog2/graylog2-server#2579

• Run WebAppNotFoundResponseFilter later and for GET requests only. Graylog2/graylog2-server#2657, Graylog2/graylog2-server#2664

• Update dependencies. Graylog2/graylog2-server#2543, Graylog2/graylog2-server#2565

• Allowing to run REST API and web interface on same port. Graylog2/graylog2-server#2515

• Changing default to make REST API and web if to listen on same host/port. Graylog2/graylog2-server#2446, Graylog2/graylog2-server#2525

• Change plugin REST resource injection to use Class instances. Graylog2/graylog2-server#2492

• Validate that Elasticsearch home/data paths are readable. Graylog2/graylog2-server#2536, Graylog2/graylog2-server#2538
• Fix Version#fromClasspathProperties() when loading from JAR plugin. Graylog2/graylog2-server#2535
• Decorator UI Fixes. Graylog2/graylog2-server#2539
• Fix timing issue in MessageCountAlertCondition. Graylog2/graylog2-server#1704, Graylog2/graylog2-server#2382, Graylog2/graylog2-server#2546
• For HttpHeadersToken pass actual remote address. Graylog2/graylog2-server#2556
• Do not blindly override permission set of ldap users. Graylog2/graylog2-server#2516, Graylog2/graylog2-server#2529
• Display original date time of index ranges on hover. Graylog2/graylog2-server#2549, Graylog2/graylog2-server#2552
• Make it possible for plugins to request a shared class loader. Graylog2/graylog2-server#2436, Graylog2/graylog2-server#2508
• Fix REST API browser after changes to the PluginRestResource injection. Graylog2/graylog2-server#2550
• Make version comparison more lenient with pre-release versions. Graylog2/graylog2-server#2462, Graylog2/graylog2-server#2548
• Always trim message field values on Message class. Graylog2/graylog2-server#1936, Graylog2/graylog2-server#2510
• Fix search results console warnings. Graylog2/graylog2-server#2527
• Fix bulk import of Grok patterns. Graylog2/graylog2-server#2229, Graylog2/graylog2-server#2561
• Add helper method to add AuditLogAppenders. Graylog2/graylog2-server#2562
• Add explanation about the configuration file format. Graylog2/graylog2-server#2563
• Display session information, fix usability in user list, editing users. Graylog2/graylog2-server#2526, Graylog2/graylog2-server#2528, Graylog2/graylog2-server#2540, Graylog2/graylog2-server#2541
• Fix issues with app prefix. Graylog2/graylog2-server#2564, Graylog2/graylog2-server#2583
• Fix extractor and static fields creation in multi-node setups. Graylog2/graylog2-server#2580, Graylog2/graylog2-server#2584
• Authentication improvements. Graylog2/graylog2-server#2572, Graylog2/graylog2-server#2573
• Move Error Prone into default build profile. Graylog2/graylog2-server#2575
• Journal info command does not work. Graylog2/graylog2-server#2493 and Graylog2/graylog2-server#2495
• Search result highlighting color similar to white. Graylog2/graylog2-server#2480
• Cannot POST on Regex Tester (error 500). Graylog2/graylog2-server#2471 and Graylog2/graylog2-server#2472
• Middle-clicking to open new tab not working for some System menu items. Graylog2/graylog2-server#2468
• Json extractor should check for valid lucene keys. Graylog2/graylog2-server#2434 and Graylog2/graylog2-server#2481
• Elasticsearch Red cluster state triggered by index rotation under some conditions. Graylog2/graylog2-server#2371, Graylog2/graylog2-server#2429 and Graylog2/graylog2-server#2477
• Report syntax error when search query contains unescaped slash. Graylog2/graylog2-server#2372 and Graylog2/graylog2-server#2450
• Allowing path prefixes in web_listen_uri so web interface is accessible via path != “/”. Graylog2/graylog2-server#2271 and Graylog2/graylog2-server#2440
• LDAP group mapping: stringwise comparison fails due to different DN formats. Graylog2/graylog2-server#1790 and Graylog2/graylog2-server#2484
• Json extractor prefix. Graylog2/graylog2-server#1646 and Graylog2/graylog2-server#2481
• LDAP users are shown a change password form. Graylog2/graylog2-server#2124, Graylog2/graylog2-server#2327 and Graylog2/graylog2-server#2485
• Switch message filters from polling to subscribing to change events. Graylog2/graylog2-server#2391 and Graylog2/graylog2-server#2496
• Make auth providers fully pluggable. Graylog2/graylog2-server#2232, Graylog2/graylog2-server#2367 and Graylog2/graylog2-server#2522
• Grok extractor: Allow returning only named captures. Graylog2/graylog2-server#1486 and Graylog2/graylog2-server#2500
• Attempt reading DSA key if RSA failed. Graylog2/graylog2-server#2503. Special thanks to @mikkolehtisalo!
• Fix session validation propagation. Graylog2/graylog2-server#2498
• A wrapper to protect from decompression bombs. Graylog2/graylog2-server#2339. Thank you again, @mikkolehtisalo!
• Make exceptions more useful by providing messages and context. Graylog2/graylog2-server#2478
• Decorate search results. Graylog2/graylog2-server#2408, Graylog2/graylog2-server#2482, Graylog2/graylog2-server#2499, Graylog2/graylog-plugin-pipeline-processor#41, Graylog2/graylog-plugin-pipeline-processor#43 and Graylog2/graylog-plugin-pipeline-processor#52
• Introduce CombinedProvider to sync actions and stores initialization. Graylog2/graylog2-server#2523
• Actually use the bluebird promise in FetchProvider. Graylog2/graylog2-server#2482
• Audit event cleanup. Graylog2/graylog2-server#2746
• Update documentation links. Graylog2/graylog2-server#2759
• Allow child elements in the search form. Graylog2/graylog2-server#2756
• Make key_prefix configuration optional. Graylog2/graylog2-server#2755, Graylog2/graylog2-server#2757
• Invalidating widget result cache cluster wide when a widget changes. Graylog2/graylog2-server#2732, Graylog2/graylog2-server#2745
• Correct documentation links in ‘misc/graylog.conf’. Graylog2/graylog2-server#2747. Thank you @supahgreg!
• Throttle LB status if journal utilization is too high. Graylog2/graylog2-server#1100, Graylog2/graylog2-server#1952 and Graylog2/graylog2-server#2312. Thank you @mikkolehtisalo!
• TLS ciphers for inputs should probably be configurable. Graylog2/graylog2-server#2051.
• SelfSignedCertificate should migrate from sun.security.*. Graylog2/graylog2-server#2132 and Graylog2/graylog2-server#2316. Thank you @mikkolehtisalo!
• Fix formatting metric names including more than one namespace prefix. Graylog2/graylog2-server#2254 and Graylog2/graylog2-server#2425.
• Waiting for index range calculation before switching deflector alias. Graylog2/graylog2-server#2264 and Graylog2/graylog2-server#2278.
• Specify application.context. Graylog2/graylog2-server#2271 and Graylog2/graylog2-server#2440.
• Add handler for / in the Graylog REST API. Graylog2/graylog2-server#2376 and Graylog2/graylog2-server#2377.
• User preferred timezone not saved. Graylog2/graylog2-server#2393 and Graylog2/graylog2-server#2395.
• Unable to delete closed index. Graylog2/graylog2-server#2419 and Graylog2/graylog2-server#2437.
• Absolute search results in widget using wrong time. Graylog2/graylog2-server#2428 and Graylog2/graylog2-server#2452.
• Upgrade to Kafka 0.9.0.1. Graylog2/graylog2-server#1912.
• RestAccessLogFilter to use X-Forwarded-For set by trusted proxies. Graylog2/graylog2-server#1981. Thank you @mikkolehtisalo!
• Upgrade to Drools 6.4.0.Final. Graylog2/graylog2-server#2106.
• Stream Rule Titles. Graylog2/graylog2-server#2244.
• Improve search with no results page. Graylog2/graylog2-server#2253.
• Alert condition titles. Graylog2/graylog2-server#2282.
• Upgrade to Jackson 2.7.4. Graylog2/graylog2-server#2304.
• Support changes for pipeline processor simulator. Graylog2/graylog2-server#2320.
• Add dependency on jna to fix chatty Elasticseach log message. Graylog2/graylog2-server#2342.
• Interfaces and simple implementations of an audit log. Graylog2/graylog2-server#2344.
• Do not init available alarm callback types, fetch them explicitly. Graylog2/graylog2-server#2353.
• Move custom analyzer into index template. Graylog2/graylog2-server#2354.
• Remove automatic private key/certificate generation. Graylog2/graylog2-server#2355.
• Improved feedback. Graylog2/graylog2-server#2357.
• Longer retention interval for journal tests. Graylog2/graylog2-server#2388.
• Remove elasticsearch_discovery_zen_ping_multicast_enabled setting. Graylog2/graylog2-server#2394.
• Fix unrequested refresh of configuration forms/Reset configuration forms on cancel. Graylog2/graylog2-server#2399.
• Web If: Updating a few dependencies which are safe to update. Graylog2/graylog2-server#2407.
• Added Information for journal partitions. Graylog2/graylog2-server#2412.
• Fix memory problems with webpack-dev-server in development mode. Graylog2/graylog2-server#2433.
• Remove _ttl in index mapping. Graylog2/graylog2-server#2435.
• Add raw message loader. Graylog2/graylog2-server#2438.
• Extracting our customized ESLint config into separate module. Graylog2/graylog2-server#2441.
• Remove deprecated MongoDB metrics reporter. Graylog2/graylog2-server#2443.
• Allow access to MongoDatabase in MongoConnection. Graylog2/graylog2-server#2444.
• Add some useful FindBugs plugins. Graylog2/graylog2-server#2447.
• Proxies deflector cycle call to make it available on every node. Graylog2/graylog2-server#2448.

Collector sidecar plugin

• Return updated configuration after changing configuration name
• Prevent crashes when failed to propagate state to the server
• Improve compatibility with old API
• Display collector IP address. Graylog2/graylog-plugin-collector#9
• Ability to clone collector configuration. Graylog2/graylog-plugin-collector#10
• NXLog GELF/TLS input should work without cert files. Graylog2/graylog-plugin-collector#13
• Add `tail_files` option
• Expand verbatim text area if value is present
• Validation improvements
• Add buffer option to NXLog outputs
• Make defaults compatible with Windows hosts
• Add support for Beats. Filebeat, Winlogbeat.
• Beats binaries are bundled with the Collector-Sidecar package
• Improve server side validation. Graylog2/graylog2-server#2247 and Graylog2/graylog-plugin-collector#7.
• Add NXLog GELF TCP and TCP/TLS output
• Add support to clone input, outputs and snippets
• Optionally display collector status information in web interface
• Optionally display log directory listing on status page
• If no node-id is given use the hostname as identification
• Linux distribution is detected and can be used in Snippet template
• Silent install on Windows works now
• Collector log files are now auto-rotated
• Collector processes are supervised and restarted on crashes
• NXLog Inputs and Outputs support free text configuration
• Fix web plugin loading on IE 11

**Pipeline processor plugin**

• Add parse error handler for precompute args failures. Graylog2/graylog-plugin-pipeline-processor#84, Graylog2/graylog-plugin-pipeline-processor#93
• Add support for DateTime comparison. Graylog2/graylog-plugin-pipeline-processor#86, Graylog2/graylog-plugin-pipeline-processor#92
• Make some small UI changes around RuleHelper. Graylog2/graylog-plugin-pipeline-processor#90
• Use shared classloader so other plugins can contribute functions. Graylog2/graylog-plugin-pipeline-processor#81, Graylog2/graylog-plugin-pipeline-processor#94
• UI improvements. Graylog2/graylog2-server#2683, Graylog2/graylog-plugin-pipeline-processor#83
• Unregister PipelineInterpreter from event bus. Graylog2/graylog-plugin-pipeline-processor#79
• Use find in the regex function. Graylog2/graylog-plugin-pipeline-processor#35, Graylog2/graylog-plugin-pipeline-processor#88
• Dynamic function list. Graylog2/graylog-plugin-pipeline-processor#89
• Unresolved functions not properly handled. Graylog2/graylog-plugin-pipeline-processor#24, Graylog2/graylog-plugin-pipeline-processor#25
• Unwrap JsonNode values. Graylog2/graylog-plugin-pipeline-processor#68, Graylog2/graylog-plugin-pipeline-processor#72

• Add optional prefix/suffix to set_fields functions. Graylog2/graylog-plugin-pipeline-processor#74, Graylog2/graylog-plugin-pipeline-processor#75

• Add key-value parsing function. Graylog2/graylog-plugin-pipeline-processor#38, Graylog2/graylog-plugin-pipeline-processor#77

• Allow selection of an input ID for the simulation message. Graylog2/graylog2-server#2610, Graylog2/graylog2-server#2650, Graylog2/graylog-plugin-pipeline-processor#78

• Support “only named captures” for pipeline grok function. Graylog2/graylog-plugin-pipeline-processor#59, Graylog2/graylog-plugin-pipeline-processor#65, Graylog2/graylog2-server#2566, Graylog2/graylog2-server#2577

• Make conversion functions more consistent. Graylog2/graylog2-server#63, Graylog2/graylog2-server#64

• Unescape string literals before using them. Graylog2/graylog-plugin-pipeline-processor#47

• Add rename_field function. Graylog2/graylog-plugin-pipeline-processor#50

• Allow null matcher group values in regex function. Graylog2/graylog-plugin-pipeline-processor#49

• Fix 500 error during simulation. Graylog2/graylog-plugin-pipeline-processor#51

• IpAddressConversion caught wrong exception. Graylog2/graylog-plugin-pipeline-processor#32

• Add syslog-related functions. Graylog2/graylog-plugin-pipeline-processor#19.

• Add concat() function. Graylog2/graylog-plugin-pipeline-processor#20.


• Streams without connections stay visible. Graylog2/graylog2-server#2322.

• Add pipeline simulator. Graylog2/graylog-plugin-pipeline-processor#34, Graylog2/graylog-plugin-pipeline-processor#36 and Graylog2/graylog-plugin-pipeline-processor#42.

• Fix page size in function list. Graylog2/graylog-plugin-pipeline-processor#97

Graylog 2.0.3

Released: 2016-06-20

https://www.graylog.org/blog/58-graylog-v2-0-3-released

Improvements

• Make Message#getStreamIds() more reliable. Graylog2/graylog2-server#2378

• Disabling a configured proxy for requests to localhost/127.0.0.1/::1. Graylog2/graylog2-server#2305

Bug fixes

• Update search query on auto refresh Graylog2/graylog2-server#2385 Graylog2/graylog2-server#2379

• Fix permission checks for non admin users Graylog2/graylog2-server#2366 Graylog2/graylog2-server#2358

• Fix display of total count of indices. Graylog2/graylog2-server#2365 Graylog2/graylog2-server#2359

• Fix base URI for API documentation Graylog2/graylog2-server#2362 Graylog2/graylog2-server#2360

• Fix link to API Browser on Node pages Graylog2/graylog2-server#2361 Graylog2/graylog2-server#2360
• Calculate keyword from and to values on the fly Graylog2/graylog2-server#2335 Graylog2/graylog2-server#2301
• Make MemoryAppender thread-safe Graylog2/graylog2-server#2307 Graylog2/graylog2-server#2302
• Use right metrics to display buffer usage Graylog2/graylog2-server#2300 Graylog2/graylog2-server#2299
• Check if props actually contain configuration fields before copying them Graylog2/graylog2-server#2298 Graylog2/graylog2-server#2297

Graylog 2.0.2

Released: 2016-05-27
https://www.graylog.org/blog/57-graylog-v2-0-2-released

Improvements
• Improved user form. Graylog2/graylog2-server#2261
• Improved logging of plugin list on server startup. Graylog2/graylog2-server#2290
• Forbid empty passwords when using LDAP. Graylog2/graylog2-server#2214 Graylog2/graylog2-server#2283
• Improved metrics page. Graylog2/graylog2-server#2250 Graylog2/graylog2-server#2255
• Improved search histogram resolution auto selection. Graylog2/graylog2-server#2148 Graylog2/graylog2-server#2289
• Improved cluster overview page. Graylog2/graylog2-server#2291

Bug Fixes
• Fixed concurrency issue with Drools. Graylog2/graylog2-server#2119 Graylog2/graylog2-server#2188 Graylog2/graylog2-server#2231
• Fixed problems with Internet Explorer. Graylog2/graylog2-server#2246
• Fixed issues with old dashboards. Graylog2/graylog2-server#2262 Graylog2/graylog2-server#2163
• Fixed changing log levels via REST API. Graylog2/graylog2-server#1904 Graylog2/graylog2-server#2277
• Fixed plugin inter-dependencies by using one class loader for all plugins. Graylog2/graylog2-server#2280

Plugin: Pipeline Processor
• Add syslog related rule functions. Graylog2/graylog-plugin-pipeline-processor#19
• Add concat rule functions. Graylog2/graylog-plugin-pipeline-processor#20
• Fixed problem with IP address function. Graylog2/graylog-plugin-pipeline-processor#28 Graylog2/graylog-plugin-pipeline-processor#32
• Properly unescape strings in raw literals. Graylog2/graylog-plugin-pipeline-processor#30 Graylog2/graylog-plugin-pipeline-processor#31

Graylog 2.0.1

Released: 2016-05-11
https://www.graylog.org/blog/56-graylog-v2-0-1-released

Improvements
• Improved session handling. Graylog2/graylog2-server#2157
• Included UPGRADEING file in the build artifact. Graylog2/graylog2-server#2170
• Added rotation/retention settings back to the config file. Graylog2/graylog2-server#2181
• Improved proxy setup configuration settings. Graylog2/graylog2-server#2156
• Forbid wildcard host in rest_transport_uri. Graylog2/graylog2-server#2205
• Improved robustness for unreachable nodes. Graylog2/graylog2-server#2206
• Use a more lightweight API to get all index names and aliases. Graylog2/graylog2-server#2194
  Graylog2/graylog2-server#2210

Bug Fixes

• Fixed some documentation links.
• Fixed inverted stream rules. Graylog2/graylog2-server#2160 Graylog2/graylog2-server#2172
• Fixed swallowed LDAP authentication exception. Graylog2/graylog2-server#2176 Graylog2/graylog2-server#2178
• Fixed insecure handling of PID files. Thanks @juergenhoetzel! Graylog2/graylog2-server#2174
• Fixed alert conditions that have been created in Graylog 1.x. Graylog2/graylog2-server#2169
  Graylog2/graylog2-server#2182
• Fixed setting of application context. Graylog2/graylog2-server#2191 Graylog2/graylog2-server#2208
• Fixed setting of custom Elasticsearch analyzer. Graylog2/graylog2-server#2209
• Fixed masking of password config values in the web interface. Graylog2/graylog2-server#2198
  Graylog2/graylog2-server#2203
• Fixed URL handling. Graylog2/graylog2-server#2200 Graylog2/graylog2-server#2213

Plugin: Collector

• Rotate nxlog logfiles once a day by default.
• Add GELF TCP output for nxlog.

Graylog 2.0.0

Released: 2016-04-27
https://www.graylog.org/blog/55-announcing-graylog-v2-0-ga

Note: Please make sure to read the Upgrade Guide before upgrading to Graylog 2.0. There are breaking changes!

Feature Highlights

See the release announcement for details on the new features.

• Web interface no longer a separate process
• Support for Elasticsearch 2.x
• Live tail support
• Message Processing Pipeline
• Map Widget Plugin
• Collector Sidecar
• Streams filter UI
• Search for surrounding messages
• Query range limit
• Configurable query time ranges
• Archiving (commercial feature)

Bug Fixes

There have been lots of bug fixes since the 1.3 releases. We only list the ones that we worked on since the 2.0 alpha phase.

• Fixed issues with search page pagination and number of returned results: Graylog2/graylog2-server#1759, Graylog2/graylog2-server#1775, and Graylog2/graylog2-server#1802
• Avoid creating MongoDB collection multiple times: Graylog2/graylog2-server#1747
• Removed number of connected nodes in login page: Graylog2/graylog2-server#1732
• Fix dynamic search result histogram resolution: Graylog2/graylog2-server#1764
• Show overlay in Graylog web interface when Graylog server is not available: Graylog2/graylog2-server#1762
• Fix metric types: Graylog2/graylog2-server#1784
• Only load all metrics on demand: Graylog2/graylog2-server#1782
• Activate search refresh after selecting a refresh interval: Graylog2/graylog2-server#1796
• Fix circular dependencies: Graylog2/graylog2-server#1789
• Only render dependencies when input type is available: Graylog2/graylog2-server#1798
• Document web interface configuration settings in graylog.conf. Graylog2/graylog2-server#1777
• Fix roles link to documentation. Graylog2/graylog2-server#1805
• Fix issue with field graphs. Graylog2/graylog2-server#1811
• Fix search result pagination. Graylog2/graylog2-server#1812
• Fix add to query button on quick values. Graylog2/graylog2-server#1797
• Fix URL to Graylog marketplace on content pack export page. Graylog2/graylog2-server#1817
• Fix elasticsearch node name for the Graylog client node. Graylog2/graylog2-server#1814 and Graylog2/graylog2-server#1820
• Fix widget sorting for dashboards.
• Use _ as default key separator in JSON Extractor. Graylog2/graylog2-server#1841
• Clarify that Graylog Collector needs access to rest_listen_uri. Graylog2/graylog2-server#1847
• Fix potential memory leak in GELF UDP handler. Graylog2/graylog2-server#1857 Graylog2/graylog2-server#1862
• Fix user with correct permissions not allowed to view stream: Graylog2/graylog2-server#1887, Graylog2/graylog2-server#1902
• Make pattern to check Graylog-managed indices stricter: Graylog2/graylog2-server#1882, Graylog2/graylog2-server#1888
• Fix throughput counter: Graylog2/graylog2-server#1876
• Fix replay search link in dashboards: Graylog2/graylog2-server#1835
• Render server unavailable page more reliably: Graylog2/graylog2-server#1867
• Fix build issue with maven. Graylog2/graylog2-server#1907 (Thanks @gitfrederic)
• Fix username in REST API access logs. Graylog2/graylog2-server#1815 Graylog2/graylog2-server#1918 (Thanks @mikkolehtisalo)
• Fix alert annotations in message histogram. Graylog2/graylog2-server#1921
• Fix problem with automatic input form reload. Graylog2/graylog2-server#1870 Graylog2/graylog2-server#1929
• Fix asset caching. Graylog2/graylog2-server#1924 Graylog2/graylog2-server#1930
• Fix issue with cursor jumps in the search bar. Graylog2/graylog2-server#1911
• Fix import of Graylog 1.x extractors. Graylog2/graylog2-server#1831 Graylog2/graylog2-server#1937
• Field charts will now use the stream and time range of the current search. Graylog2/graylog2-server#1785 Graylog2/graylog2-web-interface#1620 Graylog2/graylog2-web-interface#1618 Graylog2/graylog2-web-interface#1485 Graylog2/graylog2-server#1938
• Improve browser validations. Graylog2/graylog2-server#1885
• Fix Internet Explorer support. Graylog2/graylog2-server#1935
• Fix issue where a user was logged out when accessing an unauthorized resource. Graylog2/graylog2-server#1944
• Fix issue with surrounding search. Graylog2/graylog2-server#1946
• Fix problem deleting dashboard widget where the plugin got removed. Graylog2/graylog2-server#1943
• Fix permission issue on user edit page. Graylog2/graylog2-server#1964
• Fix histogram time range selection via mouse. Graylog2/graylog2-server#1895
• Fix problems with duplicate Reflux store instances. Graylog2/graylog2-server#1967
• Create PID file earlier in the startup process. Graylog2/graylog2-server#1969 Graylog2/graylog2-server#1978
• Fix content type detection for static assets. Graylog2/graylog2-server#1982 Graylog2/graylog2-server#1983
• Fix caching of static assets. Graylog2/graylog2-server#1982 Graylog2/graylog2-server#1983
• Show error message on malformed search query. Graylog2/graylog2-server#1896
• Fix parsing of GELF chunks. Graylog2/graylog2-server#1986
• Fix problems editing reader users profile. Graylog2/graylog2-server#1984 Graylog2/graylog2-server#1987
• Fix problem with lost extractors and static fields on input update. Graylog2/graylog2-server#1923
• Improve fetching cluster metrics to avoid multiple HTTP calls. Graylog2/graylog2-server#1974 Graylog2/graylog2-server#1990
• Properly handle empty messages. Graylog2/graylog2-server#1584 Graylog2/graylog2-server#1995
• Add 100-Continue support to HTTP inputs. Graylog2/graylog2-server#1939 Graylog2/graylog2-server#1998
• Fix setting dashboard as start page for reader users. Graylog2/graylog2-server#2005
• Allow dots (".") in LDAP group name mappings. Graylog2/graylog2-server#1458 Graylog2/graylog2-server#2009
• Update user edit form when username changes. Graylog2/graylog2-server#2000
• Fix issue with permissions in user form. Graylog2/graylog2-server#1989
• Update extractor example when message is loaded. Graylog2/graylog2-server#1957 Graylog2/graylog2-server#2013
• Disable log4j2 shutdown hooks to avoid exception on shutdown. Graylog2/graylog2-server#1795 Graylog2/graylog2-server#2015
• Fix styling issue with map widget. Graylog2/graylog2-server#2003
• Fix openstreetmap URL in map widget. Graylog2/graylog2-server#1994
• Fix problem with collector heartbeat validation. Graylog2/graylog2-server#2002 Graylog2/graylog2-web-interface#1726 Graylog2/graylog2-plugin-collector#3
• Remove unused command line parameters. Graylog2/graylog2-server#1977
• Fixed timezone issues for date time processing in JSON parser. Graylog2/graylog2-server#2007
• Fixed JavaScript error with field truncation. Graylog2/graylog2-server#2025
• Fixed redirection if user is not authorized. Graylog2/graylog2-server#1985 Graylog2/graylog2-server#2024
• Made changing the sort order in search result table work again. Graylog2/graylog2-server#2028 Graylog2/graylog2-server#2031
• Performance improvements on “System/Indices” page. Graylog2/graylog2-server#2017
• Fixed content-type settings for static assets. Graylog2/graylog2-server#2052
• Fixed return code for invalid input IDs. Graylog2/graylog2-server#1718 Graylog2/graylog2-server#1767
• Improved field analyzer UI. Graylog2/graylog2-server#2022 Graylog2/graylog2-server#2023
• Fixed login with LDAP user. Graylog2/graylog2-server#2045 Graylog2/graylog2-server#2046 Graylog2/graylog2-server#2069
• Fixed issue with bad message timestamps to avoid data loss. Graylog2/graylog2-server#2064 Graylog2/graylog2-server#2065
• Improved handling of Elasticsearch indices. Graylog2/graylog2-server#2058 Graylog2/graylog2-server#2062
• Extractor form improvements for JSON and Grok extractors. Graylog2/graylog2-server#1883 Graylog2/graylog2-server#2020
• Used search refresh to refresh field statistics. Graylog2/graylog2-server#1961 Graylog2/graylog2-server#2068
• Fixed clicking zoom button in quick values. Graylog2/graylog2-server#2040 Graylog2/graylog2-server#2067
• Web interface styling improvements.
• Replaced . in message field keys with a _ for ES 2.x compatibility. Graylog2/graylog2-server#2078
• Fixed unprocessed journal messages reload in node list. Graylog2/graylog2-server#2083
• Fixed problems with stale sessions on the login page. Graylog2/graylog2-server#2073 Graylog2/graylog2-server#2059 Graylog2/graylog2-server#1891
• Fixed issue with index retention strategies. Graylog2/graylog2-server#2100
• Fixed password change form. Graylog2/graylog2-server#2103 Graylog2/graylog2-server#2105
• Do not show search refresh controls on the sources page. Graylog2/graylog2-server#1821 Graylog2/graylog2-server#2104
• Wait for index being available before calculating index range. Graylog2/graylog2-server#2061
  Graylog2/graylog2-server#2098
• Fixed issue with sorting extractors. Graylog2/graylog2-server#2086 Graylog2/graylog2-server#2088
• Improve DataTable UI component. Graylog2/graylog-plugin-pipeline-processor#11
• Move TCP keepalive setting into AbstractTcpTransport to simplify input development. Graylog2/graylog2-
  server#2112
• Fixed issue with Elasticsearch index template update. Graylog2/graylog2-server#2089 Graylog2/graylog2-
  server#2097
• Ensure that tmpDir is writable when generating self-signed certs in TCP transports. Graylog2/graylog2-
  server#2054 Graylog2/graylog2-server#2096
• Fixed default values for plugin configuration forms. Graylog2/graylog2-server#2108 Graylog2/graylog2-
  server#2114
• Dashboard usability improvements. Graylog2/graylog2-server#2093
• Include default values in pluggable entities forms. Graylog2/graylog2-server#2122
• Ignore empty authentication tokens in LdapUserAuthenticator. Graylog2/graylog2-server#2123
• Add REST API authentication and permissions. Graylog2/graylog-plugin-pipeline-processor#15
• Require authenticated user in REST resources. Graylog2/graylog-plugin-pipeline-processor#14
• Lots of UI improvements in the web interface. Graylog2/graylog2-server#2136
• Fixed link to REST API browser. Graylog2/graylog2-server#2133
• Fixed CSV export skipping first chunk. Graylog2/graylog2-server#2128
• Fixed updating content packs. Graylog2/graylog2-server#2138 Graylog2/graylog2-server#2141
• Added missing 404 page. Graylog2/graylog2-server#2139

Graylog 1.3.4

Released: 2016-03-16

https://www.graylog.org/blog/49-graylog-1-3-4-is-now-available

• Fix security issue which allowed redirecting users to arbitrary sites on login Graylog2/graylog2-web-
  interface#1729
• Fix issue with time-based index rotation strategy Graylog2/graylog2-server#725 Graylog2/graylog2-
  server#1693
• Fix issue with IndexFailureServiceImpl Graylog2/graylog2-server#1747
• Add default Content-Type to GettingStartedResource Graylog2/graylog2-server#1700
• Improve OS platform detection Graylog2/graylog2-server#1737
• Add prefixes GRAYLOG_ (environment variables) and graylog. (system properties) for overriding configu-
  ration settings Graylog2/graylog2-server@48ed88d
• Fix URL to Graylog Marketplace on Extractor/Content Pack pages Graylog2/graylog2-server#1817
• Use monospace font on message values Graylog2/graylog2-web-interface@3cce368
Graylog 1.3.3

Released: 2016-01-14

https://www.graylog.org/graylog-1-3-3-is-now-available/

- Absolute and relative time spans give different results [Graylog2/graylog2-server#1572][Graylog2/graylog2-server#1463][Graylog2/graylog2-server#1672][Graylog2/graylog2-server#1679]
- Search result count widget not caching [Graylog2/graylog2-server#1640][Graylog2/graylog2-server#1681]
- Field Value Condition Alert, does not permit decimal values [Graylog2/graylog2-server#1657]
- Correctly handle null values in nested structures in JsonExtractor [Graylog2/graylog2-server#1676][Graylog2/graylog2-server#1677]
- Add `Content-Type` and `X-Graylog2-No-Session-Extension` to CORS headers [Graylog2/graylog2-server#1682][Graylog2/graylog2-server#1685]
- Discard Message Output [Graylog2/graylog2-server#1688]

Graylog 1.3.2

Released: 2015-12-18

https://www.graylog.org/graylog-1-3-2-is-now-available/

- Deserializing a blacklist filter (`FilterDescription`) leads to `StackOverflowError` [Graylog2/graylog2-server#1641]

Graylog 1.3.1

Released: 2015-12-17

https://www.graylog.org/graylog-1-3-1-is-now-available/

- Add option to AMQP transports to bind the queue to the exchange [Graylog2/graylog2-server#1599][Graylog2/graylog2-server#1633]
- Install a Graylog index template instead of set mappings on index creation [Graylog2/graylog2-server#1624][Graylog2/graylog2-server#1628]

Graylog 1.3.0

Released: 2015-12-09

https://www.graylog.org/graylog-1-3-ga-is-ready/

- Allow index range calculation for a single index. [Graylog2/graylog2-server#1451][Graylog2/graylog2-server#1455]
- Performance improvements for index ranges.
- Make internal server logs accessible via REST API. [Graylog2/graylog2-server#1452]
- Make specific configuration values accessible via REST API. [Graylog2/graylog2-server#1484]
• Added Replace Extractor. Graylog2/graylog2-server#1485
• Added a default set of Grok patterns. Graylog2/graylog2-server#1495
• Log operating system details on server startup. Graylog2/graylog2-server#1244 Graylog2/graylog2-server#1553
• Allow reader users to set a dashboard as start page. Graylog2/graylog2-web-interface#1681
• Auto content pack loader – download and install content packs automatically
• Appliance pre-configured for log ingestion and analysis
• Show a getting started guide on first install. Graylog2/graylog2-web-interface#1662
• Include role permissions in “/roles/{rolename}/members” REST API endpoint. Graylog2/graylog2-server#1549
• Fixed NullPointerException in GELF output. Graylog2/graylog2-server#1538
• Fixed NullPointerException in GELF input handling. Graylog2/graylog2-server#1544
• Use the root user’s timezone for LDAP users by default. Graylog2/graylog2-server#1000 Graylog2/graylog2-server#1554
• Fix display of JSON messages. Graylog2/graylog2-web-interface#1686
• Improve search robustness with missing Elasticsearch indices. Graylog2/graylog2-server#1533 Graylog2/graylog2-server#1547
• Fixed race condition between index creation and index mapping configuration. Graylog2/graylog2-server#1502 Graylog2/graylog2-server#1563
• Fixed concurrency problem in GELF input handling. Graylog2/graylog2-server#1561
• Fixed issue with widget value calculation. Graylog2/graylog2-server#1588
• Do not extend user sessions when updating widgets. Graylog2/graylog2-web-interface#1655
• Fixed compatibility mode for Internet Explorer. Graylog2/graylog2-web-interface#1661 Graylog2/graylog2-web-interface#1668
• Fixed whitespace issue in extractor example. Graylog2/graylog2-web-interface#1650
• Fixed several issues on the indices page. Graylog2/graylog2-web-interface#1691 Graylog2/graylog2-web-interface#1692
• Fixed permission issue for stream alert management. Graylog2/graylog2-web-interface#1659
• Fixed deletion of LDAP group mappings when updating LDAP settings. Graylog2/graylog2-server#1513
• Fixed dangling role references after deleting a role Graylog2/graylog2-server#1608
• Support LDAP Group Mapping for Sun Directory Server (new since beta.2) Graylog2/graylog2-server#1583

Graylog 1.2.2

Released: 2015-10-27

https://www.graylog.org/graylog-1-2-2-is-now-available/

• Fixed a whitespace issue in the extractor UI. Graylog2/graylog2-web-interface#1650
• Fixed the index description on the indices page. Graylog2/graylog2-web-interface#1653
• Fixed a memory leak in the GELF UDP handler code. (Analysis and fix contributed by @lightpriest and @onyx-master on GitHub. Thank you!) Graylog2/graylog2-server#1462, Graylog2/graylog2-server#1488
• Improved the LDAP group handling code to handle more LDAP setups. Graylog2/graylog2-server#1433, Graylog2/graylog2-server#1453, Graylog2/graylog2-server#1491, Graylog2/graylog2-server#1494

• Fixed email alerts for users with multiple email addresses. (LDAP setups) Graylog2/graylog2-server#1439, Graylog2/graylog2-server#1492

• Improve index range handling performance. Graylog2/graylog2-server#1465, Graylog2/graylog2-server#1493

• Fixed JSON extractor with null values. Graylog2/graylog2-server#1475, Graylog2/graylog2-server#1505

• Fixed role assignment when updating user via REST API. Graylog2/graylog2-server#1456, Graylog2/graylog2-server#1507

Graylog 1.2.1

Released: 2015-09-22

https://www.graylog.org/graylog-1-2-1-is-now-available/

• Fixed various issues around importing and applying content packs Graylog2/graylog2-server#1423, Graylog2/graylog2-server#1434, Graylog2/graylog2-web-interface#1605, Graylog2/graylog2-web-interface#1614

• Fixed loading existing alarm callbacks that had been created with Graylog 1.0.x or earlier Graylog2/graylog2-server#1428

• Fixed compatibility problem with Elasticsearch 1.5.x and earlier Graylog2/graylog2-server#1426

• Fixed handling of statistical functions in field graphs Graylog2/graylog2-web-interface#1604

• Use correct title when adding quick values to a dashboard Graylog2/graylog2-web-interface#1603

Graylog 1.2.0

Released: 2015-09-14

https://www.graylog.org/announcing-graylog-1-2-ga-release-includes-30-new-features/

• Make sure existing role assignments survive on LDAP account sync. Graylog2/graylog2-server#1405

• Use memberOf query for ActiveDirectory to speed up LDAP queries. Graylog2/graylog2-server#1407

• Removed disable_index_range_calculation configuration option. Graylog2/graylog2-server#1411

• Avoid potentially long-running Elasticsearch cluster-level operations by only saving an index range if it actually changed. Graylog2/graylog2-server#1412

• Allow editing the roles of LDAP users. Graylog2/graylog2-web-interface#1598

• Improved quick values widget. Graylog2/graylog2-web-interface#1487

Graylog 1.2.0-rc.4

Released: 2015-09-08

https://www.graylog.org/announcing-graylog-1-2-rc-4/
• Deprecated MongoDB storage of internal metrics feature.
• Added customizable LDAP filter for user groups lookup. Graylog2/graylog2-server#951
• Allow usage of count and cardinality statistical functions in dashboard widgets. Graylog2/graylog2-server#1376
• Disabled index range recalculation on every index rotation. Graylog2/graylog2-server#1388
• Added automatic migration of user permissions to admin or reader roles. Graylog2/graylog2-server#1389
• Fixed widget problem with invalid timestamps. Graylog2/graylog2-web-interface#1390
• Added config option to enable TLS certificate validation in REST client. Graylog2/graylog2-server#1393
• Fixed rule matching issue in stream routing engine. Graylog2/graylog2-server#1397
• Changed default titles for stream widgets. Graylog2/graylog2-web-interface#1476
• Changed data filters to be case insensitive. Graylog2/graylog2-web-interface#1585
• Improved padding for stack charts. Graylog2/graylog2-web-interface#1568
• Improved resiliency when Elasticsearch is not available. Graylog2/graylog2-web-interface#1518
• Redirect to user edit form after updating a user. Graylog2/graylog2-web-interface#1588
• Improved dashboard widgets error handling. Graylog2/graylog2-web-interface#1588
• Fixed timing issue in streams UI. Graylog2/graylog2-web-interface#1490
• Improved indices overview page. Graylog2/graylog2-web-interface#1593
• Fixed browser back button behavior. Graylog2/graylog2-web-interface#1594
• Fixed accidental type conversion for number configuration fields in alarmcallback plugins. Graylog2/graylog2-web-interface#1596
• Fixed data type problem for extracted timestamps via grok. Graylog2/graylog2-server#1403

Graylog 1.2.0-rc.2

Released: 2015-08-31

https://www.graylog.org/announcing-graylog-1-2-rc/

• Implement global Elasticsearch timeout and add elasticsearch_request_timeout configuration setting. Graylog2/graylog2-server#1220
• Fixed lots of documentation links. Graylog2/graylog2-server#1238
• Groovy shell server removed. Graylog2/graylog2-server#1266
• Lots of index range calculation fixes. Graylog2/graylog2-server#1274
• New Raw AMQP input. Graylog2/graylog2-server#1280
• New Syslog AMQP input. Graylog2/graylog2-server#1280
• Updated bundled Elasticsearch to 1.7.1.
• The fields in configuration dialogs for inputs and outputs are now ordered. Graylog2/graylog2-server#1282
• Allow server startup without working Elasticsearch cluster. Graylog2/graylog2-server#1136, Graylog2/graylog2-server#1289
• Added OR operator to stream matching. Graylog2/graylog2-server#1292, Graylog2/graylog2-web#1552

27.22. Graylog 1.2.0-rc.2 337
• New stream router engine with better stream matching performance.  Graylog2/graylog2-server#1305, Graylog2/graylog2-server#1309

• Grok pattern import/export support for content packs.  Graylog2/graylog2-server#1300, Graylog2/graylog2-web#1527

• Added MessageListCodec interface for codec implementations that can decode multiple messages from one raw message.  Graylog2/graylog2-server#1307

• Added keepalive configuration option for all TCP transports.  Graylog2/graylog2-server#1287, Graylog2/graylog2-server#1318

• Support for roles and LDAP groups.  Graylog2/graylog2-server#1321, Graylog2/graylog2-server#951

• Added timezone configuration option to date converter.  Graylog2/graylog2-server#1320, Graylog2/graylog2-server#1324

• Added alarmcallback history feature.  Graylog2/graylog2-server#1313, Graylog2/graylog2-web#1537

• Added more configuration options to GELF output.  (TCP settings, TLS support)  Graylog2/graylog2-server#1337, Graylog2/graylog2-server#979

• Store timestamp and some other internal fields in Elasticsearch as doc values.  Removed “elasticsearch_store_timestamps_as_doc_values” option from configuration file.  Graylog2/graylog2-server#1335, Graylog2/graylog2-server#1342

• Added TLS support for GELF HTTP input.  Graylog2/graylog2-server#1348

• Added JSON extractor.  Graylog2/graylog2-server#632, Graylog2/graylog2-server#1355, Graylog2/graylog2-web#1555

• Added support for TLS client certificate authentication to all TCP based inputs.  Graylog2/graylog2-server#1357, Graylog2/graylog2-server#1363

• Added stacked chart widget.  Graylog2/graylog2-server#1284, Graylog2/graylog2-web#1513

• Added cardinality option to field histograms.  Graylog2/graylog2-web#1529, Graylog2/graylog2-server#1303

• Lots of dashboard improvements.  Graylog2/graylog2-web#1550

• Replaced Gulp with Webpack.  Graylog2/graylog2-web#1548

• Updated to Play 2.3.10.

Graylog 1.1.6

Released: 2015-08-06

https://www.graylog.org/graylog-1-1-6-released/

• Fix edge case in SyslogOctetCountFrameDecoder which caused the Syslog TCP input to reset connections (Graylog2/graylog2-server#1105, Graylog2/graylog2-server#1339)

• Properly log errors in the Netty channel pipeline (Graylog2/graylog2-server#1340)

• Prevent creation of invalid alert conditions (Graylog2/graylog2-server#1332)

• Upgrade to Elasticsearch 1.6.2
Graylog 1.1.5

Released: 2015-07-27

https://www.graylog.org/graylog-1-1-5-released/

- Improve handling of exceptions in the JournallingMessageHandler (Graylog2/graylog2-server#1286)
- Upgrade to Elasticsearch 1.6.1 (Graylog2/graylog2-server#1312)
- Remove hard-coded limit for UDP receive buffer size (Graylog2/graylog2-server#1290)
- Ensure that `elasticsearch_index_prefix` is lowercase (commit 2173225)
- Add configuration option for time zone to `Date converter` (Graylog2/graylog2-server#1320)
- Fix NPE if the disk journal is disabled on a node (Graylog2/graylog2-web-interface#1520)
- Statistic and Chart error: Adding time zone offset caused overflow (Graylog2/graylog2-server#1257)
- Ignore stream alerts and throughput on serialize (Graylog2/graylog2-server#1309)
- Fix dynamic keyword time-ranges for dashboard widgets created from content packs (Graylog2/graylog2-server#1308)
- Upgraded Anonymous Usage Statistics plugin to version 1.1.1

Graylog 1.1.4

Released: 2015-06-30

https://www.graylog.org/graylog-v1-1-4-is-now-available/

- Make heartbeat timeout option for AmqpTransport optional. Graylog2/graylog2-server#1010
- Export as CSV on stream fails with “Invalid range type provided.” Graylog2/graylog2-web-interface#1504

Graylog 1.1.3

Released: 2015-06-19

https://www.graylog.org/graylog-v1-1-3-is-now-available/

- Log error message early if there is a MongoDB connection error. Graylog2/graylog2-server#1249
- Fixed field content value alert condition. Graylog2/graylog2-server#1245
- Extend warning about SO_RCVBUF size to UDP inputs. Graylog2/graylog2-server#1243
- Scroll on button dropdowns. Graylog2/graylog2-web-interface#1477
- Normalize graph widget numbers before drawing them. Graylog2/graylog2-web-interface#1479
- Fix highlight result checkbox position on old Firefox. Graylog2/graylog2-web-interface#1440
- Unescape terms added to search bar. Graylog2/graylog2-web-interface#1484
- Load another message in edit extractor page not working. Graylog2/graylog2-web-interface#1488
- Reader users aren’t able to export search results as CSV. Graylog2/graylog2-web-interface#1492
- List of streams not loaded on message details page. Graylog2/graylog2-web-interface#1496
Graylog Documentation, Release 2.2.1

Graylog 1.1.2

Released: 2015-06-10

https://www.graylog.org/graylog-v1-1-2-is-now-available/

- Get rid of NoSuchElementException if index alias doesn’t exist. Graylog2/graylog2-server#1218
- Make Alarm Callbacks API compatible to Graylog 1.0.x again. Graylog2/graylog2-server#1221, Graylog2/graylog2-server#1222, Graylog2/graylog2-server#1224
- Fixed issues with natural language parser for keyword time range. Graylog2/graylog2-server#1226
- Unable to write Graylog metrics to MongoDB Graylog2/graylog2-server#1228
- Unable to delete user. Graylog2/graylog2-server#1209
- Unable to unpause streams, despite editing permissions. Graylog2/graylog2-web-interface#1456
- Choose quick values widget size dynamically. Graylog2/graylog2-web-interface#1422
- Default field sort order is not guaranteed after reload. Graylog2/graylog2-web-interface#1436
- Toggling all fields in search list throws error and breaks pagination. Graylog2/graylog2-web-interface#1434
- Improve multi-line log messages support. Graylog2/graylog2-web-interface#612
- NPE when clicking a message from a deleted input on a stopped node. Graylog2/graylog2-web-interface#1444
- Auto created search syntax must use quotes for values with whitespaces in them. Graylog2/graylog2-web-interface#1448
- Quick Values doesn’t update for new field. Graylog2/graylog2-web-interface#1438
- New Quick Values list too large. Graylog2/graylog2-web-interface#1442
- Unloading referenced alarm callback plugin breaks alarm callback listing. Graylog2/graylog2-web-interface#1440
- Add to search button doesn’t work as expected for “level” field. Graylog2/graylog2-web-interface#1453
- Treat “*” query as empty query. Graylog2/graylog2-web-interface#1420
- Improve title overflow on widgets. Graylog2/graylog2-web-interface#1430
- Convert NaN to 0 on histograms. Graylog2/graylog2-web-interface#1417
- “&lt;&gt;” values in fields are unescaped and don’t display in Quick Values. Graylog2/graylog2-web-interface#1455
- New quickvalues are not showing number of terms. Graylog2/graylog2-web-interface#1411
- Default index for split &amp; index extractor results in an error. Graylog2/graylog2-web-interface#1464
- Improve behaviour when field graph fails to load. Graylog2/graylog2-web-interface#1276
- Unable to unpause streams, despite editing permissions. Graylog2/graylog2-web-interface#1456
- Wrong initial size of quick values pie chart. Graylog2/graylog2-web-interface#1469
- Problems refreshing data on quick values pie chart. Graylog2/graylog2-web-interface#1470
- Ignore streams with no permissions on message details. Graylog2/graylog2-web-interface#1472
Graylog 1.1.1

Released: 2015-06-05

https://www.graylog.org/graylog-v1-1-1-is-now-available/

- Fix problem with missing alarmcallbacks. Graylog2/graylog2-server#1214
- Add additional newline between messages to alert email. Graylog2/graylog2-server#1216
- Fix incorrect index range calculation. Graylog2/graylog2-server#1217, Graylog2/graylog2-web-interface#1266
- Fix sidebar auto-height on old Firefox versions. Graylog2/graylog2-web-interface#1410
- Fix “create one now” link on stream list page. Graylog2/graylog2-web-interface#1424
- Do not update StreamThroughput when unmounted. Graylog2/graylog2-web-interface#1428
- Fix position of alert annotations in search result histogram. Graylog2/graylog2-web-interface#1421
- Fix NPE when searching. Graylog2/graylog2-web-interface#1212
- Hide unlock dashboard link for reader users. Graylog2/graylog2-web-interface#1429
- Open radio documentation link on a new window. Graylog2/graylog2-web-interface#1427
- Use radio node page on message details. Graylog2/graylog2-web-interface#1423

Graylog 1.1.0

Released: 2015-06-04

https://www.graylog.org/graylog-1-1-is-now-generally-available/

- Properly set node_id on message input Graylog2/graylog2-server#1210
- Fixed handling of booleans in configuration forms in the web interface
- Various design fixes in the web interface

Graylog 1.1.0-rc.3

Released: 2015-06-02

https://www.graylog.org/graylog-v1-1-rc3-is-now-available/

- Unbreak server startup with collector thresholds set. Graylog2/graylog2-server#1194
- Adding verbal alert description to alert email templates and subject line defaults. Graylog2/graylog2-server#1158
- Fix message backlog in default body template in FormattedEmailAlertSender. Graylog2/graylog2-server#1163
- Make RawMessageEvent’s fields volatile to guard against cross-cpu visibility issues. Graylog2/graylog2-server#1207
- Set default for “disable_index_range_calculation” to “true”.
- Passing in value to text area fields in configuration forms. Graylog2/graylog2-web-interface#1340
- Stream list has no loading spinner. Graylog2/graylog2-web-interface#1309
• Showing a helpful notification when there are no active/inactive collectors.  Graylog2/graylog2-web-interface#1302
• Improve behavior when field graphs are stacked.  Graylog2/graylog2-web-interface#1348
• Keep new lines added by users on alert callbacks.  Graylog2/graylog2-web-interface#1270
• Fix duplicate metrics reporting if two components subscribed to the same metric on the same page.  Graylog2/graylog2-server#1199
• Make sidebar visible on small screens.  Graylog2/graylog2-web-interface#1390
• Showing warning and disabling edit button for output if plugin is missing.  Graylog2/graylog2-web-interface#1185
• Using formatted fields in old message loader.  Graylog2/graylog2-web-interface#1393
• Several styling and UX improvements

**Graylog 1.1.0-rc.1**

Released: 2015-05-27

https://www.graylog.org/graylog-v1-1-rc1-is-now-available/

• Unable to send email alerts.  Graylog2/graylog2-web-interface#1346
• “Show messages from this collector view” displays no messages.  Graylog2/graylog2-web-interface#1334
• Exception error in search page when using escaped characters.  Graylog2/graylog2-web-interface#1356
• Wrong timestamp on stream rule editor.  Graylog2/graylog2-web-interface#1328
• Quickvalue values are not linked to update search query.  Graylog2/graylog2-web-interface#1296
• Stream list has no loading spinner.  Graylog2/graylog2-web-interface#1309
• Collector list with only inactive collectors is confusing.  Graylog2/graylog2-web-interface#1302
• Update sockjs-client to 1.0.0.  Graylog2/graylog2-web-interface#1344
• Scroll to search bar when new query term is added.  Graylog2/graylog2-web-interface#1284
• Scroll to quick values if not visible.  Graylog2/graylog2-web-interface#1284
• Scroll to newly created field graphs.  Graylog2/graylog2-web-interface#1284
• Problems with websockets and even xhr streaming.  Graylog2/graylog2-web-interface#1344,  Graylog2/graylog2-web-interface#1353,  Graylog2/graylog2-web-interface#1338,  Graylog2/graylog2-web-interface#1322
• Add to search bar not working on sources tab.  Graylog2/graylog2-web-interface#1350
• Make field graphs work with streams.  Graylog2/graylog2-web-interface#1352
• Improved page design on outputs page.  Graylog2/graylog2-web-interface#1236
• Set startpage button missing for dashboards.  Graylog2/graylog2-web-interface#1345
• Generating chart for http response code is broken.  Graylog2/graylog2-web-interface#1358
Graylog 1.1.0-beta.3

Released: 2015-05-27

https://www.graylog.org/graylog-1-1-beta-3-is-now-available/

- Kafka inputs now support syslog, GELF and raw messages. Graylog2/graylog2-server#322
- Configurable timezone for the flexdate converter in extractors. Graylog2/graylog2-server#1166
- Allow decimal values for greater/smaller stream rules. Graylog2/graylog2-server#1101
- New configuration file option to control the default widget cache time. Graylog2/graylog2-server#1170
- Expose heartbeat configuration for AMQP inputs. Graylog2/graylog2-server#1010
- New alert condition to alert on field content. Graylog2/graylog2-server#537
- Add `<code>-Dwebsockets.enabled=false</code>` option for the web interface to disable websockets. Graylog2/graylog2-web-interface#1322
- Clicking the Graylog logo redirects to the custom startpage now. Graylog2/graylog2-web-interface#1315
- Improved reset and filter feature in sources tab. Graylog2/graylog2-web-interface#1337
- Fixed issue with stopping Kafka based inputs. Graylog2/graylog2-server#1171
- System throughput resource was always returning 0. Graylog2/graylog2-web-interface#1313
- MongoDB configuration problem with replica sets. Graylog2/graylog2-server#1173
- Syslog parser did not strip empty structured data fields. Graylog2/graylog2-server#1161
- Input metrics did not update after input has been stopped and started again. Graylog2/graylog2-server#1187
- NullPointerException with existing inputs in database fixed. Graylog2/graylog2-web-interface#1312
- Improved browser input validation for several browsers. Graylog2/graylog2-web-interface#1318
- Grok pattern upload did not work correctly. Graylog2/graylog2-web-interface#1321
- Internet Explorer 9 fixes. Graylog2/graylog2-web-interface#1319, Graylog2/graylog2-web-interface#1320
- Quick values feature did not work with reader users. Graylog2/graylog2-server#1169
- Replay link for keyword widgets was broken. Graylog2/graylog2-web-interface#1323
- Provide visual feedback when expanding message details. Graylog2/graylog2-web-interface#1283
- Allow filtering of saved searches again. Graylog2/graylog2-web-interface#1277
- Add back “Show details” link for global input metrics. Graylog2/graylog2-server#1168
- Provide visual feedback when dashboard widgets are loading. Graylog2/graylog2-web-interface#1324
- Restore preview for keyword time range selector. Graylog2/graylog2-web-interface#1280
- Fixed issue where widgets loading data looked empty. Graylog2/graylog2-web-interface#1324

Graylog 1.1.0-beta.2

Released: 2015-05-20

https://www.graylog.org/graylog-1-1-beta-is-now-available/

- CSV output streaming support including full text message
Graylog Documentation, Release 2.2.1

- Simplified MongoDB configuration with URI support
- Improved tokenizer for extractors
- Configurable UDP buffer size for incoming messages
- Enhanced Grok support with type conversions (integers, doubles and dates)
- Elasticsearch 1.5.2 support
- Added support for integrated Log Collector
- Search auto-complete
- Manual widget resize
- Auto resize of widgets based on screen size
- Faster search results
- Moved search filter for usability
- Updated several icons to text boxes for usability
- Search highlight toggle
- Pie charts (Stacked charts are coming too!)
- Improved stream management
- Output plugin and Alarm callback edit support
- Dashboard widget search edit
- Dashboard widget direct search button
- Dashboard background update support for better performance
- Log collector status UI

Graylog 1.0.2

Released: 2015-04-28

https://www.graylog.org/graylog-v1-0-2-has-been-released/

- Regular expression and Grok test failed when example message is a JSON document or contains special characters (Graylog2/graylog2-web-interface#1190, Graylog2/graylog2-web-interface#1195)
- “Show message terms” was broken (Graylog2/graylog2-web-interface#1168)
- Showing message indices was broken (Graylog2/graylog2-web-interface#1211)
- Fixed typo in SetIndexReadOnlyJob (Graylog2/graylog2-web-interface#1206)
- Consistent error messages when trying to create graphs from non-numeric values (Graylog2/graylog2-web-interface#1210)
- Fix message about too few file descriptors for Elasticsearch when number of file descriptors is unlimited (Graylog2/graylog2-web-interface#1220)
- Deleting output globally which was assigned to multiple streams left stale references (Graylog2/graylog2-server#1113)
- Fixed problem with sending alert emails (Graylog2/graylog2-server#1086)
- TokenizerConverter can now handle mixed quoted and un-quoted k/v pairs (Graylog2/graylog2-server#1083)
Graylog 1.0.1

Released: 2015-03-16

https://www.graylog.org/graylog-v1-0-1-has-been-released/

• Properly log stack traces (Graylog2/graylog2-server#970)
• Update REST API browser to new Graylog logo
• Avoid spamming the logs if the original input of a message in the disk journal can’t be loaded (Graylog2/graylog2-server#1005)
• Allows reader users to see the journal status (Graylog2/graylog2-server#1009)
• Compatibility with MongoDB 3.0 and Wired Tiger storage engine (Graylog2/graylog2-server#1024)
• Respect rest_transport_uri when generating entity URLs in REST API (Graylog2/graylog2-server#1020)
• Properly map NodeNotFoundException (Graylog2/graylog2-web-interface#1137)
• Allow replacing all existing Grok patterns on bulk import (Graylog2/graylog2-web-interface#1150)
• Configuration option for discarding messages on error in AMQP inputs (Graylog2/graylog2-server#1018)
• Configuration option of maximum HTTP chunk size for HTTP-based inputs (Graylog2/graylog2-server#1011)
• Clone alarm callbacks when cloning a stream (Graylog2/graylog2-server#990)
• Add hasField() and getField() methods to MessageSummary class (Graylog2/graylog2-server#923)
• Add per input parse time metrics (Graylog2/graylog2-web-interface#1106)
• Allow the use of https://logging.apache.org/log4j/extras/ log4j-extras classes in log4j configuration (Graylog2/graylog2-server#1042)
• Fix updating of input statistics for Radio nodes (Graylog2/graylog2-web-interface#1022)
• Emit proper error message when a regular expression in an Extractor doesn’t match example message (Graylog2/graylog2-web-interface#1157)
• Add additional information to system jobs (Graylog2/graylog2-server#920)
• Fix false positive message on LDAP login test (Graylog2/graylog2-web-interface#1138)
• Calculate saved search resolution dynamically (Graylog2/graylog2-web-interface#943)
• Only enable LDAP test buttons when data is present (Graylog2/graylog2-web-interface#1097)
• Load more than 1 message on Extractor form (Graylog2/graylog2-web-interface#1105)
• Fix NPE when listing alarm callback using non-existent plugin (Graylog2/graylog2-web-interface#1152)
• Redirect to nodes overview when node is not found (Graylog2/graylog2-web-interface#1137)
• Fix documentation links to integrations and data sources (Graylog2/graylog2-web-interface#1136)
• Prevent accidental indexing of web interface by web crawlers (Graylog2/graylog2-web-interface#1151)
• Validate grok pattern name on the client to avoid duplicate names (Graylog2/graylog2-server#937)
• Add message journal usage to nodes overview page (Graylog2/graylog2-web-interface#1083)
• Properly format numbers according to locale (Graylog2/graylog2-web-interface#1128, Graylog2/graylog2-web-interface#1129)
Graylog 1.0.0

Released: 2015-02-19

https://www.graylog.org/announcing-graylog-v1-0-ga/

  • No changes since Graylog 1.0.0-rc.4

Graylog 1.0.0-rc.4

Released: 2015-02-13

https://www.graylog.org/graylog-v1-0-rc-4-has-been-released/

  • Default configuration file locations have changed. Graylog2/graylog2-server#950
  • Improved error handling on search errors. Graylog2/graylog2-server#954
  • Dynamically update dashboard widgets with keyword range. Graylog2/graylog2-server#956, Graylog2/graylog2-web-interface#958
  • Prevent duplicate loading of plugins. Graylog2/graylog2-server#948
  • Fixed password handling when editing inputs. Graylog2/graylog2-web-interface#1103
  • Fixed issues getting Elasticsearch cluster health. Graylog2/graylog2-server#953
  • Better error handling for extractor imports. Graylog2/graylog2-server#942
  • Fixed structured syslog parsing of keys containing special characters. Graylog2/graylog2-server#845
  • Improved layout on Grok patterns page. Graylog2/graylog2-web-interface#1109
  • Improved formatting large numbers. Graylog2/graylog2-web-interface#1111
  • New Graylog logo.

Graylog 1.0.0-rc.3

Released: 2015-02-05

https://www.graylog.org/graylog-v1-0-rc-3-has-been-released/

  • Fixed compatibility with MongoDB version 2.2. Graylog2/graylog2-server#941
  • Fixed performance regression in process buffer handling. Graylog2/graylog2-server#944
  • Fixed data type for the max_size_per_index config option value. Graylog2/graylog2-web-interface#1100
  • Fixed problem with indexer error page. Graylog2/graylog2-web-interface#1102

Graylog 1.0.0-rc.2

Released: 2015-02-04

https://www.graylog.org/graylog-v1-0-rc-2-has-been-released/

  • Better Windows compatibility. Graylog2/graylog2-server#930
• Added helper methods for the plugin API to simplify plugin development.
• Fixed problem with input removal on radio nodes. Graylog2/graylog2-server#932
• Improved buffer information for input, process and output buffers. Graylog2/graylog2-web-interface#1096
• Fixed API return value incompatibility regarding node objects. Graylog2/graylog2-server#933
• Fixed reloading of LDAP settings. Graylog2/graylog2-server#934
• Fixed ordering of message input state labels. Graylog2/graylog2-web-interface#1094
• Improved error messages for journal related errors. Graylog2/graylog2-server#931
• Fixed browser compatibility for stream rules form. Graylog2/graylog2-web-interface#1095
• Improved grok pattern management. Graylog2/graylog2-web-interface#1098, Graylog2/graylog2-web-interface#1099

Graylog 1.0.0-rc.1

Released: 2015-01-28
https://www.graylog.org/graylog-v1-0-rc-1-has-been-released/
• Cleaned up internal metrics when input is terminating. Graylog2/graylog2-server#915
• Added Telemetry plugin options to example graylog.conf. Graylog2/graylog2-server#914
• Fixed problems with user permissions on streams. Graylog2/graylog2-web-interface#1058
• Added information about different rotation strategies to REST API. Graylog2/graylog2-server#913
• Added better error messages for failing inputs. Graylog2/graylog2-web-interface#1056
• Fixed problem with JVM options in bin/radioctl script. Graylog2/graylog2-server#918
• Fixed issue with updating input configuration. Graylog2/graylog2-server#919
• Fixed password updating for reader users by the admin. Graylog2/graylog2-web-interface#1075
• Enabled the message_journal_enabled config option by default. Graylog2/graylog2-server#924
• Add REST API endpoint to list reopened indices. Graylog2/graylog2-web-interface#1072
• Fixed problem with GELF stream output. Graylog2/graylog2-server#921
• Show an error message on the indices page if the Elasticsearch cluster is not available. Graylog2/graylog2-web-interface#1070
• Fixed a problem with stopping inputs. Graylog2/graylog2-server#926
• Changed output configuration display to mask passwords. Graylog2/graylog2-web-interface#1066
• Disabled message journal on radio nodes. Graylog2/graylog2-server#927
• Create new message representation format for search results in alarm callback messages. Graylog2/graylog2-server#923
• Fixed stream router to update the stream engine if a stream has been changed. Graylog2/graylog2-server#922
• Fixed focus problem in stream rule modal windows. Graylog2/graylog2-web-interface#1063
• Do not show new dashboard link for reader users. Graylog2/graylog2-web-interface#1057
• Do not show stream output menu for reader users. Graylog2/graylog2-web-interface#1059
• Do not show user forms of other users for reader users. Graylog2/graylog2-web-interface#1064
• Do not show permission settings in the user profile for reader users. Graylog2/graylog2-web-interface#1055
• Fixed extractor edit form with no messages available. Graylog2/graylog2-web-interface#1061
• Fixed problem with node details page and JVM locale settings. Graylog2/graylog2-web-interface#1062
• Improved page layout for Grok patterns.
• Improved layout for the message journal information. Graylog2/graylog2-web-interface#1084, Graylog2/graylog2-web-interface#1085
• Fixed wording on radio inputs page. Graylog2/graylog2-web-interface#1077
• Fixed formatting on indices page. Graylog2/graylog2-web-interface#1086
• Improved error handling in stream rule form. Graylog2/graylog2-web-interface#1076
• Fixed time range selection problem for the sources page. Graylog2/graylog2-web-interface#1080
• Several improvements regarding permission checks for user creation. Graylog2/graylog2-web-interface#1088
• Do not show stream alert test button for reader users. Graylog2/graylog2-web-interface#1089
• Fixed node processing status not updating on the nodes page. Graylog2/graylog2-web-interface#1090
• Fixed filename handling on Windows. Graylog2/graylog2-server#928, Graylog2/graylog2-server#732

Graylog 1.0.0-beta.2

Released: 2015-01-21

https://www.graylog.org/graylog-v1-0-beta-3-has-been-released/

• Fixed stream alert creation. Graylog2/graylog2-server#891
• Suppress warning message when PID file doesn’t exist. Graylog2/graylog2-server#889
• Fixed an error on outputs page with missing output plugin. Graylog2/graylog2-server#894
• Change default heap and garbage collector settings in scripts.
• Add extractor information to log message about failing extractor.
• Fixed problem in SplitAndIndexExtractor. Graylog2/graylog2-server#896
• Improved rendering time for indices page. Graylog2/graylog2-web-interface#1060
• Allow user to edit its own preferences. Graylog2/graylog2-web-interface#1049
• Fixed updating stream attributes. Graylog2/graylog2-server#902
• Stream throughput now shows combined value over all nodes. Graylog2/graylog2-web-interface#1047
• Fixed resource leak in JVM PermGen memory. Graylog2/graylog2-server#907
• Update to gelfclient-1.1.0 to fix DNS resolving issue. Graylog2/graylog2-server#882
• Allow arbitrary characters in user names (in fact in any resource url). Graylog2/graylog2-web-interface#1005, Graylog2/graylog2-web-interface#1006
• Fixed search result CSV export. Graylog2/graylog2-server#901
• Skip GC collection notifications for parallel collector. Graylog2/graylog2-server#899
• Shorter reconnect timeout for Radio AMQP connections. Graylog2/graylog2-server#900
• Fixed random startup error in Radio. Graylog2/graylog2-server#911
• Fixed updating an alert condition. Graylog2/graylog2-server#912
• Add system notifications for journal related warnings. Graylog2/graylog2-server#897
• Add system notifications for failing outputs. Graylog2/graylog2-server#741
• Improve search result pagination. Graylog2/graylog2-web-interface#834
• Improved regex error handling in extractor testing. Graylog2/graylog2-web-interface#1044
• Wrap long names for node metrics. Graylog2/graylog2-web-interface#1028
• Fixed node information progress bars. Graylog2/graylog2-web-interface#1046
• Improve node buffer utilization readability. Graylog2/graylog2-web-interface#1046
• Fixed username alert receiver form field. Graylog2/graylog2-web-interface#1050
• Wrap long messages without break characters. Graylog2/graylog2-web-interface#1052
• Hide list of node plugins if there aren’t any plugins installed.
• Warn user before leaving page with unpinned graphs. Graylog2/graylog2-web-interface#808

**Graylog 1.0.0-beta.2**

Released: 2015-01-16

https://www.graylog.org/graylog-v1-0-0-beta2/

• SIGAR native libraries are now found correctly (for getting system information)
• plugins can now state if they want to run in server or radio
• Fixed LDAP settings testing. Graylog2/graylog2-web-interface#1026
• Improved RFC5425 syslog message parsing. Graylog2/graylog2-server#845
• JVM arguments are now being logged on start. Graylog2/graylog2-server#875
• Improvements to log messages when Elasticsearch connection fails during start.
• Fixed an issue with AMQP transport shutdown. Graylog2/graylog2-server#874
• After index cycling the System overview page could be broken. Graylog2/graylog2-server#880
• Extractors can now be edited. Graylog2/graylog2-web-interface#549
• Fixed saving user preferences. Graylog2/graylog2-web-interface#1027
• Scripts now return proper exit codes. Graylog2/graylog2-server#886
• Grok patterns can now be uploaded in bulk. Graylog2/graylog2-server#377
• During extractor creation the test display could be offset. Graylog2/graylog2-server#804
• Performance fix for the System/Indices page. Graylog2/graylog2-web-interface#1035
• A create dashboard link was shown to reader users, leading to an error when followed. Graylog2/graylog2-web-interface#1032
• Content pack section was shown to reader users, leading to an error when followed. Graylog2/graylog2-web-interface#1033
• Failing stream outputs were being restarted constantly. Graylog2/graylog2-server#741
Graylog2 0.92.4

Released: 2015-01-14

https://www.graylog.org/graylog2-v0-92-4/

- [SERVER] Ensure that Radio inputs can only be started on server nodes (Graylog2/graylog2-server#843)
- [SERVER] Avoid division by zero when finding rotation anchor in the time-based rotation strategy (Graylog2/graylog2-server#836)
- [SERVER] Use username as fallback if display name in LDAP is empty (Graylog2/graylog2-server#837)

Graylog 1.0.0-beta.1

Released: 2015-01-12

https://www.graylog.org/graylog-v1-0-0-beta1/

- Message Journaling
- New Widgets
- Grok Extractor Support
- Overall stability and resource efficiency improvements
- Single binary for graylog2-server and graylog2-radio
- Inputs are now editable
- Order of field charts rendered inside the search results page is now maintained.
- Improvements in focus and keyboard behaviour on modal windows and forms.
- You can now define whether to disable expensive, frequent real-time updates of the UI in the settings of each user. (For example the updating of total messages in the system)
- Experimental search query auto-completion that can be enabled in the user preferences.
- The API browser now documents server response payloads in a better way so you know what to expect as an answer to your call.
- Now using the standard Java ServiceLoader for plugins.

Graylog2 0.92.3

Released: 2014-12-23

https://www.graylog.org/graylog2-v0-92-3/

- [SERVER] Removed unnecessary instrumentation in certain places to reduce GC pressure caused by many short living objects (Graylog2/graylog2-server#800)
- [SERVER] Limit Netty worker thread pool to 16 threads by default (see rest_worker_threads_max_pool_size in graylog2.conf)
- [WEB] Fixed upload of content packs when a URI path prefix (application.context in graylog2-web-interface.conf) is being used (Graylog2/graylog2-web-interface#1009)
- [WEB] Fixed display of metrics of type Counter (Graylog2/graylog2-server#795)
Graylog2 0.92.1

Released: 2014-12-11

https://www.graylog.org/graylog2-v0-92-1/

- [SERVER] Fixed name resolution and overriding sources for network inputs.
- [SERVER] Fixed wrong delimiter in GELF TCP input.
- [SERVER] Disabled the output cache by default. The output cache is the source of all sorts of interesting problems. If you want to keep using it, please read the upgrade notes.
- [SERVER] Fixed message timestamps in GELF output.
- [SERVER] Fixed connection counter for network inputs.
- [SERVER] Added warning message if the receive buffer size (SO_RECV) couldn’t be set for network inputs.
- [WEB] Improved keyboard shortcuts with most modal dialogs (e.g. hitting Enter submits the form instead of just closing the dialogs).
- [WEB] Upgraded to play2-graylog2 1.2.1 (compatible with Play 2.3.x and Java 7).

Graylog2 0.92.0

Released: 2014-12-01

https://www.graylog.org/graylog2-v0-92/

- [SERVER] IMPORTANT SECURITY FIX: It was possible to perform LDAP logins with crafted wildcards. (A big thank you to Jose Tozo who discovered this issue and disclosed it very responsibly.)
- [SERVER] Generate a system notification if garbage collection takes longer than a configurable threshold.
- [SERVER] Added several JVM-related metrics.
- [SERVER] Added support for Elasticsearch 1.4.x which brings a lot of stability and resilience features to Elasticsearch clusters.
- [SERVER] Made version check of Elasticsearch version optional. Disabling this check is not recommended.
- [SERVER] Added an option to disable optimizing Elasticsearch indices on index cycling.
- [SERVER] Added an option to disable time-range calculation for indices on index cycling.
- [SERVER] Lots of other performance enhancements for large setups (i.e. involving several Radio nodes and multiple Graylog2 Servers).
- [SERVER] Support for Syslog Octet Counting, as used by syslog-ng for syslog via TCP (#743)
- [SERVER] Improved support for structured syslog messages (#744)
- [SERVER] Bug fixes regarding IPv6 literals in mongodb_replica_set and elasticsearch_discovery_zen_ping_unicast_hosts
- [WEB] Added additional details to system notification about Elasticsearch max. open file descriptors.
- [WEB] Fixed several bugs and inconsistencies regarding time zones.
- [WEB] Improved graphs and diagrams
- [WEB] Allow to update dashboards when browser window is not on focus (#738)
- [WEB] Bug fixes regarding timezone handling
Graylog Documentation, Release 2.2.1

• Numerous internal bug fixes

Graylog2 0.92.0-rc.1

Released: 2014-11-21
https://www.graylog.org/graylog2-v0-92-rc-1/

• [SERVER] Generate a system notification if garbage collection takes longer than a configurable threshold.
• [SERVER] Added several JVM-related metrics.
• [SERVER] Added support for Elasticsearch 1.4.x which brings a lot of stability and resilience features to Elasticsearch clusters.
• [SERVER] Made version check of Elasticsearch version optional. Disabling this check is not recommended.
• [SERVER] Added an option to disable optimizing Elasticsearch indices on index cycling.
• [SERVER] Added an option to disable time-range calculation for indices on index cycling.
• [SERVER] Lots of other performance enhancements for large setups (i.e. involving several Radio nodes and multiple Graylog2 Servers).
• [WEB] Upgraded to Play 2.3.6.
• [WEB] Added additional details to system notification about Elasticsearch max. open file descriptors.
• [WEB] Fixed several bugs and inconsistencies regarding time zones.
• Numerous internal bug fixes

Graylog2 0.91.3

Released: 2014-11-05
https://www.graylog.org/graylog2-v0-90-3-and-v0-91-3-has-been-released/

• Fixed date and time issues related to DST changes
• Requires Elasticsearch 1.3.4; Elasticsearch 1.3.2 had a bug that can cause index corruptions.
• The mongodb_replica_set configuration variable now supports IPv6
• Messages read from the on-disk caches could be stored with missing fields

Graylog2 0.91.3

Released: 2014-11-05
https://www.graylog.org/graylog2-v0-90-3-and-v0-91-3-has-been-released/

• Fixed date and time issues related to DST changes
• The mongodb_replica_set configuration variable now supports IPv6
• Messages read from the on-disk caches could be stored with missing fields
Graylog2 0.92.0-beta.1

Released: 2014-11-05

https://www.graylog.org/graylog2-v0-92-beta-1/

- Content packs
- [SERVER] SSL/TLS support for Graylog2 REST API
- [SERVER] Support for time based retention cleaning of your messages. The old message count based approach is still the default.
- [SERVER] Support for Syslog Octet Counting, as used by syslog-ng for syslog via TCP (Graylog2/graylog2-server#743)
- [SERVER] Improved support for structured syslog messages (Graylog2/graylog2-server#744)
- [SERVER] Bug fixes regarding IPv6 literals in mongodb_replica_set and elasticsearch_discovery_zen_ping_unicast_hosts
- [WEB] Revamped “Sources” page in the web interface
- [WEB] Improved graphs and diagrams
- [WEB] Allow to update dashboards when browser window is not on focus (Graylog2/graylog2-web-interface#738)
- [WEB] Bug fixes regarding timezone handling
- Numerous internal bug fixes

Graylog2 0.91.1

Released: 2014-10-17

https://www.graylog.org/two-new-graylog2-releases/

- Messages written to the persisted master caches were written to the system with unreadable timestamps, leading to
- errors when trying to open the message.
- Extractors were only being deleted from running inputs but not from all inputs
- Output plugins were not always properly loaded
- You can now configure the alert_check_interval in your graylog2.conf
- Parsing of configured Elasticsearch unicast discovery addresses could break when including spaces

Graylog2 0.90.1

Released: 2014-10-17

https://www.graylog.org/two-new-graylog2-releases/

- Messages written to the persisted master caches were written to the system with unreadable timestamps, leading to
- errors when trying to open the message.
- Extractors were only being deleted from running inputs but not from all inputs
• Output plugins were not always properly loaded
• You can now configure the alert_check_interval in your graylog2.conf
• Parsing of configured Elasticsearch unicast discovery addresses could break when including spaces

Graylog2 0.91.0-rc.1

Released: 2014-09-23
https://www.graylog.org/graylog2-v0-90-has-been-released/
• Optional ElasticSearch v1.3.2 support

Graylog2 0.90.0

Released: 2014-09-23
https://www.graylog.org/graylog2-v0-90-has-been-released/
• Real-time data forwarding to Splunk or other systems
• Alert callbacks for greater flexibility
• New disk-based architecture for buffering in load spike situations
• Improved graphing
• Plugin API
• Huge performance and stability improvements across the whole stack
• Small possibility of losing messages in certain scenarios has been fixed
• Improvements to internal logging from threads to avoid swallowing Graylog2 error messages
•Paused streams are no longer checked for alerts
• Several improvements to timezone handling
• JavaScript performance fixes in the web interface and especially a fixed memory leak of charts on dashboards
• The GELF HTTP input now supports CORS
• Stream matching now has a configurable timeout to avoid stalling message processing in case of too complex rules or erroneous regular expressions
• Stability improvements for Kafka and AMQP inputs
• Inputs can now be paused and resumed
• Dozens of bug fixes and other improvements

Graylog2 0.20.3

Released: 2014-08-09
https://www.graylog.org/graylog2-v0-20-3-has-been-released/
• Bugfix: Storing saved searches was not accounting custom application contexts
• Bugfix: Editing stream rules could have a wrong a pre-filled value
• Bugfix: The create dashboard link was shown even if the user has no permission to so. This caused an ugly error page because of the missing permissions.
• Bugfix: graylog2-radio could lose numeric fields when writing to the message broker
• Better default batch size values for the Elasticsearch output
• Improved rest_transport_uri default settings to avoid confusion with loopback interfaces
• The deflector index is now also using the configured index prefix

Graylog2 0.20.2

Released: 2014-05-24

https://www.graylog.org/graylog2-v0-20-2-has-been-released/
• Search result highlighting
• Reintroduces AMQP support
• Extractor improvements and sharing
• Graceful shutdowns, Lifecycles, Load Balancer integration
• Improved stream alert emails
• Alert annotations
• CSV exports via the REST API now support chunked transfers and avoid heap size problems with huge result sets
• Login now redirects to page you visited before if there was one
• More live updating information in node detail pages
• Empty dashboards no longer show lock/unlock buttons
• Global inputs now also show IO metrics
• You can now easily copy message IDs into native clipboard with one click
• Improved message field selection in the sidebar
• Fixed display of floating point numbers in several places
• Now supporting application contexts in the web interface like http://example.org/graylog2
• Several fixes for LDAP configuration form
• Message fields in the search result sidebar now survive pagination
• Only admin users are allowed to change the session timeout for reader users
• New extractor: Copy whole input
• New converters: uppercase/lowercase, flexdate (tries to parse any string as date)
• New stream rule to check for presence or absence of fields
• Message processing now supports trace logging
• Better error message for ES discovery problems
• Fixes to GELF HTTP input and it holding open connections
• Some timezone fixes
• CSV exports now only contain selected fields
• Improvements for bin/graylog* control scripts
• UDP inputs now allow for custom receive buffer sizes
• Numeric extractor converter now supports floating point values
• Bugfix: Several small fixes to system notifications and closing them
• Bugfix: Carriage returns were not escaped properly in CSV exports
• Bugfix: Some AJAX calls redirected to the startpage when they failed
• Bugfix: Wrong sorting in sources table
• Bugfix: Quickvalues widget was broken with very long values
• Bugfix: Quickvalues modal was positioned wrong in some cases
• Bugfix: Indexer failures list could break when you had a lot of failures
• Custom application prefix was not working for field chart analytics
• Bugfix: Memory leaks in the dashboards
• Bugfix: NullPointerException when Elasticsearch discovery failed and unicast discovery was disabled
• Message backlog in alert emails did not always include the correct number of messages
• Improvements for message outputs: No longer only waiting for filled buffers but also flushing them regularly. This avoids problems that make Graylog2 look like it misses messages in cheap benchmark scenarios combined with only little throughput.
Graylog Enterprise, built on top of the Graylog open source platform, offers additional features that enable users to deploy Graylog at enterprise scale and apply Graylog to processes and workflows across the whole organization.

Please see the Graylog Enterprise Page for details.
Graylog Enterprise comes as a set of Graylog server plugins which need to be installed in addition to the Graylog open source setup.

**Requirements**

The following list shows the minimum required Graylog versions for the Graylog Enterprise plugins.

<table>
<thead>
<tr>
<th>Enterprise Version</th>
<th>Required Graylog Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0.0</td>
<td>2.0.0, 2.0.1</td>
</tr>
<tr>
<td>1.0.1</td>
<td>2.0.2, 2.0.3</td>
</tr>
<tr>
<td>1.2.0</td>
<td>2.1.0, 2.1.1, 2.1.2</td>
</tr>
<tr>
<td>1.2.1</td>
<td>2.1.3</td>
</tr>
<tr>
<td>2.2.0</td>
<td>2.2.0</td>
</tr>
<tr>
<td>2.2.1</td>
<td>2.2.1</td>
</tr>
</tbody>
</table>

**Installation**

Once you purchase a license for Graylog Enterprise, you will get download links for the tarballs and DEB/RPM packages in the confirmation mail.

**Note:** The Graylog Enterprise plugins need to be installed on all your Graylog nodes!

**Tarball**

The tarball includes the enterprise plugin JAR files.

```
$ tar -tzf graylog-enterprise-plugins-1.0.0.tgz
graylog-enterprise-plugins-1.0.0/plugin/graylog-plugin-archive-1.0.0.jar
graylog-enterprise-plugins-1.0.0/plugin/graylog-plugin-license-1.0.0.jar
```

Depending on the Graylog setup method you have used, you have to install the plugins into different locations.
Table 29.2: Plugin Installation Locations

<table>
<thead>
<tr>
<th>Installation Method</th>
<th>Directory</th>
</tr>
</thead>
<tbody>
<tr>
<td>Virtual Machine Appliances</td>
<td>/opt/graylog/plugins/</td>
</tr>
<tr>
<td>Operating System Packages</td>
<td>/usr/share/graylog-server/plugin/</td>
</tr>
<tr>
<td>Manual Setup</td>
<td>/&lt;extracted-graylog-tarball-path&gt;/plugin/</td>
</tr>
</tbody>
</table>

Also check the plugin_dir config option in your Graylog server configuration file. The default might have been changed.

Make sure to install the enterprise plugin JAR files alongside the other Graylog plugins. Your plugin directory should look similar to this after installing the enterprise plugins.

```
plugin/
|-- graylog-plugin-archive-1.0.0.jar
|-- graylog-plugin-collector-1.0.2.jar
|-- graylog-plugin-enterprise-integration-1.0.2.jar
|-- graylog-plugin-license-1.0.0.jar
|-- graylog-plugin-map-widget-1.0.2.jar
|-- graylog-plugin-pipeline-processor-1.0.0-beta.4.jar
|-- usage-statistics-2.0.2.jar
```

**DEB / RPM Package**

We also provide DEB and RPM packages for the enterprise plugins.

**Note:** These packages can only be used when you installed Graylog via the *Operating System Packages*!

**DEB**

Use the following command to install the plugins via dpkg.

```
$ sudo dpkg -i graylog-enterprise-plugins-1.0.0.deb
```

**RPM**

Use the following command to install the plugins via rpm.

```
$ sudo rpm -Uvh graylog-enterprise-plugins-1.0.0-1.noarch.rpm
```

**Server Restart**

After you installed the Graylog Enterprise plugins you have to restart each of your Graylog servers to load the plugins.

**Note:** We recommend restarting one server at a time!

You should see something like the following in your Graylog server logs. It indicates that the plugins have been successfully loaded.
Cluster Setup

If you run a Graylog cluster you need to add the enterprise plugins to every Graylog node. Additionally your load-balancer must route `/api/plugins/org.graylog.plugins.archive/` only to the Graylog master node. Future versions of Graylog will forward these requests automatically to the correct node.

License Installation

The Graylog Enterprise plugins require a valid license to use the additional features. Once you have purchased a license you can import it into your Graylog setup by going through the following steps.

1. As an admin user, open the System/License page from the menu in the web interface.
2. Click the Import new license button in the top right hand corner.
3. Copy the license text from the confirmation email and paste it into the text field.
4. The license should be valid and a preview of your license details should appear below the text field.
5. Click Import to activate the license.

The license automatically applies to all nodes in your cluster without the need to restart your server nodes.

Note: If there are errors, please check that you copied the entire license from the email without line breaks. The same license is also attached as a text file in case it is wrongly formatted in the email.
Graylog License Management

Import new Graylog license

License

Please add the license key for your cluster in the text field above.

Cancel  Import
Graylog enables you to configure a retention period to automatically delete older messages - this is to help you control the costs of storage in Elasticsearch. But we know it’s not ideal deciding between keeping less messages in Graylog or paying more for hardware. Additionally, many of you are required to store data for long periods of time due to compliance requirements like PCI or HIPAA.

The Archiving functionality allows you to archive log messages until you need to re-import them into Graylog for analysis. You can instruct Graylog to automatically archive log messages to compressed flat files on the local filesystem before retention cleaning kicks in and messages are deleted from Elasticsearch. Archiving also works through a REST call or the web interface if you don’t want to wait until retention cleaning to happen. We chose flat files for this because they are vendor agnostic so you will always be able to access your data.

You can then do whatever you want with the archived files: move them to cheap storage, write them on tape, or even print them out if you need to! If you need to search through archived data in the future, you can move any selection of archived messages back into the Graylog archive folder, and the web interface will enable you to temporarily import the archive so you can analyze the messages again in Graylog.

**Note:** The archive plugin is a commercial feature and part of Graylog Enterprise.

---

**Setup**

The archive plugin is a commercial Graylog feature that can be installed in addition to the Graylog open source server.

**Installation**

Please see the *Graylog Enterprise setup page* for details on how to install the Archive plugin.

**Configuration**

The archive plugin can be configured via the Graylog web interface and does not need any changes in the Graylog server configuration file.

In the web interface menu navigate to “System/Archives” and click “Configuration” to adjust the configuration.
Archive Options

There are several configuration options to configure the archive plugin.

Table 30.1: Configuration Options

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Backend</td>
<td>Backend on the master node where the archive files will be stored.</td>
</tr>
<tr>
<td>Max Segment Size</td>
<td>Maximum size (in bytes) of archive segment files.</td>
</tr>
<tr>
<td>Compression Type</td>
<td>Compression type that will be used to compress the archives.</td>
</tr>
<tr>
<td>Checksum Type</td>
<td>Checksum algorithm that is used to calculate the checksum for archives.</td>
</tr>
<tr>
<td>Restore index batch size</td>
<td>Elasticsearch batch size when restoring archive files.</td>
</tr>
<tr>
<td>Streams to archive</td>
<td>Streams that should be included in the archive.</td>
</tr>
</tbody>
</table>

Backend

The archived indices will be stored in a backend. A backend that stores the data in /tmp/graylog-archive is created when the server starts for the first time but you can create a new backend if you want to store the data in a different path.

Max Segment Size

When archiving an index, the archive job writes the data into segments. The Max Segment Size setting sets the size limit for each of these data segments.

This allows control over the file size of the segment files to make it possible to process them with tools which have a size limit for files.

Once the size limit is reached, a new segment file will be started.
Example:

```
/path/to/archive/
greylog_201/
  archive-metadata.json
  archive-segment-0.gz
  archive-segment-1.gz
  archive-segment-2.gz
```

### Compression Type

Archives will be compressed with gzip by default. This option can be changed to use a different compression type. The selected compression type has a big impact on the time it takes to archive an index. Gzip for example is pretty slow but has a great compression rate. Snappy and LZ4 are way faster but the archives will be bigger.

Here is a comparison between the available compression algorithms with test data.

<table>
<thead>
<tr>
<th>Type</th>
<th>Index Size</th>
<th>Archive Size</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>gzip</td>
<td>1 GB</td>
<td>134 MB</td>
<td>15 minutes, 23 seconds</td>
</tr>
<tr>
<td>Snappy</td>
<td>1 GB</td>
<td>291 MB</td>
<td>2 minutes, 31 seconds</td>
</tr>
<tr>
<td>LZ4</td>
<td>1 GB</td>
<td>266 MB</td>
<td>2 minutes, 25 seconds</td>
</tr>
</tbody>
</table>

**Note**: Results with your data may vary! Make sure to test the different compression types to find the one that is best for your data.

**Warning**: The current implementation of LZ4 is not compatible with the LZ4 CLI tools, thus decompressing the LZ4 archives outside of Graylog is currently not possible.

### Checksum Type

When writing archives Graylog computes a CRC32 checksum over the files. This option can be changed to use a different checksum algorithm.

The type of checksum depends on the use case. CRC32 and MD5 are quick to compute and a reasonable choice to be able to detect damaged files, but neither is suitable as protection against malicious changes in the files. Graylog also supports using SHA-1 or SHA-256 checksums which can be used to make sure the files were not modified, as they are cryptographic hashes.

The best choice of checksum types depends on whether the necessary system tools are installed to compute them later (not all systems come with a SHA-256 utility for example), speed of checksum calculation for larger files as well as the security considerations.

### Restore Index Batch Size

This setting controls the batch size for re-indexing archive data into Elasticsearch. When set to 1,000, the restore job will re-index the archived data in document batches of 1,000.

You can use this setting to control the speed of the restore process and also how much load it will generate on the Elasticsearch cluster. The higher the batch size, the faster the restore will progress and the more load will be put on your Elasticsearch cluster in addition to the normal message processing.
Make sure to tune this **carefully** to avoid any negative impact on your message indexing throughput and search speed!

### Streams To Archive

This option can be used to select which streams should be included in the archive. With this you are able to archive only your important data instead of archiving everything that is arriving in Graylog.

**Note:** New streams will be archived automatically. If you create a new stream and don’t want it to be archived, you have to disable it in this configuration dialog.

### Backends

A backend can be used to store the archived data. For now, we only support a single file system backend type.

### File System

The archived indices will be stored in the *Output base path* directory. This directory **needs to exist and be writable for the Graylog server process** so the files can be stored.

**Note:** Only the **master** node needs access to the *Output base path* directory because the archiving process runs on the master node.

We recommend to put the *Output base path* directory onto a **separate disk or partition** to avoid any negative impact on the message processing should the archiving fill up the disk.

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>A simple title to identify the backend.</td>
</tr>
<tr>
<td>Description</td>
<td>Longer description for the backend.</td>
</tr>
<tr>
<td>Output base path</td>
<td>Directory path where the archive files should be stored.</td>
</tr>
</tbody>
</table>

### Output base path

The output base path can either be a simple directory path string or a template string to build dynamic paths.
You could use a template string to store the archive data in a directory tree that is based on the archival date.

Example:

```
# Template
/data/graylog-archive/${year}/${month}/${day}

# Result
/data/graylog-archive/2017/04/01/graylog_0
```

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>${year}</td>
<td>Archival date year. (e.g. “2017”)</td>
</tr>
<tr>
<td>${month}</td>
<td>Archival date month. (e.g. “04”)</td>
</tr>
<tr>
<td>${day}</td>
<td>Archival date day. (e.g. “01”)</td>
</tr>
<tr>
<td>${hour}</td>
<td>Archival date hour. (e.g. “23”)</td>
</tr>
<tr>
<td>${minute}</td>
<td>Archival date minute. (e.g. “24”)</td>
</tr>
<tr>
<td>${second}</td>
<td>Archival date second. (e.g. “59”)</td>
</tr>
<tr>
<td>${index-name}</td>
<td>Name of the archived index. (e.g. “graylog_0”)</td>
</tr>
</tbody>
</table>

**Index Retention**

Graylog is using configurable index retention strategies to delete old indices. By default indices can be *closed* or *deleted* if you have more than the configured limit.

The archive plugin offers a new index retention strategy that you can configure to automatically archive an index before closing or deleting it.

Index retention strategies can be configured in the system menu under “System/Indices”. Select an index set and click “Edit” to change the index rotation and retention strategies.
As with the regular index retention strategies, you can configure a max number of Elasticsearch indices. Once there are more indices than the configured limit, the oldest ones will be archived into the backend and then closed or deleted. You can also decide to not do anything (NONE) after archiving an index. In that case no cleanup of old indices will happen and you have to take care of that yourself!

Usage

Creating Archives

There are three ways to create archives from the Graylog Elasticsearch indices.

- **Web Interface**
- **Index Retention**
- **REST API**

**Web Interface**

You can manually create an archive on the “System/Archives” page in the web interface.
On the “Create Archive for Index” section of the page is a form where you can select an index and archive it by pressing “Archive Index”.

Using this will just archive the index to disk and does not close it or delete it. This is a great way to test the archiving feature without changing your index retention configuration.

Index Retention

The archive plugin ships with an index retention strategy that can be used to automatically create archives before closing or deleting Elasticsearch indices.

This is the easiest way to automatically create archives without custom scripting.

Please see the Index Retention Configuration on how to configure it.

REST API

The archive plugin also offers a REST API that you can use to automate archive creation if you have some special requirements and need a more flexible way to do this.

An index can be archived with a simple curl command:

```bash
$ curl -s -u admin -X POST http://127.0.0.1:9000/api/plugins/org.graylog.plugins.archive/archives/graylog_386
Enter host password for user 'admin': ***************
```

```
{
  "archive_job_config": {
    "archive_path": "/tmp/graylog-archive",
  }
```
That command started a system job in the Graylog server to create an archive for index graylog_386. The system_job.id can be used to check the progress of the job.

The REST API can be used to automate other archive related tasks as well, like restoring and deleting archives or updating the archive config. See the REST API browser on your Graylog server for details.

### Restoring Archives

**Note:** The restore process adds load to your Elasticsearch cluster because all messages are basically re-indexed. Please make sure to keep this in mind and test with smaller archives to see how your cluster behaves. Also use the **Restore Index Batch Size** setting to control the Elasticsearch batch size on re-index.

The archive plugin offer two ways to restore archived indices.

- **Web Interface**
- **REST API**

The archive plugin restores all indices into the “Restored Archives” index set to avoid conflicts with the original indices. (should those still exist)

Restored indices are also marked as **reopened** so they are **ignored** by index retention jobs and are not closed or deleted. That means you have to manually delete any restored indices **manually** once you do not need them anymore.
Web Interface

In the web interface you can restore an archive on the “System/Archives” page by selecting an archive from the list, open the archive details and clicking the “Restore Index” button.

REST API

As with archive creation you can also use the REST API to restore an archived index into the Elasticsearch cluster:

```bash
$ curl -s -u admin -X POST http://127.0.0.1:9000/api/plugins/org.graylog.plugins.archive/archives/graylog_386/restore
```

Enter host password for user 'admin': ***************

```
{
    "archive_metadata": {
        "archive_id": "graylog_307",
        "index_name": "graylog_307",
        "document_count": 491906,
        "created_at": "2016-04-14T14:31:50.787Z",
        "creation_duration": 142663,
        "timestamp_min": "2016-04-14T14:00:01.008Z",
        "timestamp_max": "2016-04-14T14:29:27.639Z",
        "id_mappings": {
            "streams": {
            "56fbafe0fbl21a5309cef297": "nginx requests"
            },
            "inputs": {
            "56fbafe0fbl21a5309cef290": "nginx error_log",
            "56fbafe0fbl21a5309cef28d": "nginx access_log"
            },
            "nodes": {
            "c5df7bff-cafd-4546-ac0a-5cc2ba4c847": "graylog.example.org"
            }
        }
    }
}
```
The returned JSON payload contains the archive metadata and the system job description that runs the index restore process.

### Restore into a separate cluster

As said earlier, restoring archived indices slow down your indexing speed because of added load. If you want to completely avoid adding more load to your Elasticsearch cluster, you can restore the archived indices on a different cluster.

To do that, you only have to transfer the archived indices to a different machine and put them into a configured Backend.

Each index archive is in a separate directory, so if you only want to transfer one index to a different machine, you only have to copy the corresponding directory into the backend.

Example:

```
$ tree /tmp/graylog-archive
/tmp/graylog-archive
 -- graylog_171
  | -- archive-metadata.json
  | -- archive-segment-0.gz
  -- graylog_201
  | -- archive-metadata.json
  | -- archive-segment-0.gz
  -- graylog_268
```
Searching in Restored Indices

Once an index has been restored from an archive it will be used by search queries automatically.

Every message that gets restored into an Elasticsearch index gets a special `gl2_archive_restored` field with value `true`. This allows you to only search in restored messages by using a query like:

```
exists:gl2_archive_restored AND <your search query>
```

Example:

```
_if exists (gl2_archive_restored) { _exists_:gl2_archive_restored AND <your search query> }
```

If you want to exclude all restored messages from your query you can use:

```
_missing_:gl2_archive_restored AND <your search query>
```
The Audit Log plugin keeps track of changes made by users to a Graylog system. It records all state changes into the database and makes it possible to search, filter and export all audit log entries.

**Note:** The audit log plugin is a commercial feature and part of Graylog Enterprise.

### Setup

The Audit Log plugin is a commercial Graylog feature that can be installed in addition to the Graylog open source server.

### Installation

Please see the [Graylog Enterprise setup page](#) for details on how to install the Audit Log plugin.

**Note:** Make sure the Audit Log plugin is installed on every node in your Graylog cluster.

### Configuration

The audit log plugin provides two ways of writing audit log entries:

1. Database
2. Log file via log4j2 appender

Logging to the database is always enabled and cannot be disabled.

**Note:** All configuration needs to be done in the Graylog server configuration file and in the logging configuration. (only if the log4j2 appender is enabled) Check the [default file locations page](#) for details.

The web interface can show the current configuration.
Database Configuration Options

The default MongoDB audit log has a few configuration options available.

Table 31.1: Configuration Options

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>auditlog_mongodb_keep_entries</code></td>
<td>delete audit log entries older that configured interval</td>
</tr>
<tr>
<td><code>auditlog_mongodb_cleanup_interval</code></td>
<td>interval of the audit log entry cleanup job</td>
</tr>
<tr>
<td><code>auditlog_mongodb_collection</code></td>
<td>the MongoDB collection to store the audit log entries in</td>
</tr>
</tbody>
</table>

`auditlog_mongodb_keep_entries`

This configures the interval after which old audit log entries in the MongoDB database will be deleted. You have to use values like `90d` (90 days) to configure the interval.

**Warning:** Make sure to configure this to fit your needs. Deleted audit log entries are gone forever!

The default value for this is `365d`.

Example:

```
auditlog_mongodb_keep_entries = 365d
```

`auditlog_mongodb_cleanup_interval`

This configures the interval of the background job that periodically deletes old audit log entries from the MongoDB database. You have to use values like `1h` (1 hour) to configure the interval.

The default value for this is `1h`. 

---

376 Chapter 31. Audit Log
Example:

```
auditlog_mongodb_cleanup_interval = 1h
```

**auditlog_mongodb_collection**

This configures the name of the MongoDB collection where the audit log plugin stores the audit log entries. The default value for this is `audit_log`.

Example:

```
auditlog_mongodb_collection = audit_log
```

### Log4j2 Configuration Options

The optional log4j2 audit log appender has a few configuration options available.

**Note:** To configure the log4j2 appender you have to edit the Graylog server configuration file and the log4j2.xml file for your setup!

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>auditlog_log4j_enabled</code></td>
<td>whether the log4j2 appender is enabled or not</td>
</tr>
<tr>
<td><code>auditlog_log4j_logger_name</code></td>
<td>log4j2 logger name</td>
</tr>
<tr>
<td><code>auditlog_log4j_marker_name</code></td>
<td>log4j2 marker name</td>
</tr>
</tbody>
</table>

**auditlog_log4j_enabled**

The log4j2 audit log appender is disabled by default and can be enabled by setting this option to `true`. The default value for this is `false`.

Example:

```
auditlog_log4j_enabled = true
```

**auditlog_log4j_logger_name**

This configures the log4j2 logger name of the audit log. The default value for this is `gl.org.graylog.plugins.auditlog`.

Example:

```
auditlog_log4j_logger_name = graylog-auditlog
```
auditlog_log4j_marker_name

This configures the log4j2 marker name for the audit log.

The default value for this is AUDIT_LOG.

Example:

```
auditlog_log4j_marker_name = AUDIT_LOG
```

Log4j2 Appender Configuration

To write audit log entries into a file you have to enable the log4j2 appender in your Graylog configuration file and add some configuration to the log4j2.xml file that is used by your server process.

The log4j2.xml file location is dependent on your deployment method. So please check the default file locations page.

An existing log4j2.xml config file needs another <Logger/> statement in the <Loggers/> section and an additional appender in the <Appenders/> section of the file.

**Warning:** The file on your system might look different than the following example. Make sure to only add the audit log related snippets to your config and do not remove anything else!

Example log4j2.xml file with audit log enabled:

```xml
<?xml version="1.0" encoding="UTF-8"?>
<Configuration packages="org.graylog2.log4j" shutdownHook="disable">
  <Appenders>
    <!-- Graylog server log file appender -->
    <RollingFile name="rolling-file" fileName="/var/log/graylog-server/server.log" filePattern="/var/log/graylog-server/server.log.%i.gz">
      <PatternLayout pattern="%d{yyyy-MM-dd'T'HH:mm:ss.SSSXXX} %-5p [%c{1}] %m%n"/>
      <Policies>
        <SizeBasedTriggeringPolicy size="50MB"/>
      </Policies>
      <DefaultRolloverStrategy max="10" fileIndex="min"/>
    </RollingFile>
    <!-- ##################################################### -->
    <!-- Rotate audit logs daily -->
    <RollingFile name="AUDITLOG" fileName="/var/log/graylog-server/audit.log" filePattern="/var/log/graylog-server/audit-%d{yyyy-MM-dd}.log.gz">
      <PatternLayout>%d - %m - %X%n</PatternLayout>
      <Policies>
        <TimeBasedTriggeringPolicy />
      </Policies>
    </RollingFile>
    <!-- ##################################################### -->
  </Appenders>
  <Loggers>
    <Logger name="org.graylog2" level="info"/>
    <!-- Graylog Audit Log. The logger name has to match the "auditlog_log4j_logger_name" setting in the Graylog configuration file
    <Logger name="graylog-auditlog" level="info" additivity="false">
      <AppenderRef ref="AUDITLOG"/>
```
The config snippets between the <!-- # --> tags have been added to the existing log4j2.xml file. Make sure that the name in the <Logger /> tag matches the configured auditlog_log4j_logger_name in your Graylog server configuration. Otherwise you will not see any log entries in the log file.

Caveats

You have to make sure that the log4j2 related settings in the Graylog server config file and the log4j2.xml file are the same on every node in your cluster!

Since every Graylog server writes its own audit log entries when the plugin is installed, the log files configured in the log4j2.xml file are written on every node. But only the entries from the local node will show up in that file.

If you have more than one node, you have to search in all configured files on all nodes to get a complete view of the audit trail.

Usage

Once you installed the Audit Log plugin, Graylog will automatically write audit log entries into the database.

View Audit Log Entries

The plugin adds a new page to the web interface which can be reached via “System/Audit Log”. You can view and export existing audit log entries in the database.

It also provides a simple search form to search and filter for audit events you are interested in.
Expand Event Details

Every row in the audit event entry table is clickable. Once clicked it will reveal the details of the audit event. All audit events have static fields like *actor*, *object* and others. In addition to that, every event has some event specific fields.

The fields on the left side in the details are the static fields every event has and the fields on the right side are the event specific fields.
Search & Filter

To make it easier to get to the audit log entries you need, the audit log UI provides a simple query language to search and filter the audit log entries.

You can either enter one or more words into the search field or choose to look for some specific fields in the audit log entries.

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>actor</td>
<td>the user that triggered the audit event</td>
</tr>
<tr>
<td>namespace</td>
<td>the namespace of the audit event; might be different in plugins</td>
</tr>
<tr>
<td>object</td>
<td>the object of the audit event; what has been changed</td>
</tr>
<tr>
<td>action</td>
<td>name of the action that has been executed on the object</td>
</tr>
<tr>
<td>success_status</td>
<td>if the action failed or succeeded</td>
</tr>
<tr>
<td>message</td>
<td>the actual audit event message</td>
</tr>
</tbody>
</table>

Search for text in the message

If you just want to find some text in the audit event message, you can enter the word you are looking for into the search bar.

Search for specific fields

You can also filter the entries for specific fields like the `actor`.

If you want to filter for all events triggered by the user `jane` you can enter `actor:jane` into the search bar.

Maybe you want to filter for events for more than one actor. That can be done by using either `actor:jane,john` or `actor:jane actor:john`.

Or you want to find all audit events which have not been triggered by a user. Add a `−` in front of the field name to negate the condition. To show all events except those created by user `jane` you can add `−actor:jane` to the search field.

You can mix and match several field queries to find the entries you need. Here are some more examples.

- `actor:jane,john −namespace:server` get all events by users `jane` and `john` which are not in the `server` namespace
- `index action:create` get all events which have the word `index` in the event message and where the action is `create`
- `message:index action:create` same as above, just with an explicit field selector for the message field
Export Entries

If the simple entry viewer is not enough, you can also export the result of your query as JSON or CSV to further process it.

The “Export Results” button next to the search bar can be used to do that.

**Note:** The export from the UI is currently limited to the newest 10,000 entries. Use the REST API if you need a bigger export.

Export via REST API

If you want to backup the audit log entries or make them available to another system, you can use the REST API to export them.

Example:

```
# Export 20,000 audit log entries in JSON format

# Export 5,000 audit log entries with actor "jane" in CSV format
```

**Note:** Make sure the `query` parameter is properly escaped if it contains whitespace.
Graylog Enterprise 2.2.3
Released: 2017-04-04

**Plugin: Archive**
- Metadata is now stored in MongoDB
- Preparation for storage backend support

Graylog Enterprise 2.2.2
Released: 2017-03-02

**Plugin: Audit Log**
- Extend integration with the Archive plugin

Graylog Enterprise 2.2.1
Released: 2017-02-20

**Plugin: Archive**
- Improve stability and smaller UI fixes

Graylog Enterprise 2.2.0
Released: 2017-02-09
Plugin: Archive

- Improve index set support

Graylog Enterprise 1.2.1

Released: 2017-01-26

Plugin: Archive

- Prepare the plugin to be compatible with the new default stream.

Plugin: Audit Log

- Add support for index sets and fix potential NPEs.
- Smaller UI improvements.

Graylog Enterprise 1.2.0

Released: 2016-09-14

https://www.graylog.org/blog/70-announcing-graylog-enterprise-v1-2

Plugin: Archive

- Add support for selecting which streams should be included in your archives.

Plugin: Audit Log

New plugin to keep track of changes made by users to a Graylog system by automatically saving them in MongoDB.

Graylog Enterprise 1.1

Released: 2016-09-01

- Added support for Graylog 2.1.0.

Graylog Enterprise 1.0.1

Released: 2016-06-08

Bugfix release for the archive plugin.
Plugin: Archive

Fixed problem when writing multiple archive segments

There was a problem when exceeding the max segment size so that multiple archive segments are written. The problem has been fixed and wrongly written segments can be read again.

Graylog Enterprise 1.0.0

Released: 2016-05-27
Initial Release including the Archive plugin.

Plugin: Archive

New features since the last beta plugin:

- Support for multiple compression strategies. (Snappy, LZ4, Gzip, None)