Log management with Graylog2

Lennart Koopmann, Kieker Days 2012

About me

24 years old, Software Engineer at XING AG
Hamburg, Germany

@_lennart

Graylog2

Free and open source log management system
Started in early 2010 by me
100.000 downloads in total until today
v0.10.0 will be released in a few days
Just started a company behind it

www.graylog2.org / @graylog2

What is this log management stuff?

Even grepping over flat files can be log management.

Log Management Maturity Scale

Log management can be done on different levels - Raffael Marty set up a scale for that.

Level 0

Do not collect logs at all.

Level 1

Collect logs. Mostly simple log files from email or HTTP servers.

Level 2

Use the logs for forensics and troubleshooting. Why was that email (not) sent out? Why was that HTTP 500 thrown?

Level 3

Save searches. The most basic case would be to save a grep command you used.

Level 4

Share searches. Store that search command somewhere so co-workers can find and use it to solve similar problems.

Level 5

Reporting. Easily generate reports from your logs. How many exceptions did we have this week, how many last week? Charts, PDF, stuff managers love!

Level 6

Alerting. Automate some of your troubleshooting tasks. Be warned automatically instead of waiting for a user to complain.

Level 7

Collect more logs! We may need more log sources for some use cases. Firewall logs, Router logs, even physical access logs.

Level 8

Correlation. Manual analysis of all that new data may take too long. Correlate different sources.

Level 9

Visual analysis.

Next levels

Pattern detection, interactive visualization, dynamic queries, anomaly detection, more sharing, ...

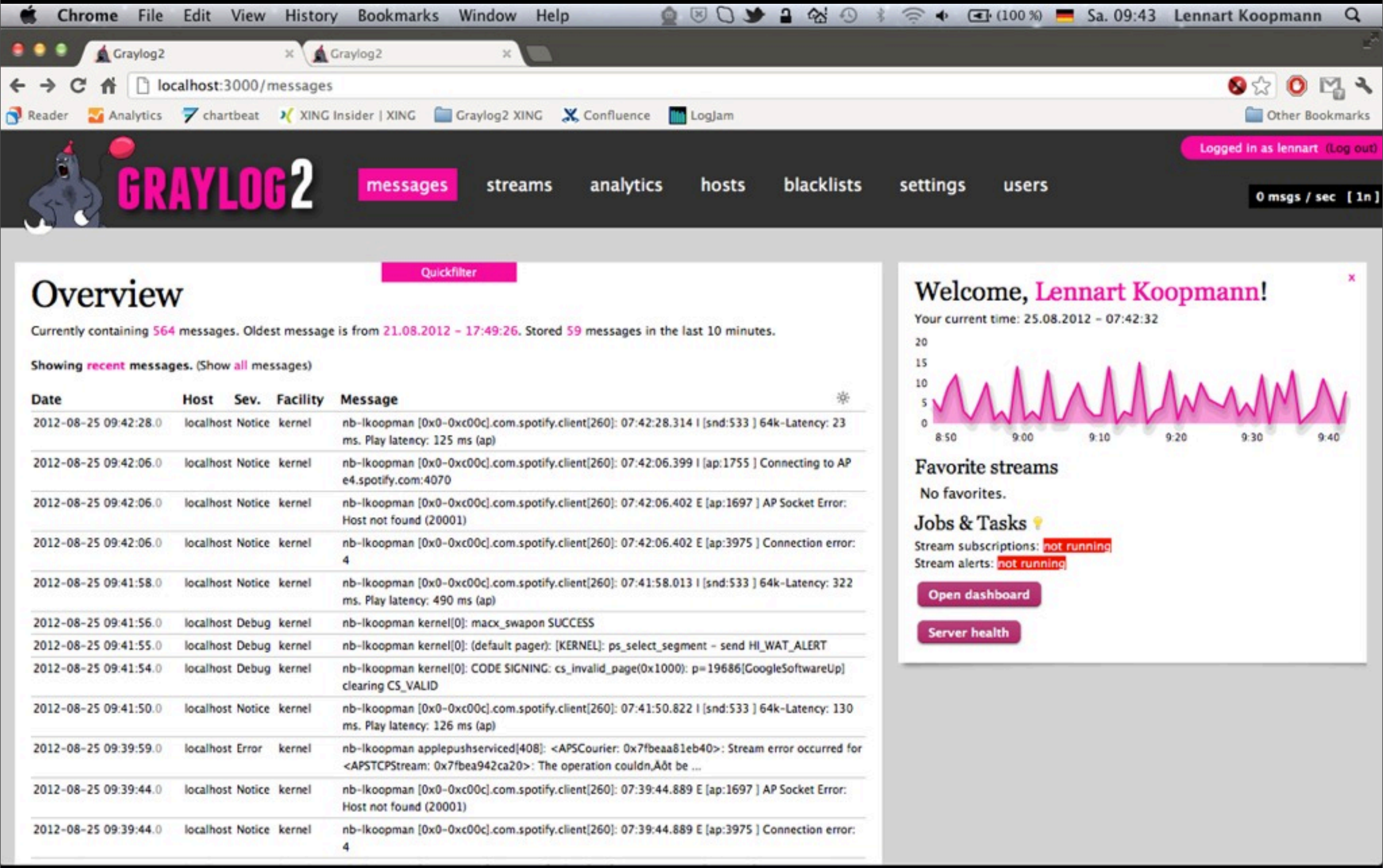
You need a central place to send your logs to

You need a central place to send your logs to

graylog2-server

You need a central place to make use of those logs

graylog2-web-interface



Overview

Currently containing 564 messages. Oldest message is from 21.08.2012 - 17:49:26. Stored 59 messages in the last 10 minutes.

Showing recent messages. (Show all messages)

Date	Host	Sev.	Facility	Message
2012-08-25 09:42:28.0	localhost	Notice	kernel	nb-lkoopman [0x0-0xc00c].com.spotify.client[260]: 07:42:28.314 I [snd:533] 64k-Latency: 23 ms. Play latency: 125 ms (ap)
2012-08-25 09:42:06.0	localhost	Notice	kernel	nb-lkoopman [0x0-0xc00c].com.spotify.client[260]: 07:42:06.399 I [ap:1755] Connecting to AP e4.spotify.com:4070
2012-08-25 09:42:06.0	localhost	Notice	kernel	nb-lkoopman [0x0-0xc00c].com.spotify.client[260]: 07:42:06.402 E [ap:1697] AP Socket Error: Host not found (20001)
2012-08-25 09:42:06.0	localhost	Notice	kernel	nb-lkoopman [0x0-0xc00c].com.spotify.client[260]: 07:42:06.402 E [ap:3975] Connection error: 4
2012-08-25 09:41:58.0	localhost	Notice	kernel	nb-lkoopman [0x0-0xc00c].com.spotify.client[260]: 07:41:58.013 I [snd:533] 64k-Latency: 322 ms. Play latency: 490 ms (ap)
2012-08-25 09:41:56.0	localhost	Debug	kernel	nb-lkoopman kernel[0]: macx_swapon SUCCESS
2012-08-25 09:41:55.0	localhost	Debug	kernel	nb-lkoopman kernel[0]: (default pager): [KERNEL]: ps_select_segment - send HI_WAT_ALERT
2012-08-25 09:41:54.0	localhost	Debug	kernel	nb-lkoopman kernel[0]: CODE SIGNING: cs_invalid_page(0x1000): p=19686[GoogleSoftwareUp] clearing CS_VALID
2012-08-25 09:41:50.0	localhost	Notice	kernel	nb-lkoopman [0x0-0xc00c].com.spotify.client[260]: 07:41:50.822 I [snd:533] 64k-Latency: 130 ms. Play latency: 126 ms (ap)
2012-08-25 09:39:59.0	localhost	Error	kernel	nb-lkoopman applepushserviced[408]: <APSCourier: 0x7fbeaa81eb40>: Stream error occurred for <APSTCPStream: 0x7fbea942ca20>: The operation couldn,Äôt be ...
2012-08-25 09:39:44.0	localhost	Notice	kernel	nb-lkoopman [0x0-0xc00c].com.spotify.client[260]: 07:39:44.889 E [ap:1697] AP Socket Error: Host not found (20001)
2012-08-25 09:39:44.0	localhost	Notice	kernel	nb-lkoopman [0x0-0xc00c].com.spotify.client[260]: 07:39:44.889 E [ap:3975] Connection error: 4

Welcome, Lennart Koopmann!

Your current time: 25.08.2012 - 07:42:32

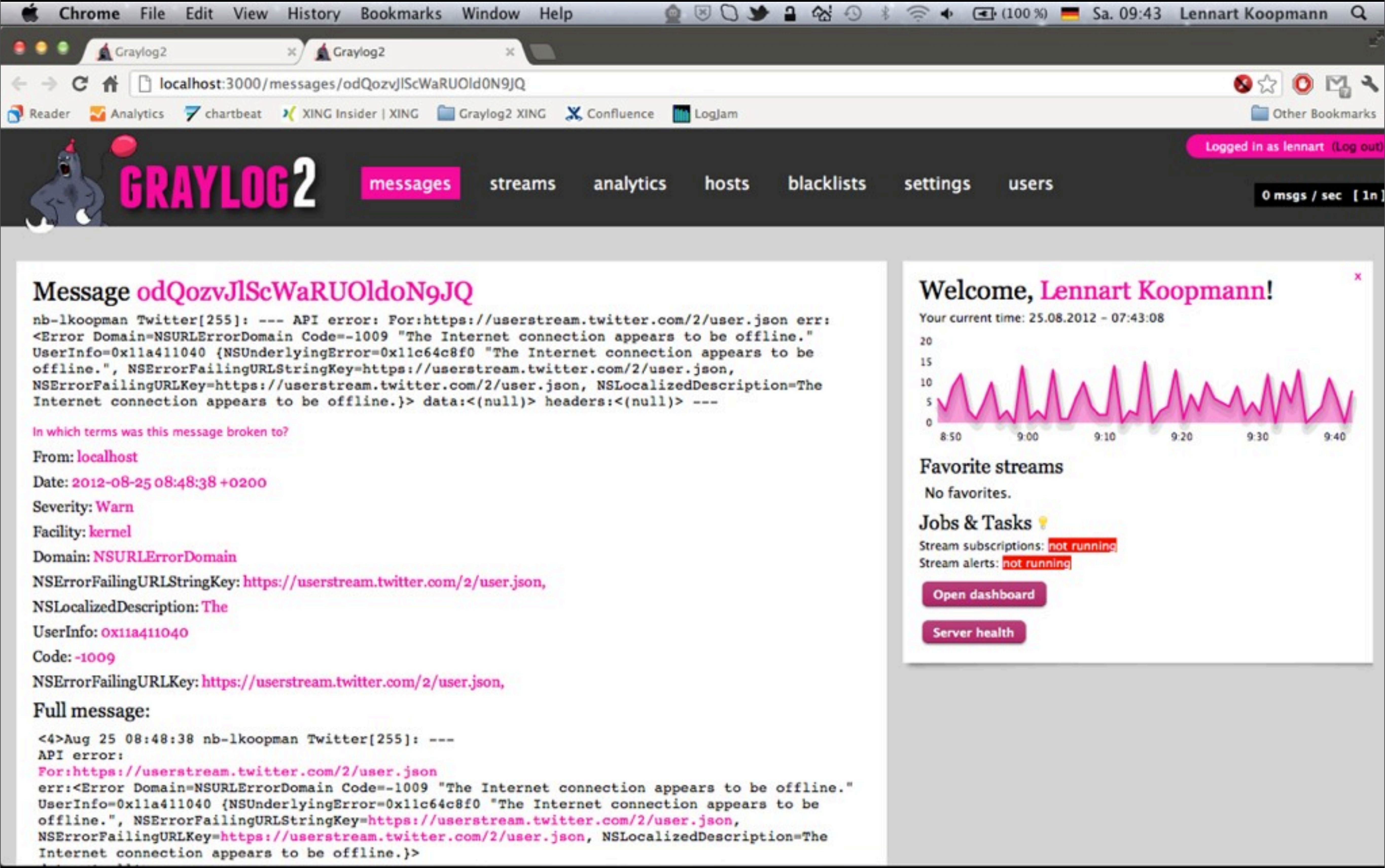
Favorite streams

No favorites.

Jobs & Tasks

Stream subscriptions: not running

Stream alerts: not running



Message odQozvJlScWaRUOldoN9JQ

```
nb-lkoopman Twitter[255]: --- API error: For:https://userstream.twitter.com/2/user.json err:
<Error Domain=NSURLErrorDomain Code=-1009 "The Internet connection appears to be offline."
UserInfo=0x11a411040 {NSUnderlyingError=0x11c64c8f0 "The Internet connection appears to be
offline.", NSErrorFailingURLStringKey=https://userstream.twitter.com/2/user.json,
NSErrorFailingURLKey=https://userstream.twitter.com/2/user.json, NSLocalizedDescription=The
Internet connection appears to be offline.}> data:<(null)> headers:<(null)> ---
```

In which terms was this message broken to?

From: localhost

Date: 2012-08-25 08:48:38 +0200

Severity: Warn

Facility: kernel

Domain: NSURLErrorDomain

NSErrorFailingURLStringKey: https://userstream.twitter.com/2/user.json,

NSLocalizedDescription: The

UserInfo: 0x11a411040

Code: -1009

NSErrorFailingURLKey: https://userstream.twitter.com/2/user.json,

Full message:

```
<4>Aug 25 08:48:38 nb-lkoopman Twitter[255]: ---
API error:
For:https://userstream.twitter.com/2/user.json
err:<Error Domain=NSURLErrorDomain Code=-1009 "The Internet connection appears to be offline."
UserInfo=0x11a411040 {NSUnderlyingError=0x11c64c8f0 "The Internet connection appears to be
offline.", NSErrorFailingURLStringKey=https://userstream.twitter.com/2/user.json,
NSErrorFailingURLKey=https://userstream.twitter.com/2/user.json, NSLocalizedDescription=The
Internet connection appears to be offline.}>
```

Welcome, Lennart Koopmann!

Your current time: 25.08.2012 - 07:43:08

Favorite streams

No favorites.

Jobs & Tasks

Stream subscriptions: not running

Stream alerts: not running

How to send your logs

Classic syslog via TCP/UDP
GELF via TCP/UDP/HTTP
both via AMQP

...or write your own input plugins.

GELF

Graylog Extended Log Format - Lets you structure your logs.

Many libraries for different systems and languages available.

Example GELF message

```
{
  ‘short_message’ : ‘Something went wrong’,
  ‘host’ : ‘some-host-1.example.org’,
  ‘severity’ : 2,
  ‘facility’ : ‘some subsystem’,
  ‘full_message’ : ‘Stacktrace and stuff’,
  ‘file’ : ‘some_controller.rb’,
  ‘line’ : 7,
  ‘_from_load_balancer’ : ‘lb-3’,
  ‘_user_id’ :  9001,
  ‘_http_response_code’ : 500
}
```

Two different types of log messages

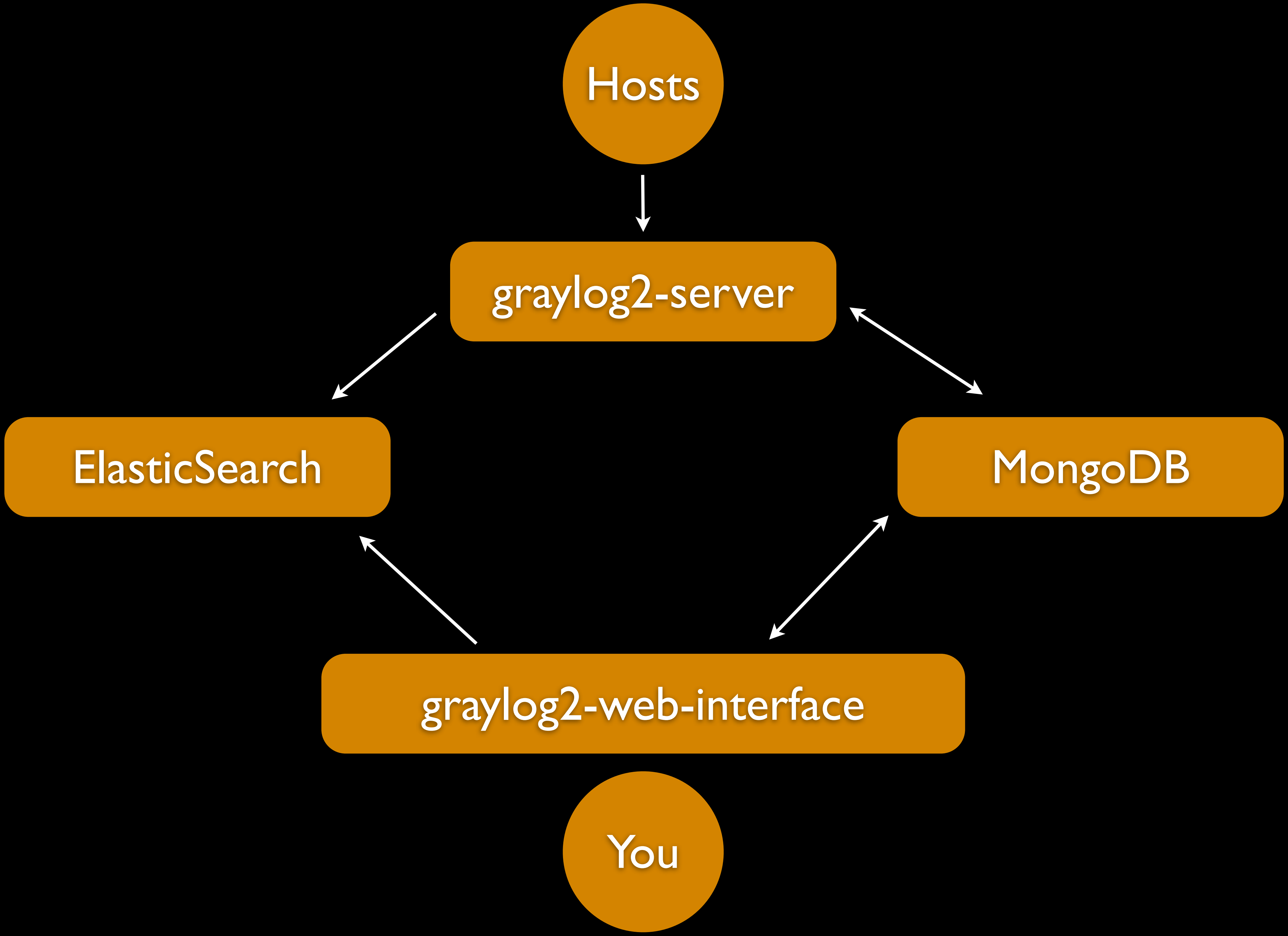
Type I

Automatically generated from a service. Usually huge amount of structured but raw data. You have only limited control about what is logged.

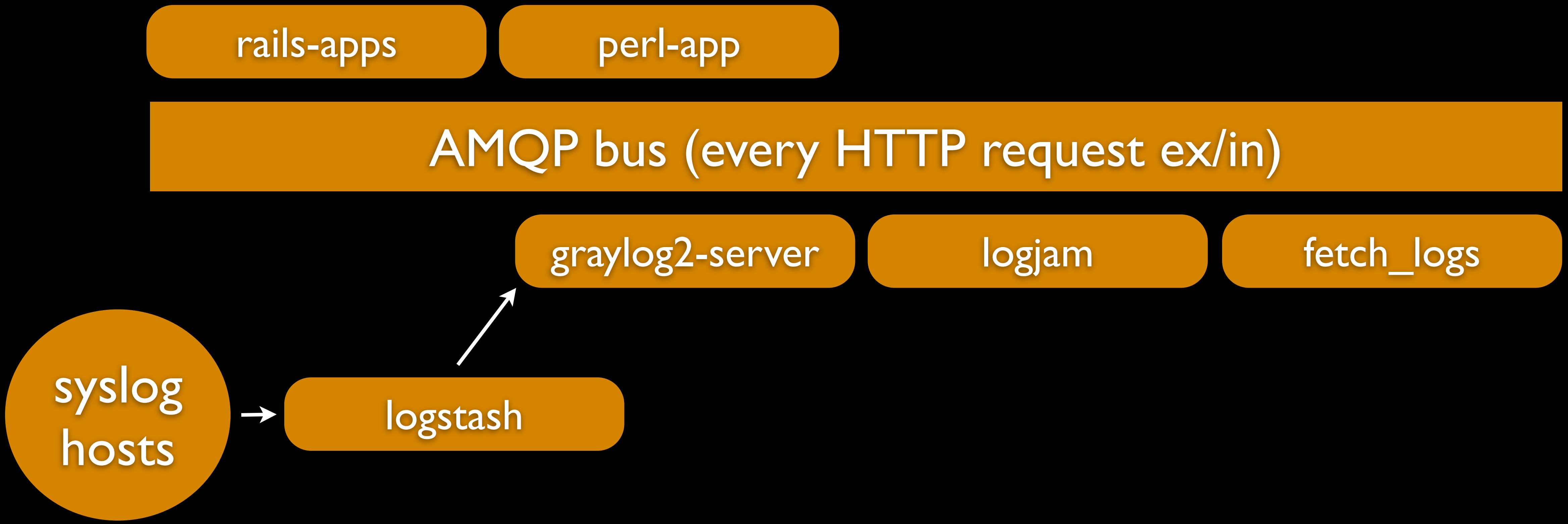
Type II

Logs directly sent from within your applications. Triggered for example by a log.error() call or an exception catcher. Possible to send highly structured via GELF.

Architecture



How we do it



Around 5000 msgs / sec

Decouple logging

Sending of a log message must never interrupt application flow. Think about timeouts, latency, DNS problems, ...

Extract information to fields

Have fields like user_id, http_response_code, _processed_controller, _processed_action, ...

...and get powerful analytics

```
_oauth_consumer_key = ‘acbd18db4cc2f85cedef654fccc4a4d8’
_processed_controller = ‘NewsfeedController’
_processed_action = ‘comment’
_http_method = ‘DELETE’
```

Is the iPhone application using a specific API call?
Old versions using that call getting used less over time?

facility = ‘rails-app’
_source = ‘applogs’
_exception_type = ‘CSRFProtect::AuthenticityTokenMissingException’

You enforced the authenticity token CSRF protection system wide. This will go live next Wednesday. Prepare a stream that catches all Exceptions related to that. See where you forgot to adapt view/form behavior.

```
facility = ‘rails-app’
_source = ‘applogs’
_exception_type = ‘CSRFProtect::AuthenticityTokenMissingException’
```

...and get a distribution of controllers that produce this error / were forgotten to adapt. Analytics shell:

```
stream(csrf).distribution({_processed_controller})

-> Events::PostingsController (5347), Profile::SettingsController (459)
```

```
facility = ‘rails-app’
_source = ‘applogs’
_exception_type = ‘CSRFProtect::AuthenticityTokenMissingException’
```

...or the specific actions in a controller

```
stream(csrf).distribution({_processed_action}, _processed_controller = ‘Events::PostingsController’)
```

-> destroy (42), create (5302)

```
_http_response_code = 404
_from_lb = ‘lb-3’
```

Is there an imbalance of HTTP 404s caused by different load balancers? (possible config problem)

```
all.distribution({_from_lb}, _http_response_code = 404)
```

```
> lb-1 (5732), lb-2 (69), lb-3 (45), lb-4 (22)
```

host = “router-1”
level = 3

Are there routers with especially high error rates?

all.distribution({host}, level >= 4)

> router-1 (0), router-2 (0), router-3 (0)

```
_processed_controller = “Session”
_processed_action = “create”
_http_return_code = 301
```

How many signups did you have today?

```
all.count(_processed_controller = ‘Session’, _processed_action = ‘create’, _http_return_code = 301)

> 294358
```

```
_processed_controller = “Session”
_processed_action = “create”
_http_return_code = 301
```

...and how many failed? Graph this!

```
all.count(_processed_controller = ‘Session’, _processed_action = ‘create’, _http_return_code = 200)

> 10452
```

Use streams for monitoring

Forward counts to Graphite, Librato or other tools

Be alerted if certain watermarks are reached or subscribe to streams in total.

Forward messages to other hosts, feed Graphite, feed team dashboards, unleash the power of your logs!

Extract everything you could ever need! How much you can do with the logs strongly depends on the message quality.

Coming in v0.10.0:

Complete server re-write

- Plugin system
- Better performance
- Possible to run multiple server instances in parallel

Better web interface performance

- Multiple indices
- Recent index

More new stuff like LDAP integration, more visualizations and analytics, JavaScript fixes, improvements and bugfixes!