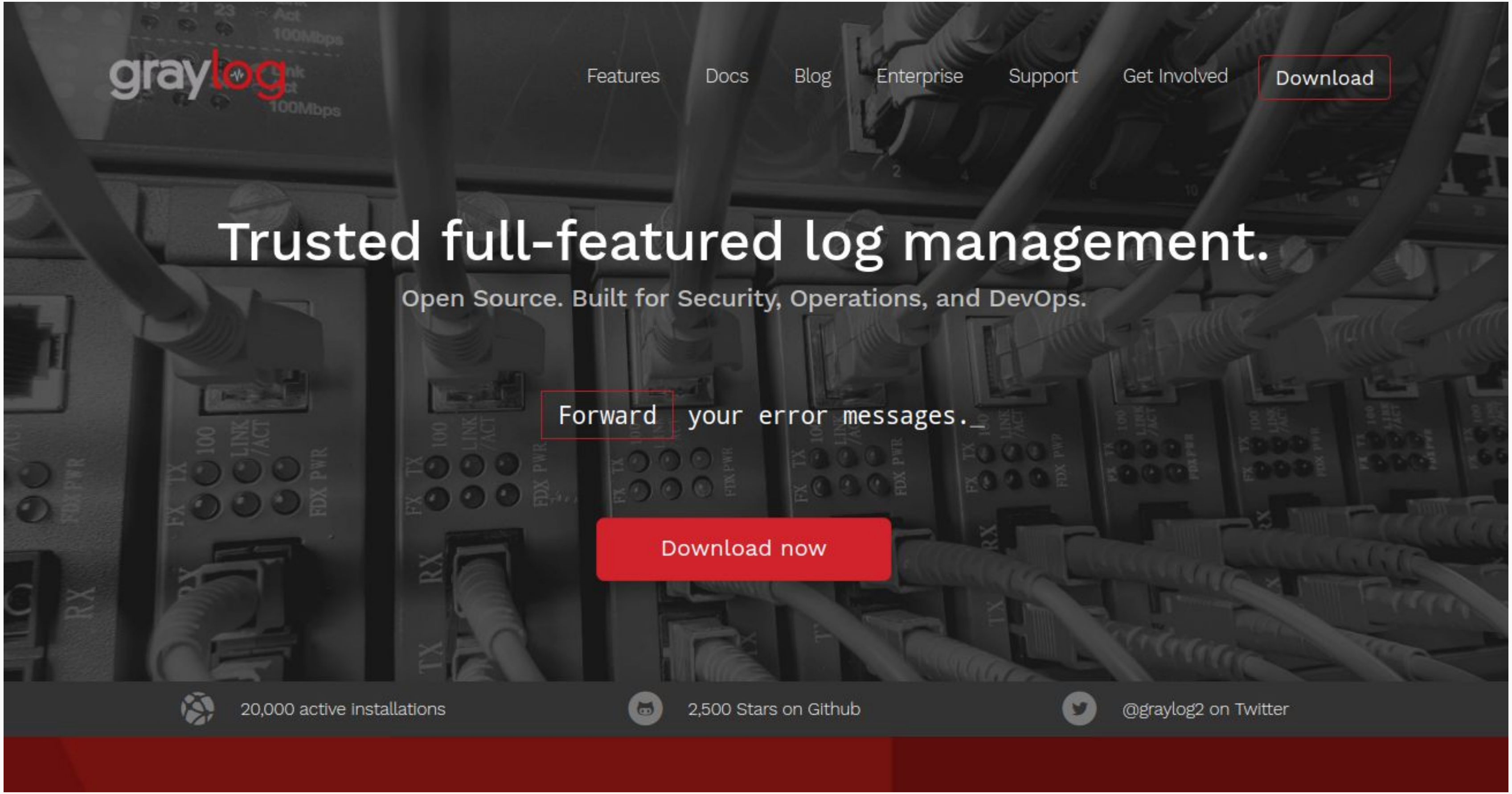
Log-Monitoring mit graylog

Thomas Darimont
Java User Group Saarland - 28. Treffen

16.02.2017

Graylog Website

https://www.graylog.org



Trusted full-featured log management.

Open Source. Built for Security, Operations, and DevOps.

Forward your error messages._

20,000 active installations

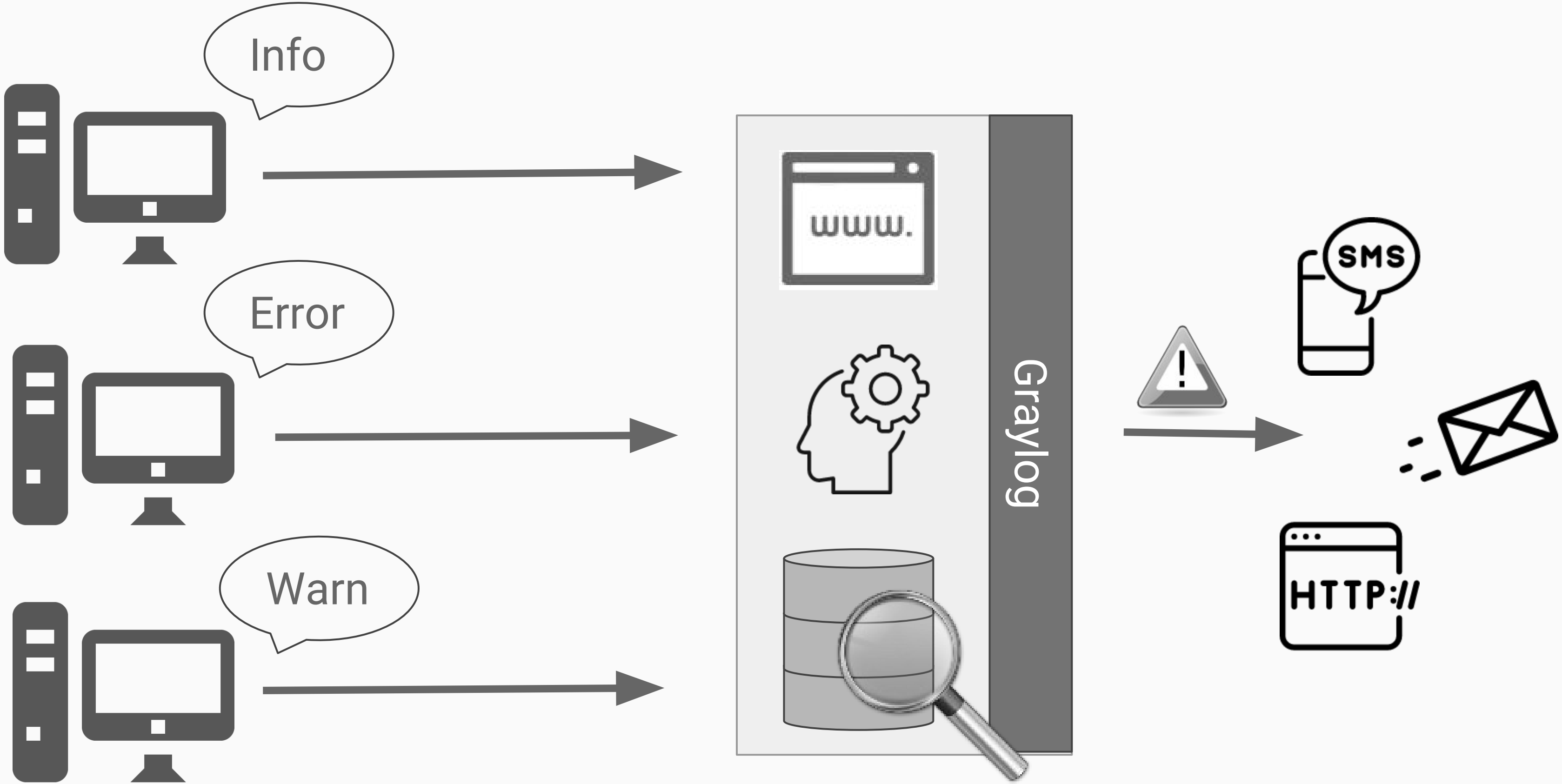
2,500 Stars on Github

@graylog2 on Twitter

Graylog in a Nutshell

- Log Management Platform
- Collect, Index and Analyze Log data
- Structured and Unstructured
- Java based, Open Source GPLv3
- Uses Elasticsearch & MongoDB
- Multi-User

What is Graylog?

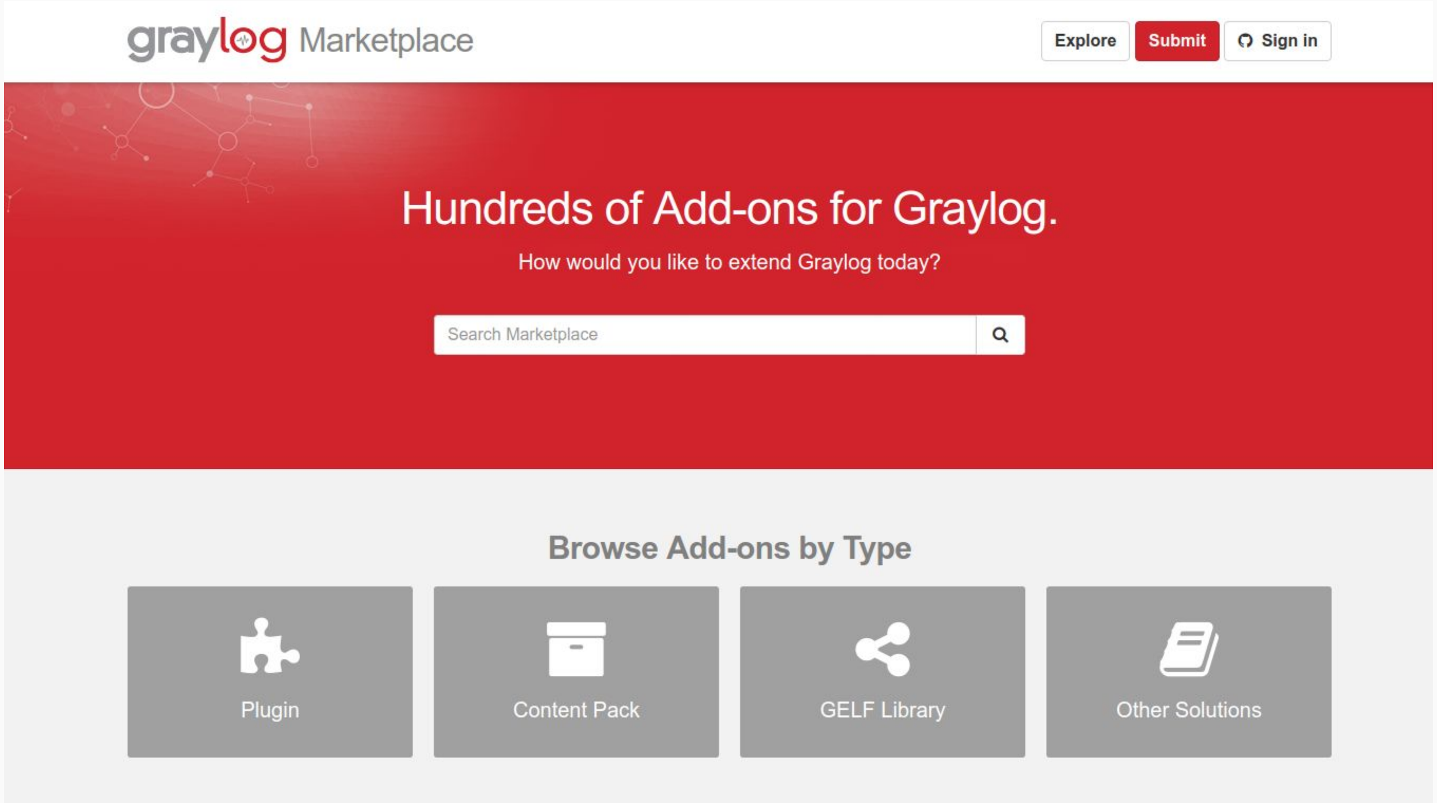


Graylog Facts

- Current Version 2.2.0 (Released 14. February)
- Very mature project > 6 years
- Docker, OVA Appliance, Standalone
- Free & Commercial (Graylog Inc.)
- Free Version quite powerful
- Enterprise: Support, Audit-Trail, Archiving++
- Trusted by Leading Companies (> 20.000 Installs)
- Graylog Marketplace

Graylog Marketplace

https://marketplace.graylog.org

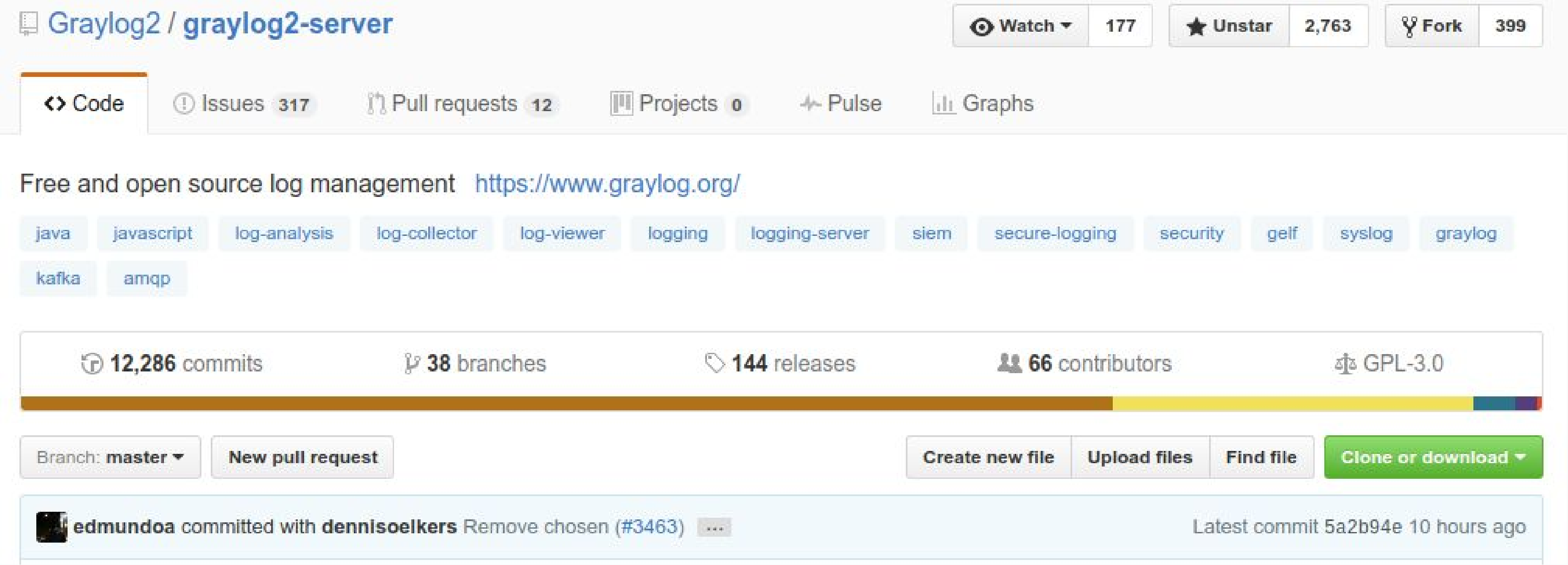


Hundreds of Add-ons for Graylog.

How would you like to extend Graylog today?

Browse Add-ons by Type

Graylog on Github



Graylog2 / graylog2-server

Free and open source log management https://www.graylog.org/

edmundoa committed with dennisoelkers Remove chosen (#3463) … Latest commit 5a2b94e 10 hours ago

https://github.com/Graylog2/graylog2-server

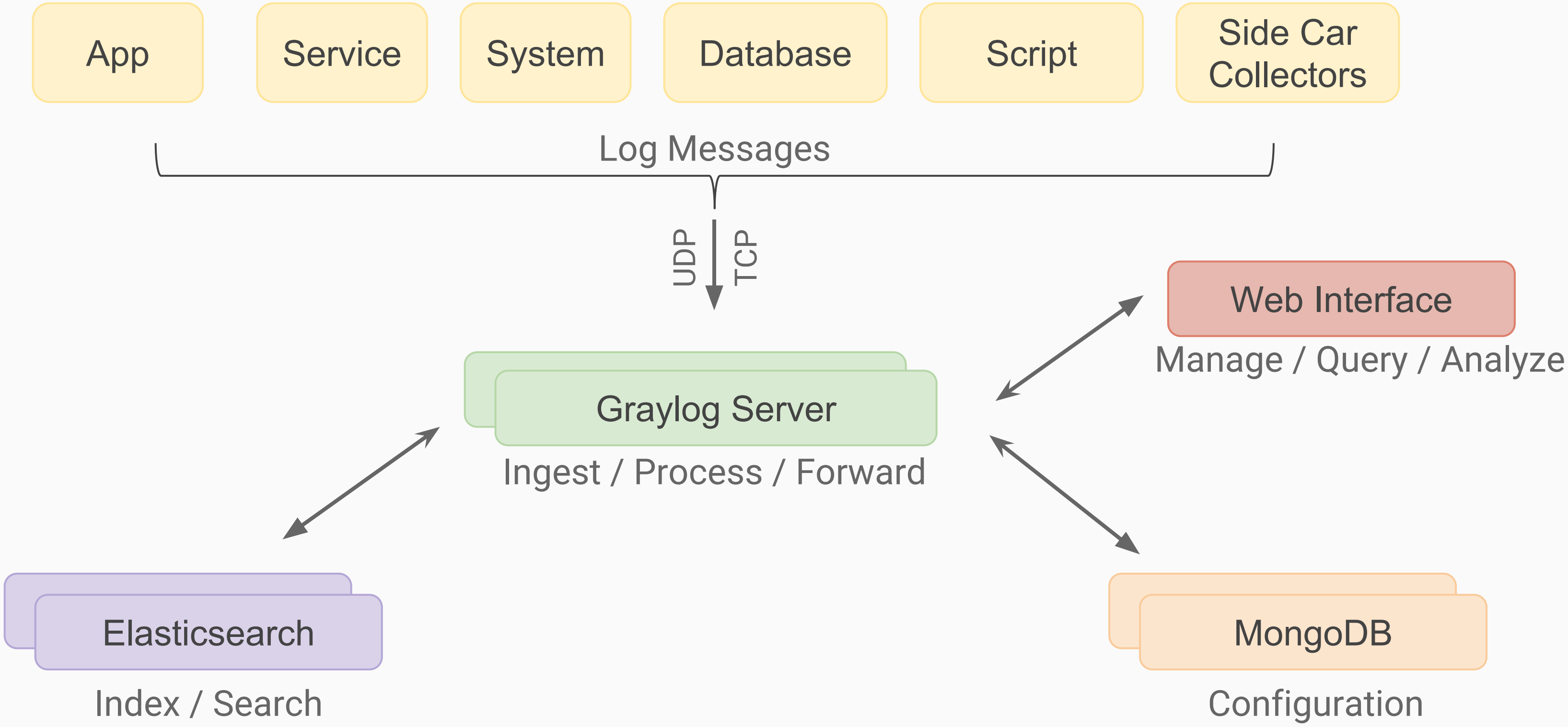
Graylog Features

- Multiple Formats
  - SYSLOG, GELF, Beats, JSON, Plaintext, Raw,...
- Multiple Protocols
  - TCP, UDP, HTTP, AMQP, Kafka, ...
- Log Message Classification
- User Management and Access Control
- Scalable Architecture with HA support
- High Performance Log Processing

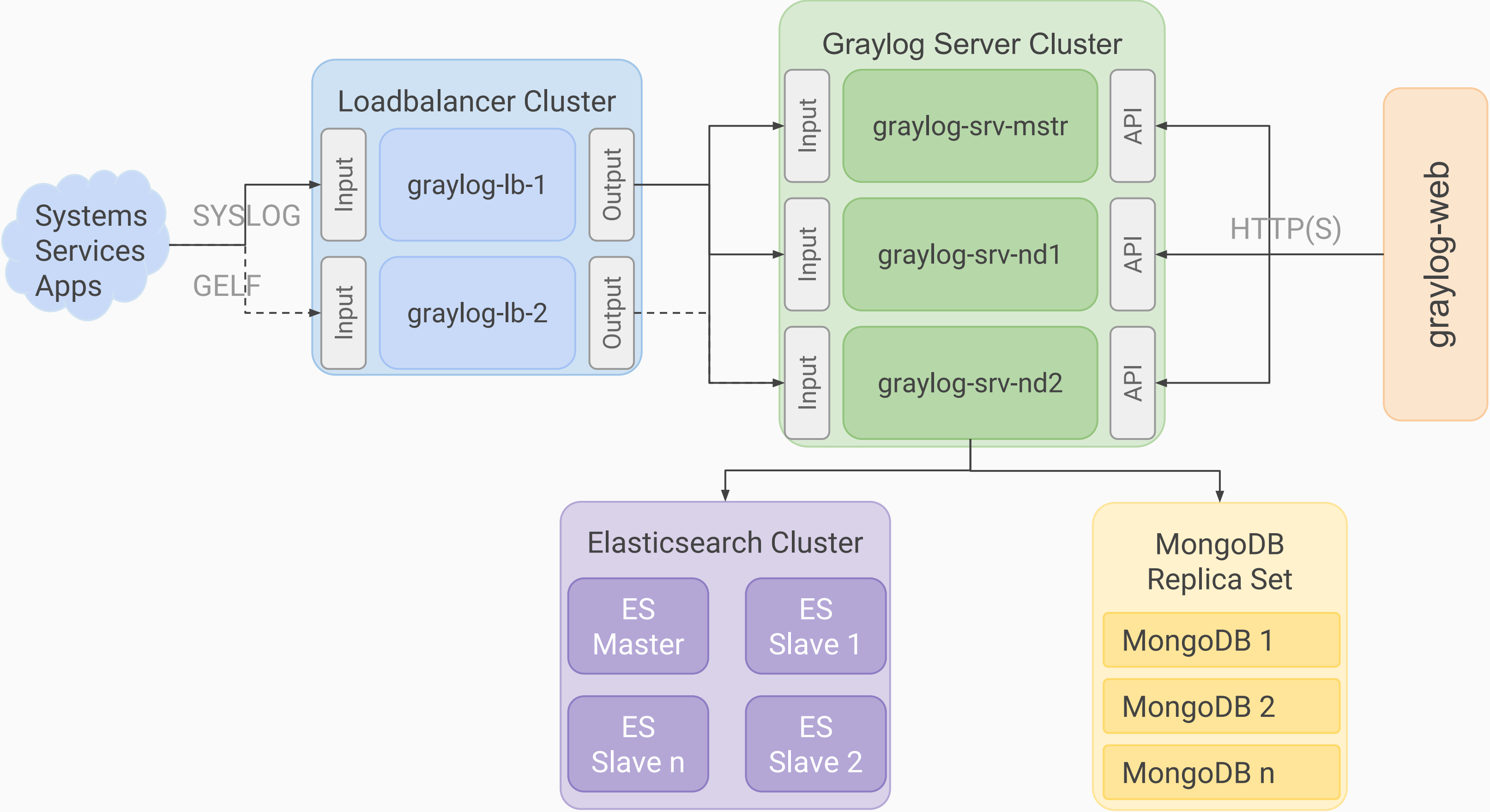
Graylog Concepts

● System	Config, Nodes, Indices, AuthN
● Inputs	Endpoints for receiving log data
● Indices	Store log data, controls log retention
● Streams	Rule based message routing & filtering
● Dashboards	Aggregated views on log data
● Alerts	Conditionally trigger & send notifications
● Outputs	Forward log data
● Pipelines	Stackable Pipes & Filters for log processing

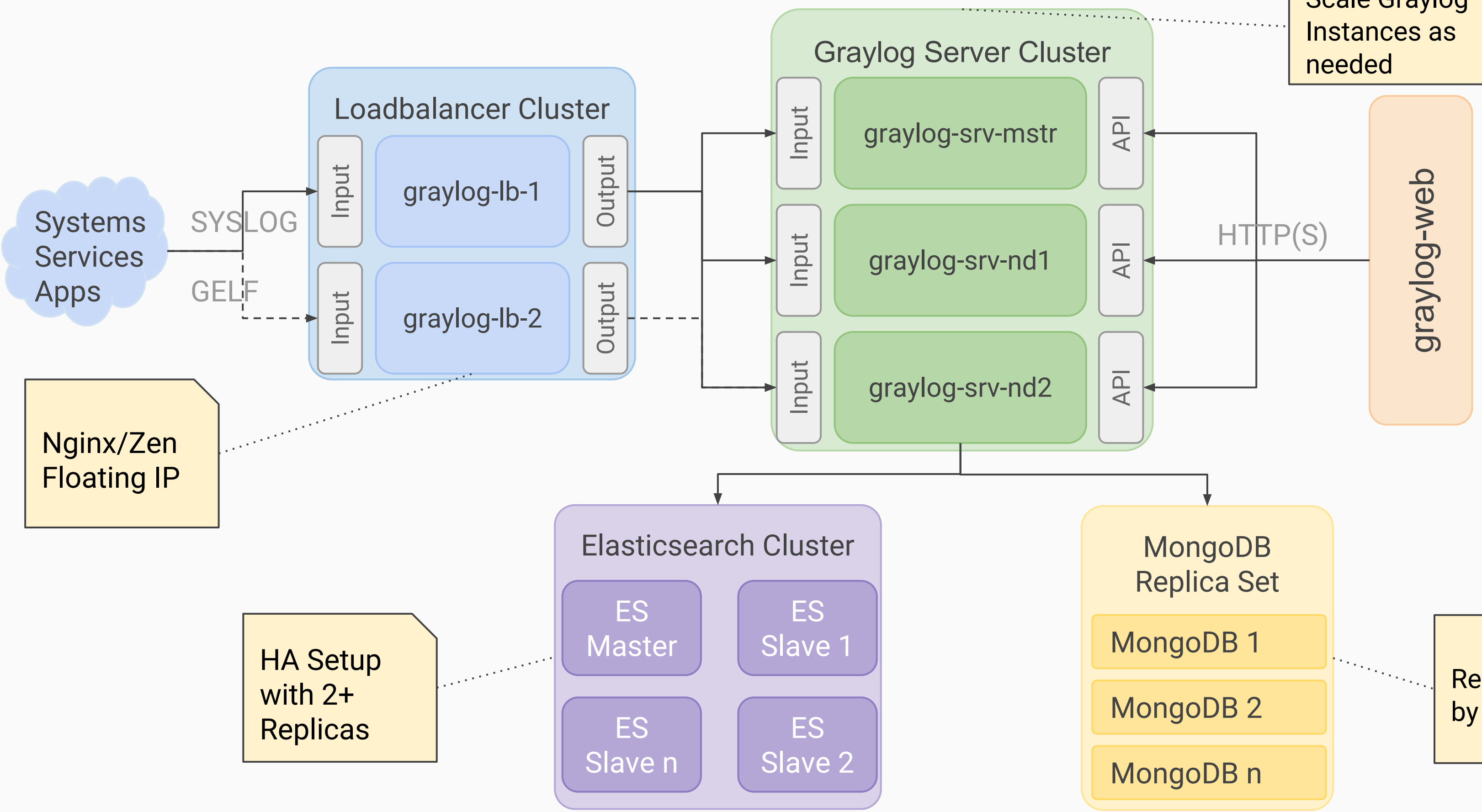
Graylog Component Overview



Graylog Canonical HA-Setup



Graylog Canonical HA-Setup



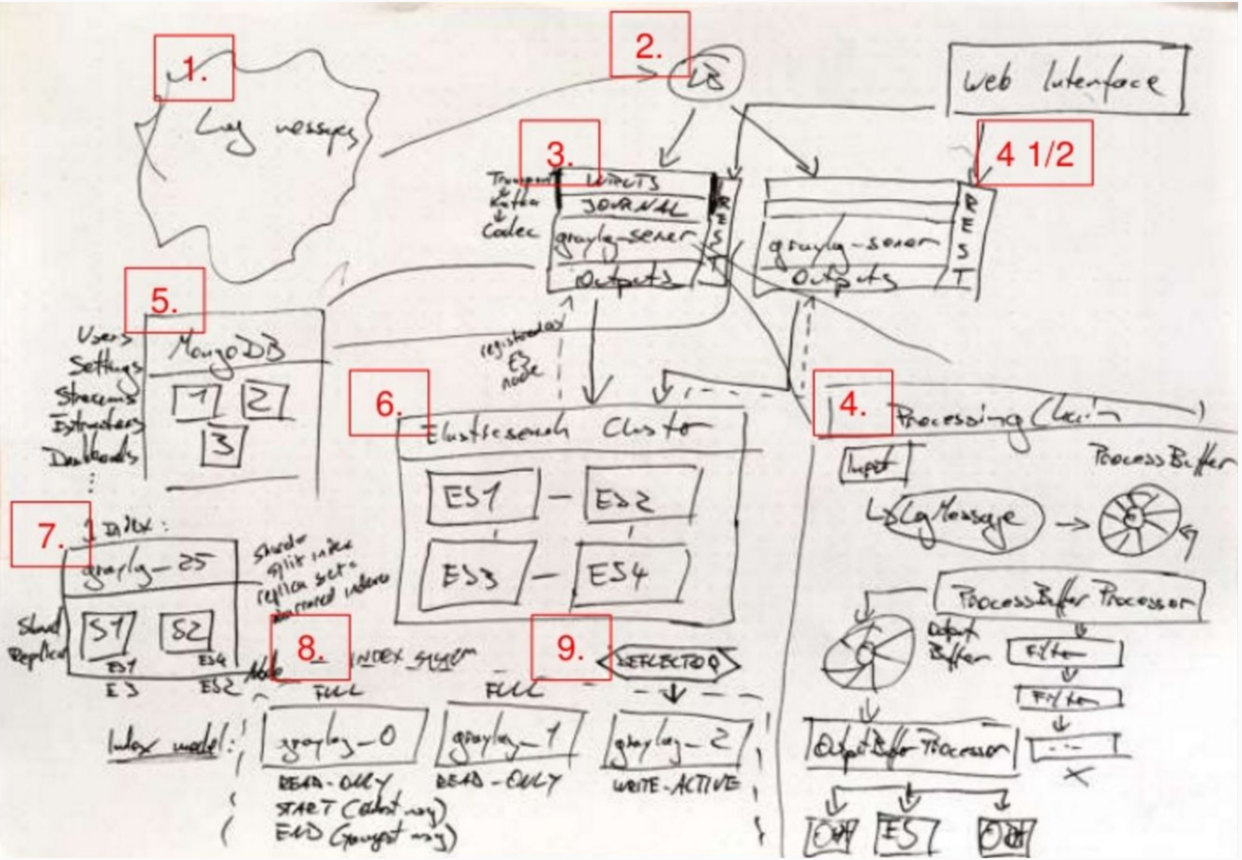
Scale Graylog Instances as needed

Nginx/Zen Floating IP

HA Setup with 2+ Replicas

Recommended by MongoDB

Graylog Architecture



1. Log Messages
2. Load Balancer
3. Transport Layer
4. Processing Chain

4.1/2 REST API

5. MongoDB ReplicaSet
6. Elasticsearch Cluster
7. Anatomy of an Index
8. Index Model
9. Deflector Queue

Graylog Engineering
Design your Architecture

Interesting Architecture Bits...

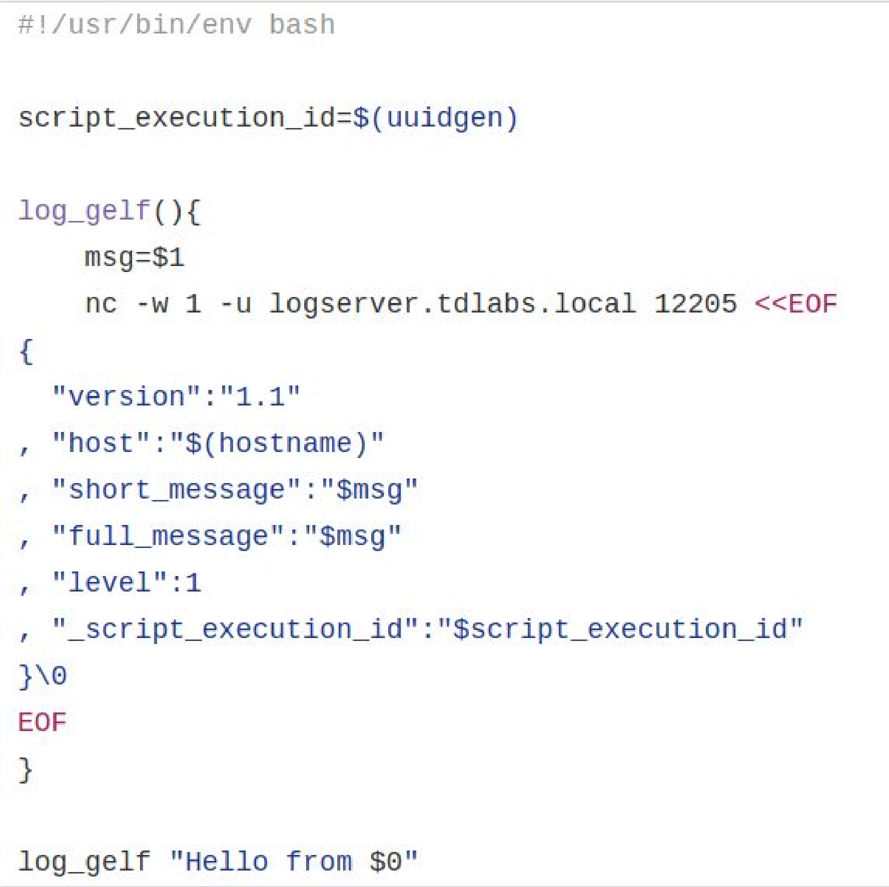
- Uses Apache Kafka for the append-only log journal on disk
  - Allows fast writes to disk
  - Avoids losing messages during spikes
- Uses LMAX Disruptor RingBuffer
  - Allows fast data ingestion and processing with low-latency
- Graylog Node acts as non-data Elasticsearch Node
  - Allows faster native protocols instead of HTTP/JSON
- Designed for Horizontal Scalability and HA
  - Graylog Nodes (2n+1 Processor nodes)
  - MongoDB (Shards + Replicas)
  - Elasticsearch (Shards + Replicas)
- Frontend build with React
- Custom Log Format GELF for more flexibility

Graylog Extended Log Format

- JSON String
- Avoids shortcomings of classic plain syslog
- Structured Log Message with Types
- Supports custom fields
- UDP and TCP
- Chunking
- Compression
- … GELF reference

```
{
  "version": "1.1",
  "host": "example.org",
  "short_message": "A short message",
  "full_message": "Backtrace here\n\nmore stuff",
  "timestamp": 1385053862.3072,
  "level": 1,
  "_user_id": 9001,
  "_some_info": "foo",
  "_some_env_var": "bar"
}
```

Demo Send GELF Message from a Shell Script



```
#!/usr/bin/env bash

script_execution_id=$(uuidgen)

log_gelf(){
    msg=$1
    nc -w 1 -u logserver.tdlabs.local 12205 <<EOF
{
  "version":"1.1"
, "host":"$(hostname)"
, "short_message":"$msg"
, "full_message":"$msg"
, "level":1
, "_script_execution_id":"$script_execution_id"
}\0
EOF
}

log_gelf "Hello from $0"
```

GELF with
netcat and heredoc
Gist

Demo use GELF logging with Docker

```
docker run -dit \
           --name nginx \
           -p 28080:80 \
             --log-driver=gelf \
             --log-opt gelf-address=udp://logserver.tdlabs.local:12205 \
             nginx:1.11.9-alpine
```

See: https://docs.docker.com/engine/admin/logging/overview

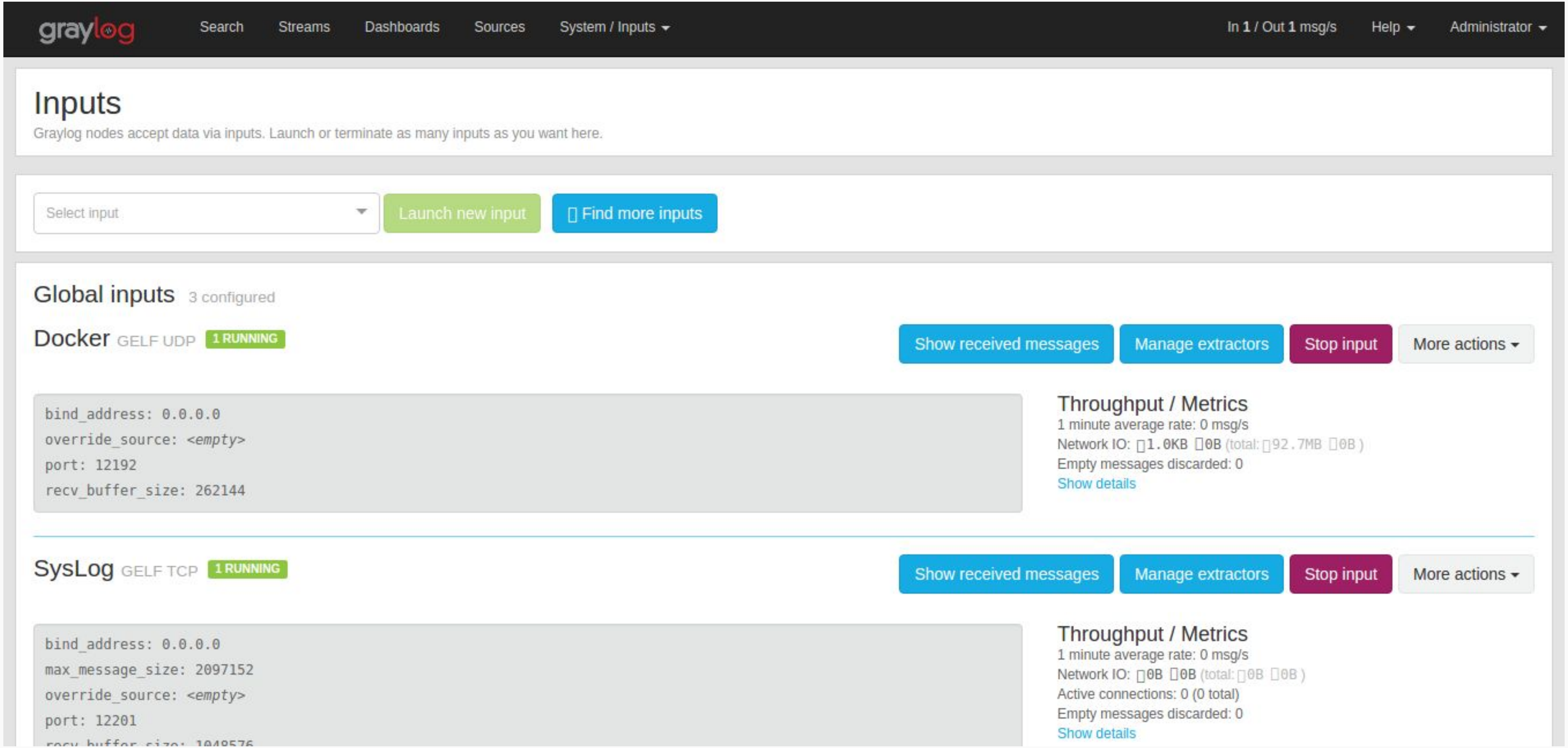
DEMO

Graylog in Action

Recap

- System
- Inputs
- Streams
- Searches
- Dashboards
- Alerts
- REST API Browser

Inputs



Inputs

Graylog nodes accept data via inputs. Launch or terminate as many inputs as you want here.

Global inputs 3 configured

Docker GELF UDP 1 RUNNING

```
bind_address: 0.0.0.0
override_source: <empty>
port: 12192
recv_buffer_size: 262144
```

Throughput / Metrics

1 minute average rate: 0 msg/s

Network IO: ▯1.0KB ▯0B (total: ▯92.7MB ▯0B)

Empty messages discarded: 0

Show details

SysLog GELF TCP 1 RUNNING

```
bind_address: 0.0.0.0
max_message_size: 2097152
override_source: <empty>
port: 12201
recv_buffer_size: 1048576
```

Throughput / Metrics

1 minute average rate: 0 msg/s

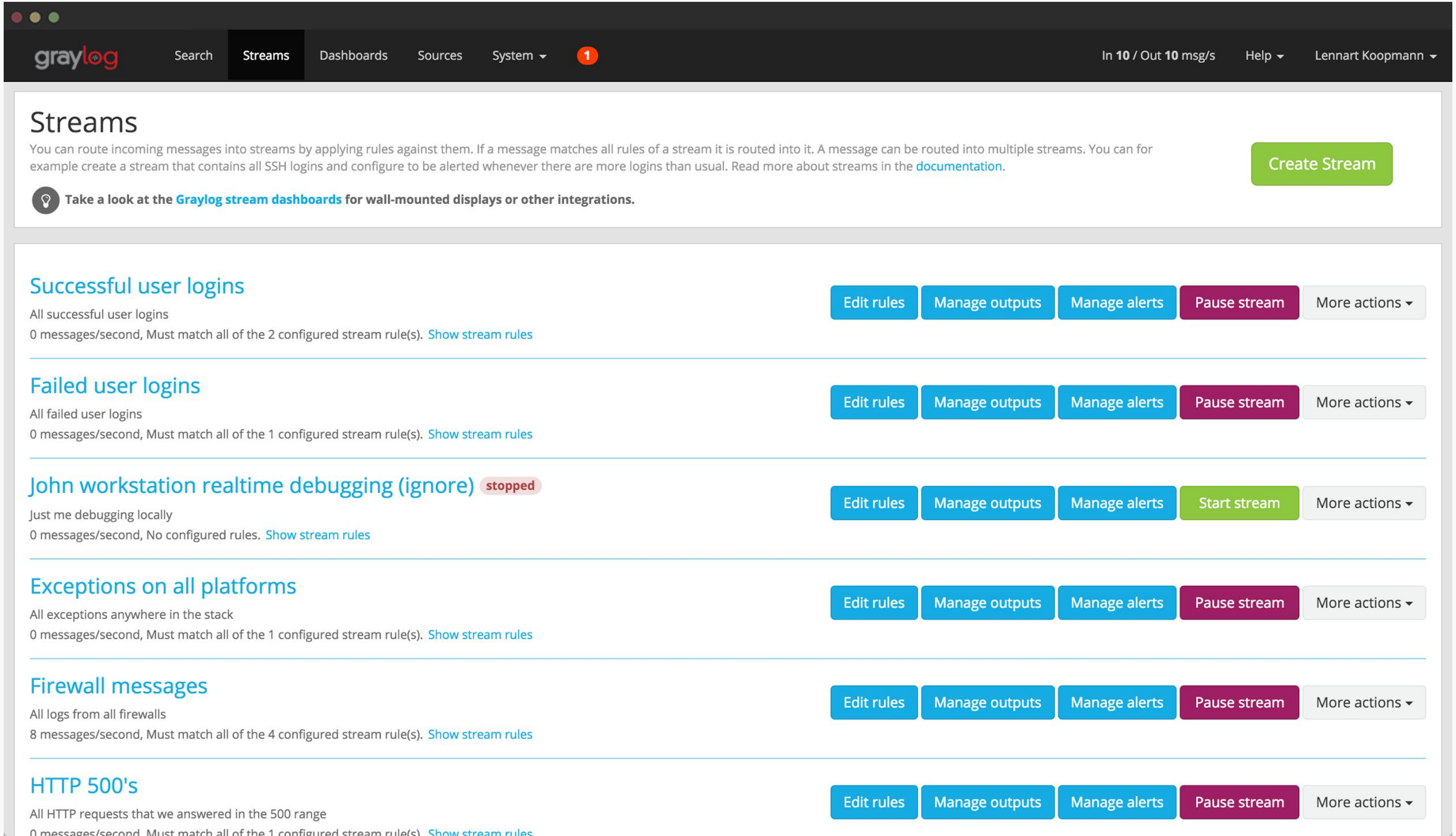
Network IO: ▯0B ▯0B (total: ▯0B ▯0B)

Active connections: 0 (0 total)

Empty messages discarded: 0

Show details

Streams



Streams

You can route incoming messages into streams by applying rules against them. If a message matches all rules of a stream it is routed into it. A message can be routed into multiple streams. You can for example create a stream that contains all SSH logins and configure to be alerted whenever there are more logins than usual. Read more about streams in the documentation.

Take a look at the Graylog stream dashboards for wall-mounted displays or other integrations.

Successful user logins

All successful user logins

0 messages/second, Must match all of the 2 configured stream rule(s). Show stream rules

Failed user logins

All failed user logins

0 messages/second, Must match all of the 1 configured stream rule(s). Show stream rules

John workstation realtime debugging (ignore) stopped

Just me debugging locally

0 messages/second, No configured rules. Show stream rules

Exceptions on all platforms

All exceptions anywhere in the stack

0 messages/second, Must match all of the 1 configured stream rule(s). Show stream rules

Firewall messages

All logs from all firewalls

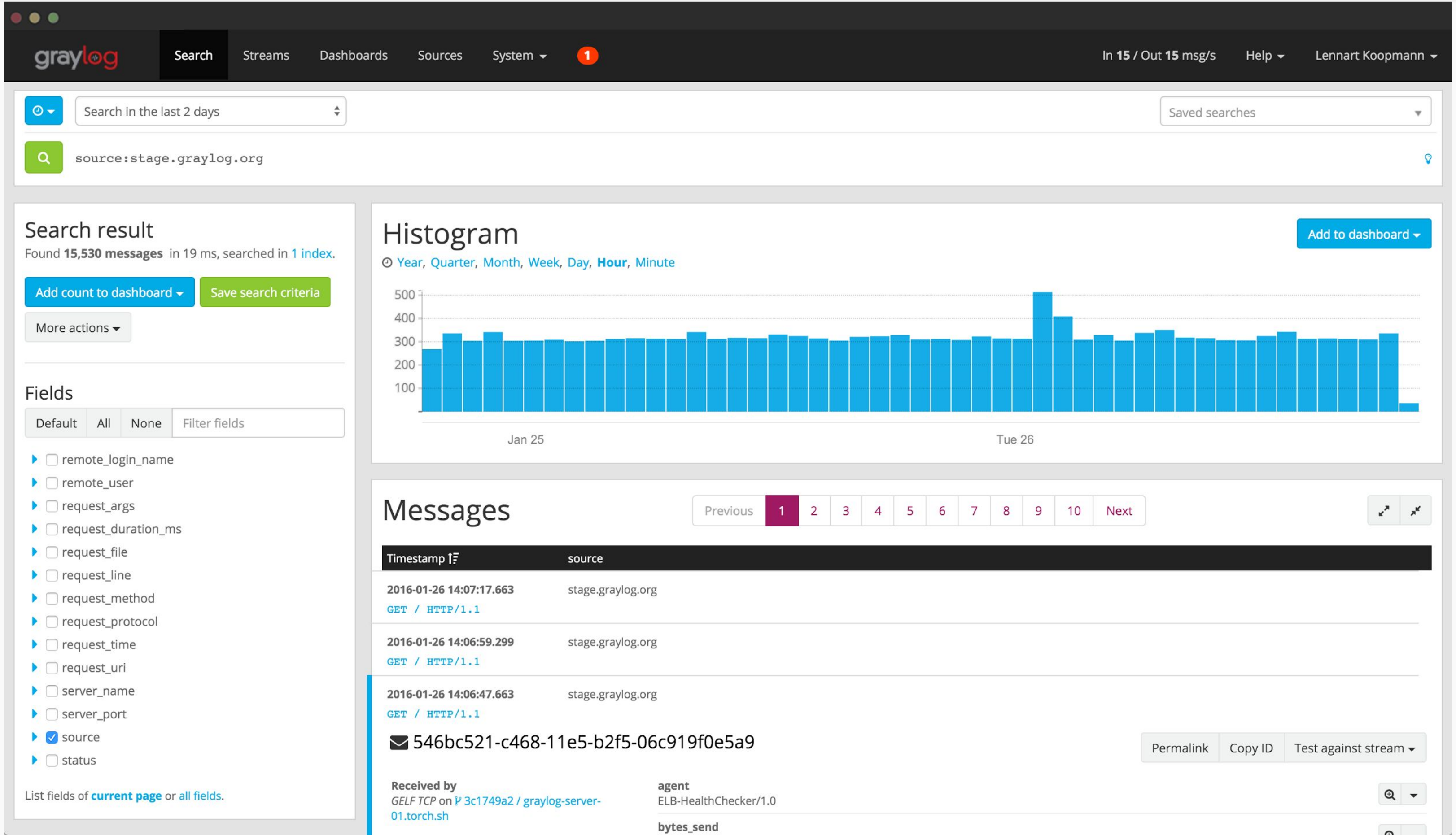
8 messages/second, Must match all of the 4 configured stream rule(s). Show stream rules

HTTP 500's

All HTTP requests that we answered in the 500 range

0 messages/second, Must match all of the 1 configured stream rule(s). Show stream rules

Log Message Search



Search in the last 2 days

```
source:stage.graylog.org
```

Search result

Found **15,530 messages** in 19 ms, searched in 1 index.

Fields

- ☐ remote_login_name
- ☐ remote_user
- ☐ request_args
- ☐ request_duration_ms
- ☐ request_file
- ☐ request_line
- ☐ request_method
- ☐ request_protocol
- ☐ request_time
- ☐ request_uri
- ☐ server_name
- ☐ server_port
- ☑ source
- ☐ status

List fields of **current page** or all fields.

Histogram

Messages

Timestamp	source
2016-01-26 14:07:17.663 GET / HTTP/1.1	stage.graylog.org
2016-01-26 14:06:59.299 GET / HTTP/1.1	stage.graylog.org
2016-01-26 14:06:47.663 GET / HTTP/1.1	stage.graylog.org

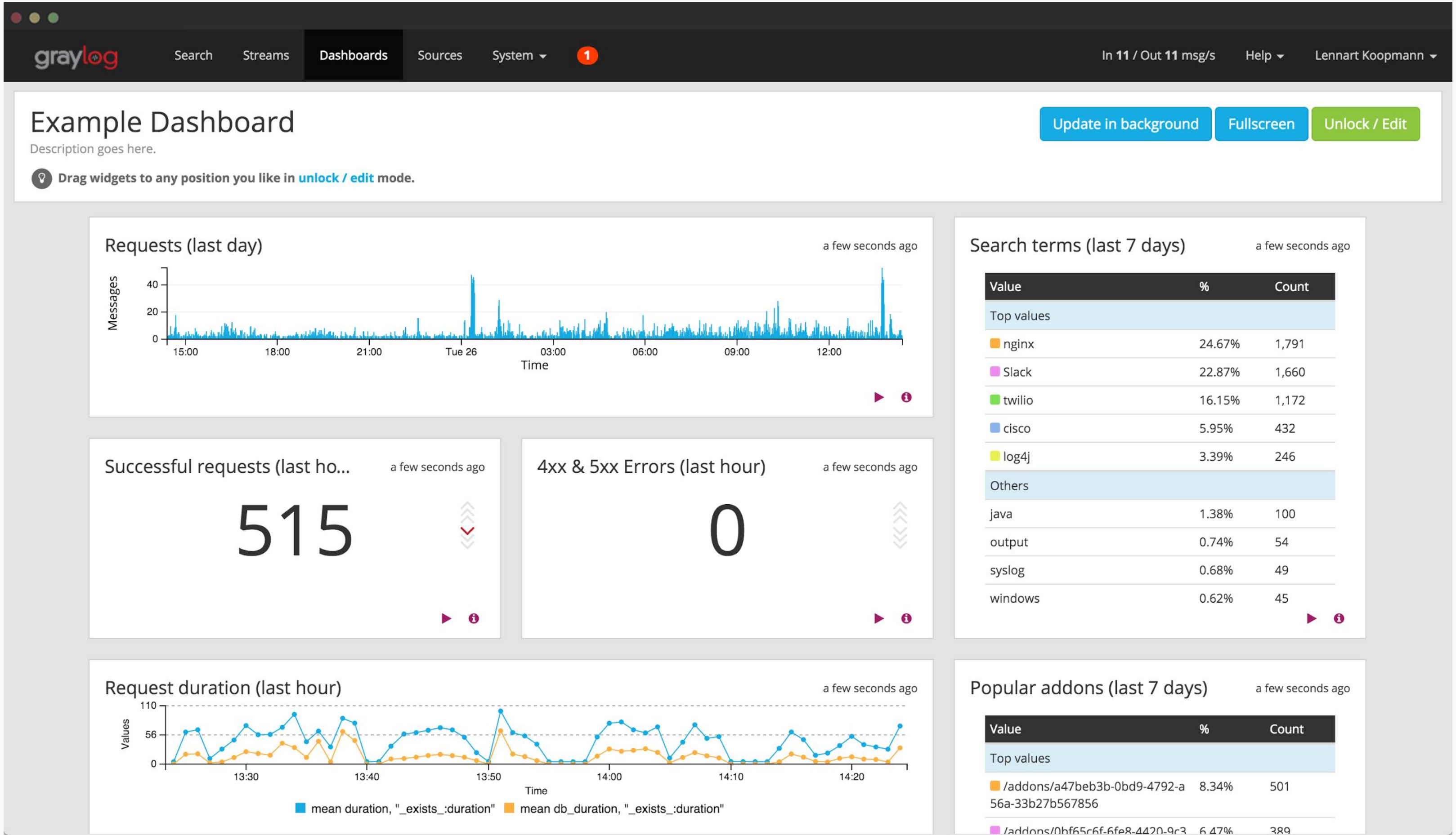
546bc521-c468-11e5-b2f5-06c919f0e5a9

Received by
GELF TCP on 3c1749a2 / graylog-server-01.torch.sh

agent
ELB-HealthChecker/1.0

bytes_send

Dashboards



Example Dashboard

Description goes here.

Drag widgets to any position you like in **unlock / edit** mode.

Requests (last day)

a few seconds ago

Successful requests (last ho...

a few seconds ago

515

4xx & 5xx Errors (last hour)

a few seconds ago

0

Search terms (last 7 days)

a few seconds ago

Value	%	Count
Top values		
🟧 nginx	24.67%	1,791
🟪 Slack	22.87%	1,660
🟩 twilio	16.15%	1,172
🟦 cisco	5.95%	432
🟨 log4j	3.39%	246
Others		
java	1.38%	100
output	0.74%	54
syslog	0.68%	49
windows	0.62%	45

Request duration (last hour)

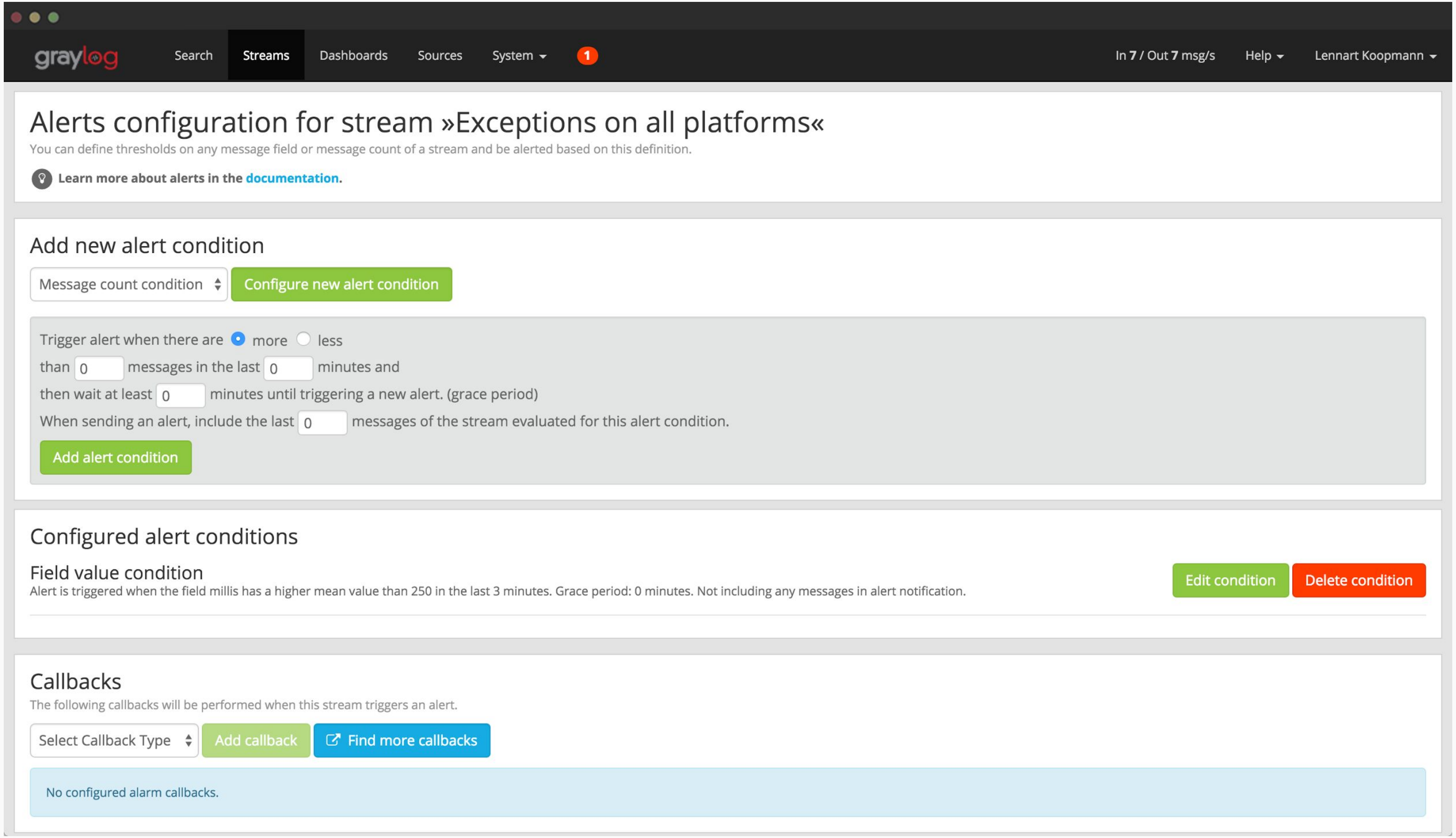
a few seconds ago

Popular addons (last 7 days)

a few seconds ago

Value	%	Count
Top values		
🟧 /addons/a47beb3b-0bd9-4792-a56a-33b27b567856	8.34%	501
🟪 /addons/0bf65c6f-6fe8-4420-9c3	6.47%	389

Alerts



Alerts configuration for stream »Exceptions on all platforms«

You can define thresholds on any message field or message count of a stream and be alerted based on this definition.

Learn more about alerts in the documentation.

Add new alert condition

Message count condition | Configure new alert condition

Trigger alert when there are ● more ○ less
than 0 messages in the last 0 minutes and
then wait at least 0 minutes until triggering a new alert. (grace period)
When sending an alert, include the last 0 messages of the stream evaluated for this alert condition.

Add alert condition

Configured alert conditions

Field value condition

Alert is triggered when the field millis has a higher mean value than 250 in the last 3 minutes. Grace period: 0 minutes. Not including any messages in alert notification.

Edit condition | Delete condition

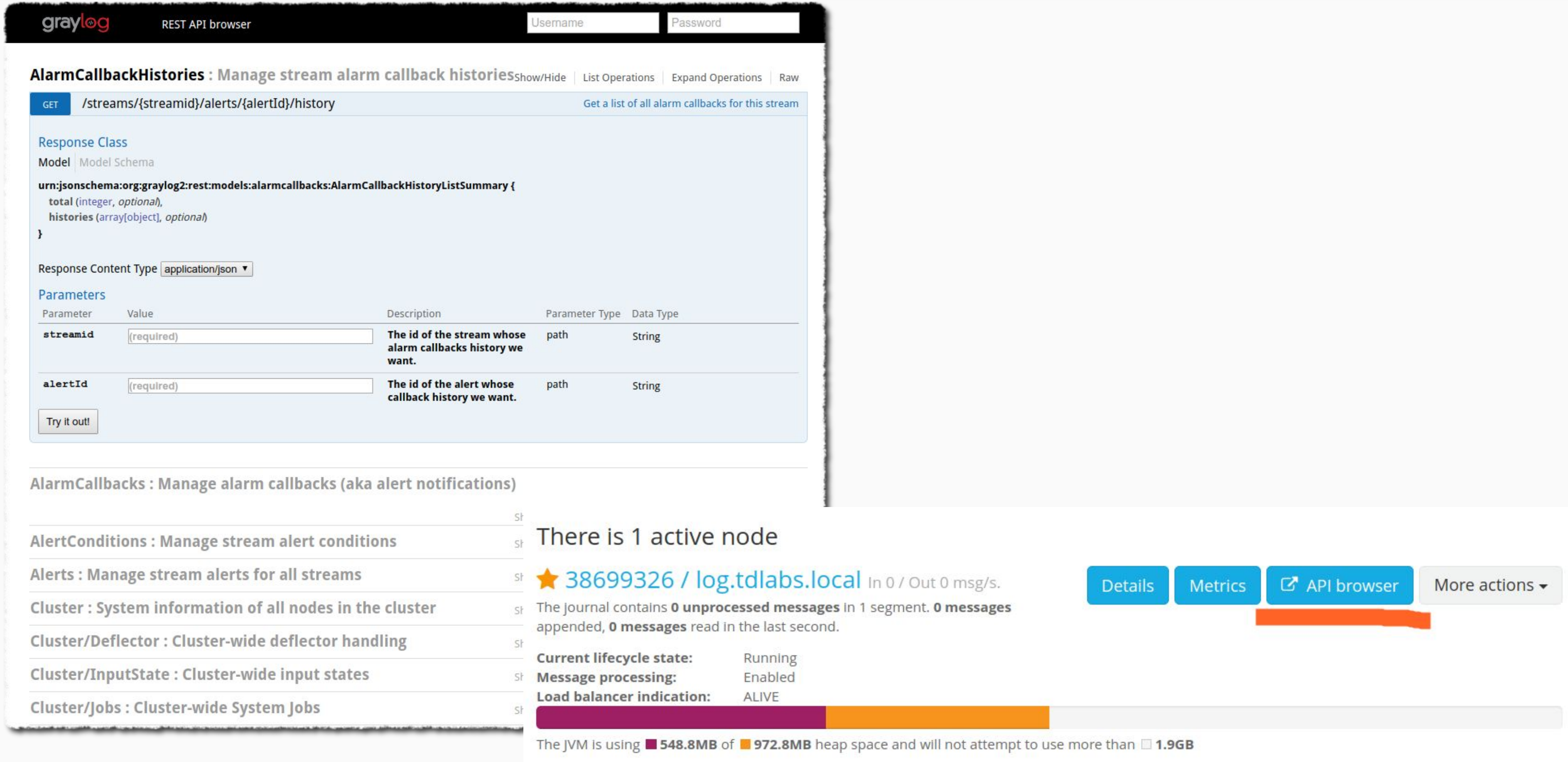
Callbacks

The following callbacks will be performed when this stream triggers an alert.

Select Callback Type | Add callback | Find more callbacks

No configured alarm callbacks.

API Browser



AlarmCallbackHistories : Manage stream alarm callback histories

GET /streams/{streamid}/alerts/{alertId}/history — Get a list of all alarm callbacks for this stream

Response Class

Model | Model Schema

```
urn:jsonschema:org:graylog2:rest:models:alarmcallbacks:AlarmCallbackHistoryListSummary {
  total (integer, optional),
  histories (array[object], optional)
}
```

Response Content Type: application/json ▾

Parameters

Parameter	Value	Description	Parameter Type	Data Type
streamid	(required)	The id of the stream whose alarm callbacks history we want.	path	String
alertId	(required)	The id of the alert whose callback history we want.	path	String

AlarmCallbacks : Manage alarm callbacks (aka alert notifications)

AlertConditions : Manage stream alert conditions

Alerts : Manage stream alerts for all streams

Cluster : System information of all nodes in the cluster

Cluster/Deflector : Cluster-wide deflector handling

Cluster/InputState : Cluster-wide input states

Cluster/Jobs : Cluster-wide System Jobs

There is 1 active node

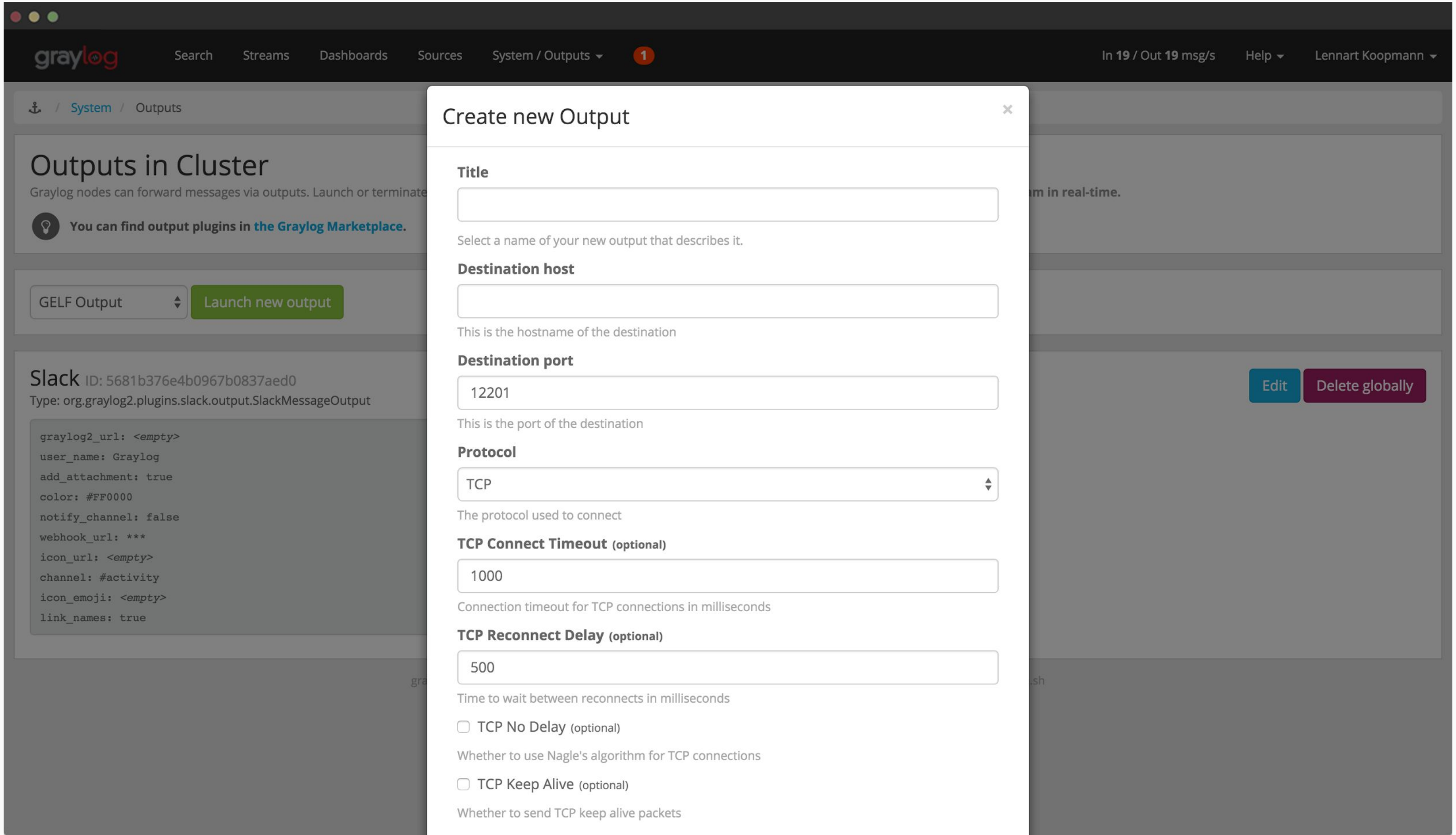
★ 38699326 / log.tdlabs.local In 0 / Out 0 msg/s.

The journal contains **0 unprocessed messages** in 1 segment. **0 messages** appended, **0 messages** read in the last second.

Current lifecycle state: Running
Message processing: Enabled
Load balancer indication: ALIVE

The JVM is using ■ **548.8MB** of ■ **972.8MB** heap space and will not attempt to use more than □ **1.9GB**

Outputs



Outputs in Cluster

Graylog nodes can forward messages via outputs. Launch or terminate … m in real-time.

You can find output plugins in **the Graylog Marketplace**.

GELF Output

Slack ID: 5681b376e4b0967b0837aed0

Type: org.graylog2.plugins.slack.output.SlackMessageOutput

```
graylog2_url: <empty>
user_name: Graylog
add_attachment: true
color: #FF0000
notify_channel: false
webhook_url: ***
icon_url: <empty>
channel: #activity
icon_emoji: <empty>
link_names: true
```

Create new Output

Title

Select a name of your new output that describes it.

Destination host

This is the hostname of the destination

Destination port

12201

This is the port of the destination

Protocol

TCP

The protocol used to connect

TCP Connect Timeout (optional)

1000

Connection timeout for TCP connections in milliseconds

TCP Reconnect Delay (optional)

500

Time to wait between reconnects in milliseconds

□ TCP No Delay (optional)

Whether to use Nagle's algorithm for TCP connections

□ TCP Keep Alive (optional)

Whether to send TCP keep alive packets

Integrations

- Java
  - logstash-gelf Library
    - Support for multiple Logging Frameworks
    - Website http://logging.paluch.biz/
    - Github https://github.com/mp911de/logstash-gelf
    - Examples mp911de/logstash-gelf src/test/java/biz/paluch/logging/gelf
- .Net
  - gelf4net https://github.com/jjchiw/gelf4net
- Go
  - go-gelf https://github.com/Graylog2/go-gelf
- Windows
  - winlogbeat, Graylog Collector Sidecar
  - nxlog https://nxlog.co/products/nxlog-community-edition
- Linux
  - filebeat, Graylog Collector Sidecar
  - nxlog, syslog

DEMO

GELF & Java

Logback & GELF example configuration

```
<appender name="GELF" class="biz.paluch.logging.gelf.logback.GelfLogbackAppender">
    <host>${LOG_PROTO:-udp}:${LOG_HOSTNAME:-localhost}</host>
    <port>${LOG_PORT:-12201}</port>
    <version>1.1</version>
    <timestampPattern>yyyy-MM-dd HH:mm:ss,SSSS</timestampPattern>
    <maximumMessageSize>8192</maximumMessageSize>
    <facility>-</facility>
    <extractStackTrace>true</extractStackTrace>
    <filterStackTrace>true</filterStackTrace>
    <mdcProfiling>false</mdcProfiling>
    <additionalFields>org=tdlabs,ctx=demo,svc=hello-world-svc,env=test</additionalFields>
    <additionalFieldTypes>org=String,ctx=String,svc=String,env=String</additionalFieldTypes>
    <mdcFields>APP_STAGE</mdcFields>
    <dynamicMdcFields>svc_.*</dynamicMdcFields>
    <filter class="ch.qos.logback.classic.filter.ThresholdFilter">
        <level>${LOG_LEVEL_GELF:-INFO}</level>
    </filter>
</appender>

<root level="INFO">
    <appender-ref ref="GELF"/>
    <appender-ref ref="CONSOLE"/>
</root>
```

Log-Server Destination

StackTrace handling

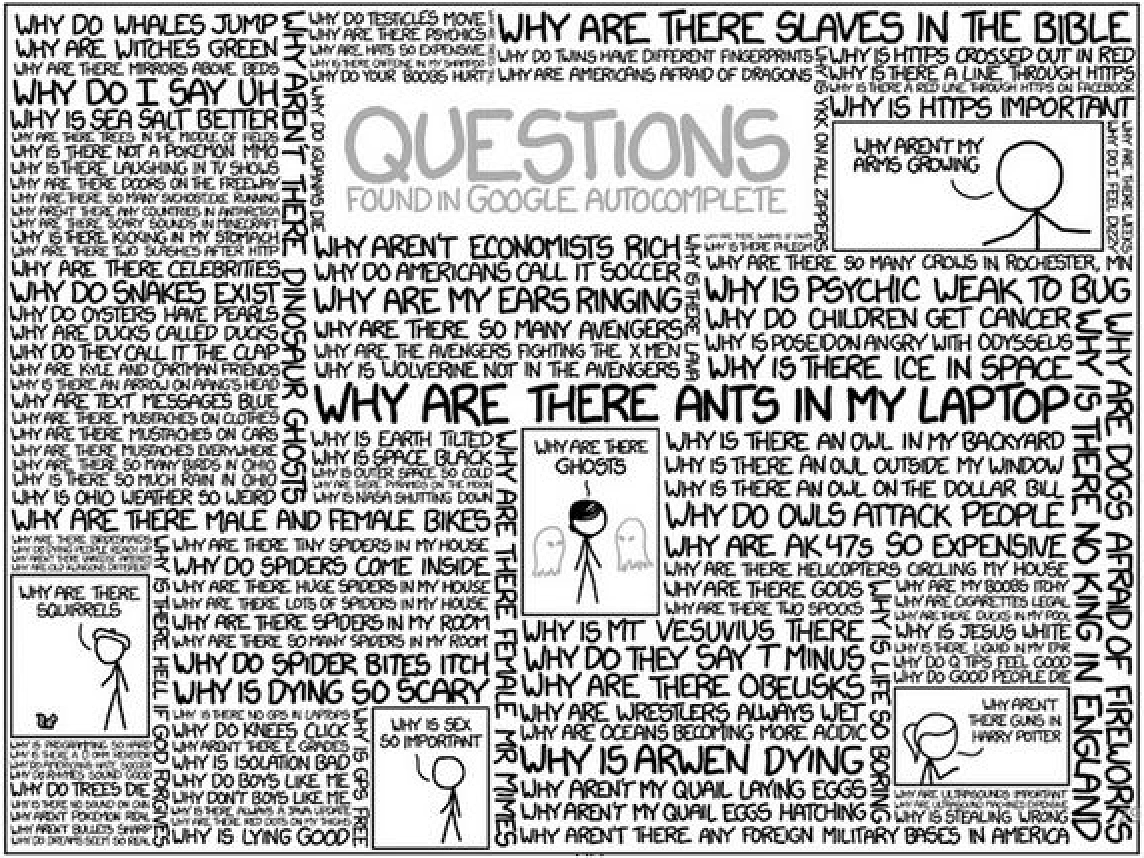
Thread-Local **M**apped **D**iagnostic **C**ontext fields

`logback.xml`

Further reading

- Graylog 2.2 Design Documents
- Blog Post Monitoring Graylog
- Blog Post Processing 250GB Log Data / Day
- German Article in IT-Administrator 2015/09
- German Article in IT-Administrator 2015/10
- German Article in IT-Administrator 2015/11
- Youtube Windows Event log with Graylog

Questions


QUESTIONS
FOUND IN GOOGLE AUTOCOMPLETE
WHY DO WHALES JUMP
WHY ARE WITCHES GREEN
WHY ARE THERE MIRRORS ABOVE BEDS
WHY DO I SAY UH
WHY IS SEA SALT BETTER
WHY ARE THERE TREES IN THE MIDDLE OF FIELDS
WHY IS THERE NOT A POKEMON MMO
WHY IS THERE LAUGHING IN TV SHOWS
WHY ARE THERE DOORS ON THE FREEWAY
WHY ARE THERE SO MANY SVCHOST.EXE RUNNING
WHY AREN'T THERE ANY COUNTRIES IN ANTARCTICA
WHY ARE THERE SCARY SOUNDS IN MINECRAFT
WHY IS THERE KICKING IN MY STOMACH
WHY ARE THERE TWO SLASHES AFTER HTTP
WHY ARE THERE CELEBRITIES
WHY DO SNAKES EXIST
WHY DO OYSTERS HAVE PEARLS
WHY ARE DUCKS CALLED DUCKS
WHY DO THEY CALL IT THE CLAP
WHY ARE KYLE AND CARTMAN FRIENDS
WHY IS THERE AN ARROW ON AANG'S HEAD
WHY ARE TEXT MESSAGES BLUE
WHY ARE THERE MUSTACHES ON CLOTHES
WHY ARE THERE MUSTACHES ON CARS
WHY ARE THERE MUSTACHES EVERYWHERE
WHY ARE THERE SO MANY BIRDS IN OHIO
WHY IS THERE SO MUCH RAIN IN OHIO
WHY IS OHIO WEATHER SO WEIRD
WHY AREN'T THERE DINOSAUR GHOSTS
WHY DO IGUANAS DIE
WHY DO TESTICLES MOVE
WHY ARE THERE PSYCHICS
WHY ARE HATS SO EXPENSIVE
WHY IS THERE CAFFEINE IN MY SHAMPOO
WHY DO YOUR BOOBS HURT
WHY ARE THERE SLAVES IN THE BIBLE
WHY DO TWINS HAVE DIFFERENT FINGERPRINTS
WHY ARE AMERICANS AFRAID OF DRAGONS
WHY IS HTTPS CROSSED OUT IN RED
WHY IS THERE A LINE THROUGH HTTPS
WHY IS THERE A RED LINE THROUGH HTTPS ON FACEBOOK
WHY IS HTTPS IMPORTANT
WHY YKK ON ALL ZIPPERS
WHY AREN'T MY ARMS GROWING
WHY ARE THERE WEEKS
WHY DO I FEEL DIZZY
WHY AREN'T ECONOMISTS RICH
WHY DO AMERICANS CALL IT SOCCER
WHY ARE MY EARS RINGING
WHY ARE THERE SO MANY AVENGERS
WHY ARE THE AVENGERS FIGHTING THE X MEN
WHY IS WOLVERINE NOT IN THE AVENGERS
WHY IS THERE LAVA
WHY ARE THERE SWARMS OF GNATS
WHY IS THERE PHLEGM
WHY ARE THERE SO MANY CROWS IN ROCHESTER, MN
WHY IS PSYCHIC WEAK TO BUG
WHY DO CHILDREN GET CANCER
WHY IS POSEIDON ANGRY WITH ODYSSEUS
WHY IS THERE ICE IN SPACE
WHY ARE THERE ANTS IN MY LAPTOP
WHY IS EARTH TILTED
WHY IS SPACE BLACK
WHY IS OUTER SPACE SO COLD
WHY ARE THERE PYRAMIDS ON THE MOON
WHY IS NASA SHUTTING DOWN
WHY ARE THERE GHOSTS
WHY IS THERE AN OWL IN MY BACKYARD
WHY IS THERE AN OWL OUTSIDE MY WINDOW
WHY IS THERE AN OWL ON THE DOLLAR BILL
WHY DO OWLS ATTACK PEOPLE
WHY ARE AK 47s SO EXPENSIVE
WHY ARE THERE HELICOPTERS CIRCLING MY HOUSE
WHY ARE THERE GODS
WHY ARE THERE TWO SPOCKS
WHY IS THERE NO KING IN ENGLAND
WHY ARE DOGS AFRAID OF FIREWORKS
WHY ARE THERE MALE AND FEMALE BIKES
WHY ARE THERE TINY SPIDERS IN MY HOUSE
WHY DO SPIDERS COME INSIDE
WHY ARE THERE HUGE SPIDERS IN MY HOUSE
WHY ARE THERE LOTS OF SPIDERS IN MY HOUSE
WHY ARE THERE SPIDERS IN MY ROOM
WHY ARE THERE SO MANY SPIDERS IN MY ROOM
WHY DO SPIDER BITES ITCH
WHY IS DYING SO SCARY
WHY IS THERE HELL IF GOD FORGIVES
WHY ARE THERE SQUIRRELS
WHY IS THERE NO GPS IN LAPTOPS
WHY DO KNEES CLICK
WHY AREN'T THERE E GRADES
WHY IS ISOLATION BAD
WHY DO BOYS LIKE ME
WHY DON'T BOYS LIKE ME
WHY IS THERE ALWAYS A JAVA UPDATE
WHY ARE THERE RED DOTS ON MY THIGHS
WHY IS LYING GOOD
WHY IS GPS FREE
WHY IS SEX SO IMPORTANT
WHY ARE THERE FEMALE MR MIMES
WHY IS PROGRAMMING SO HARD
WHY IS THERE A 0 OHM RESISTOR
WHY DO AMERICANS HATE SOCCER
WHY DO RHYMES SOUND GOOD
WHY DO TREES DIE
WHY IS THERE NO SOUND ON CNN
WHY AREN'T POKEMON REAL
WHY AREN'T BULLETS SHARP
WHY DO DREAMS SEEM SO REAL
WHY IS MT VESUVIUS THERE
WHY DO THEY SAY T MINUS
WHY ARE THERE OBELISKS
WHY ARE WRESTLERS ALWAYS WET
WHY ARE OCEANS BECOMING MORE ACIDIC
WHY IS ARWEN DYING
WHY AREN'T MY QUAIL LAYING EGGS
WHY AREN'T MY QUAIL EGGS HATCHING
WHY AREN'T THERE ANY FOREIGN MILITARY BASES IN AMERICA
WHY IS LIFE SO BORING
WHY ARE MY BOOBS ITCHY
WHY ARE CIGARETTES LEGAL
WHY ARE THERE DUCKS IN MY POOL
WHY IS JESUS WHITE
WHY IS THERE LIQUID IN MY EAR
WHY DO Q TIPS FEEL GOOD
WHY DO GOOD PEOPLE DIE
WHY AREN'T THERE GUNS IN HARRY POTTER
WHY ARE ULTRASOUNDS IMPORTANT
WHY ARE ULTRASOUND MACHINES EXPENSIVE
WHY IS STEALING WRONG

Best Practices: Log Message Enrichment

- System Context
  - Where did the log message originate?
  - → Associate context information with the log Message
- Request Context
  - Who processed the message?
  - Follow the request processing through multiple layers (Request Id)
  - ... or even accros multiple nodes (Trace Id) → http://zipkin.io
- Audit Information
  - Which user did produce the log message?
  - Beware of privacy law!
- First Failure Data Capture
  - Create a unique id for each particular error instance
  - → Makes it easier to refer to the error

Best Practices: Log Message Enrichment

System Context

`source`	*Host*	dborac1a.db.internal.acme.com
`org`	*Organization / Tenant*	acme, customer1, tdlabs
`ctx`	*Context / System Boundary*	idm, net, accounting, clearing
`env`	*Environment*	dev, local, test, qa, prod
`svc`	*Logical Service name*	sso, booking, sla-monitoring
`inst`	*Service Instance*	1, 1a, 2b

Best Practices: Log Message Enrichment

Request Context

● **rid**	*Request Id*	6caae423-64f8-326d
● tid	*Trace Id*	12321-23231-2133-23

Audit Information

● **usr_id**	*Global/Tenant User Id*	c8609423-66d8-485d
● usr_name	*Tenant User Name*	ameier

First Failure Data Capture

● **err_id**	*UUID per Error*	aa2a-4e10609b95a1
● **err_code**	*Logical Error Code*	BILLING_ERROR_BANK_IFACE_UNAVIL

Thomas Darimont
Software Architect eurodata AG

Java User Group
Saarland

t.darimont@eurodata.de
@thomasdarimont