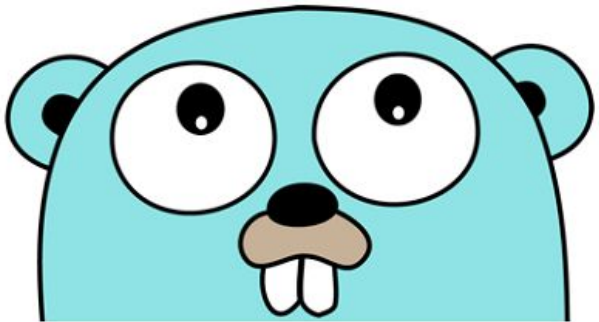
Система сбора логов с кластера kubernetes

Антон Галицын

Tinkoff.ru

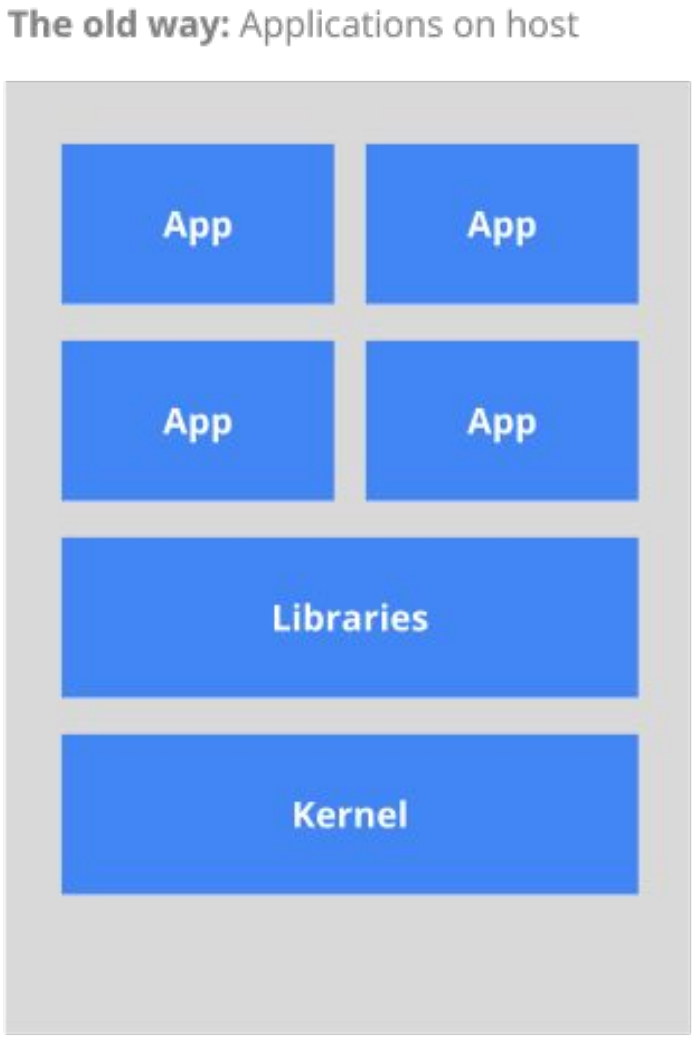
О чем доклад?

1. Кратко о Kubernetes
2. Задача сбора логов
3. Сбор логов у docker контейнеров
4. Что сделали мы?
5. В конце про Go

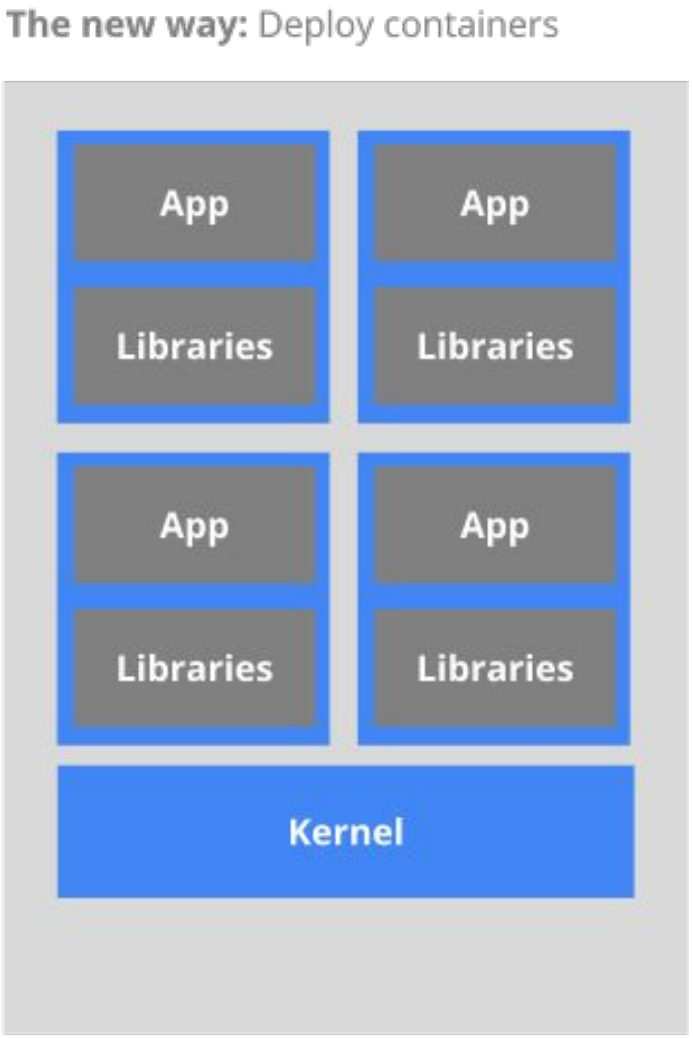


Kubernetes ;TLDR

Контейнеры



Heavyweight, non-portable
Relies on OS package manager



Small and fast, portable
Uses OS-level virtualization

Контейнеры

1. Изоляция - Linux namespaces
   сеть, юзеры, пространство процессов, диск и тд

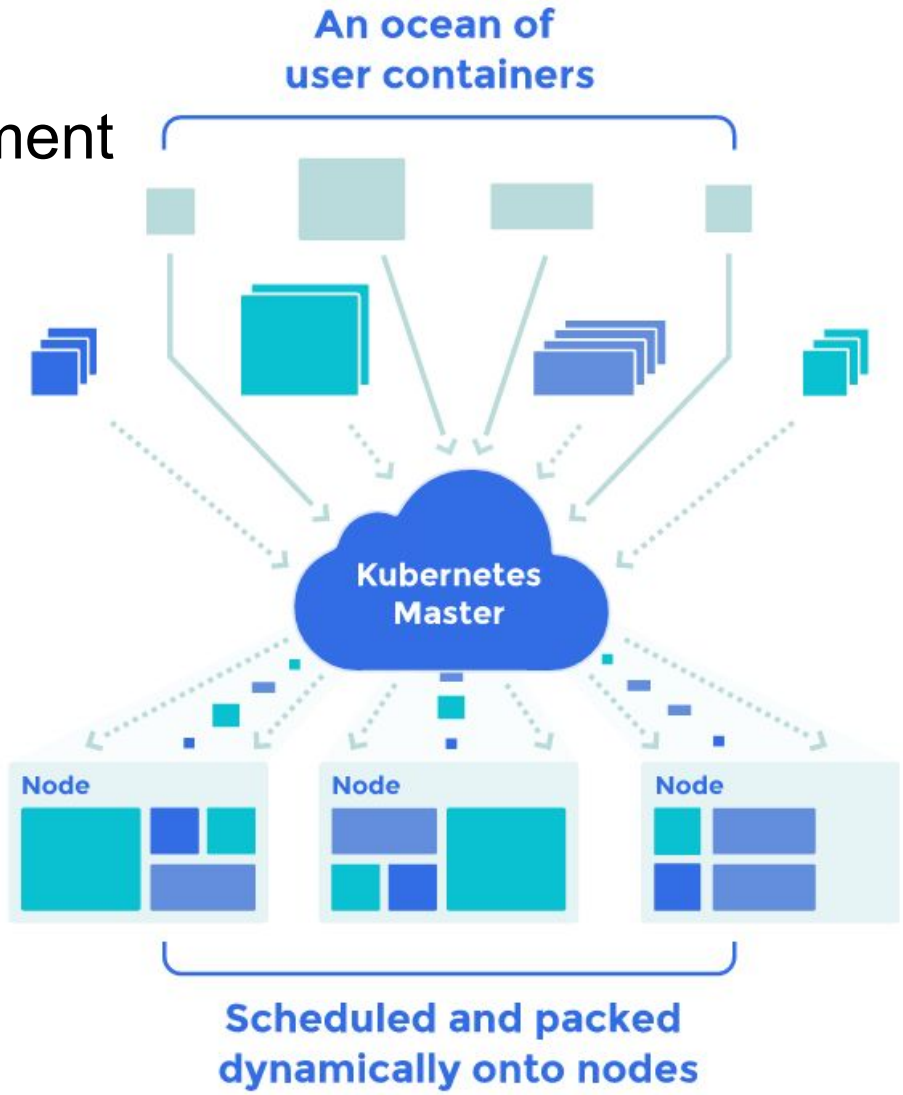
2. Ограничение действий - Linux capabilities,
   apparmor / selinux, seccomp

3. Ограничение ресурсов - Linux cgroups
   Память, CPU и тд.

Kubernetes

Container cluster management

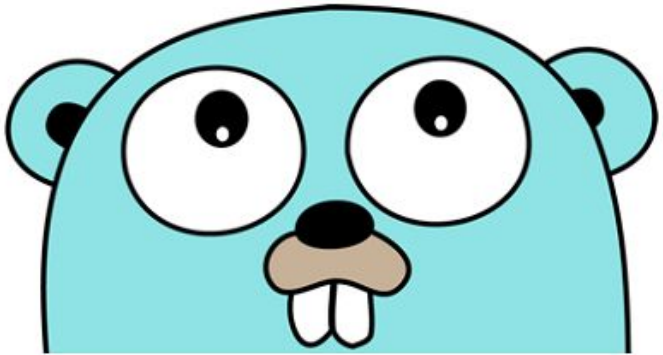
- Orchestration
- Deployment
- Scaling
- Load balancing
- Logging and monitoring



Система сбора логов

От кого можно собрать логи?

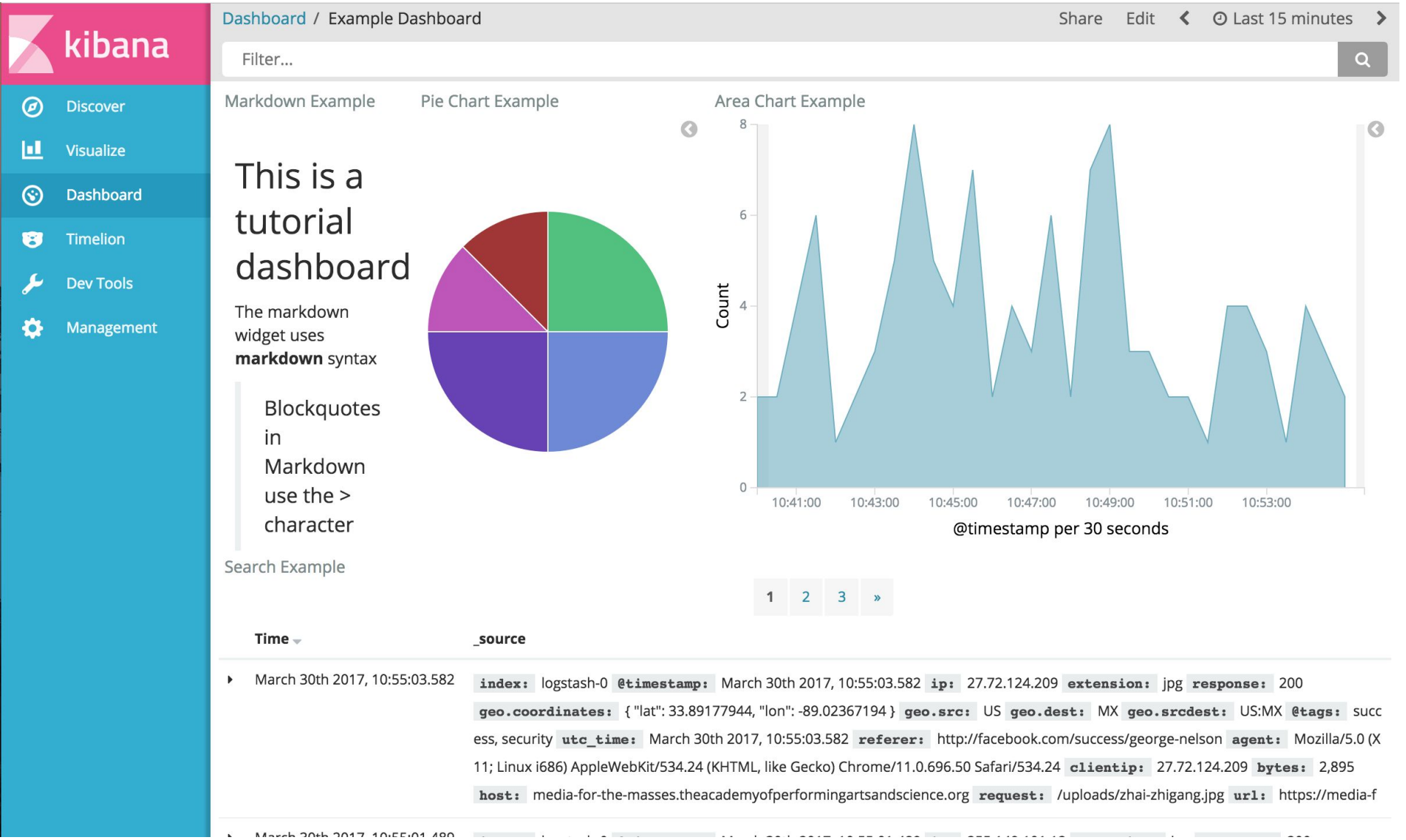
- Балансировщики
- Базы данных
- Приложения
- Kernel
- и тд



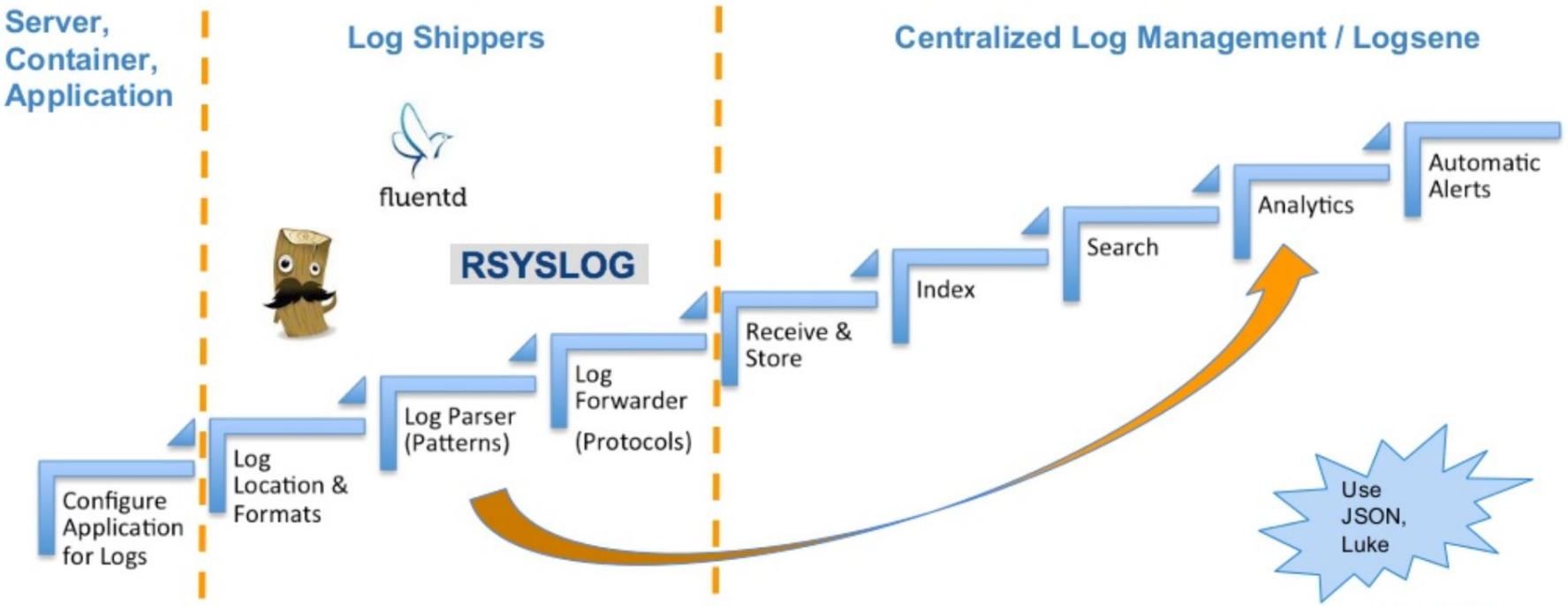
Centralized log management

- Поиск
- Аналитика
- Дашборды с агрегатами
- Алерты

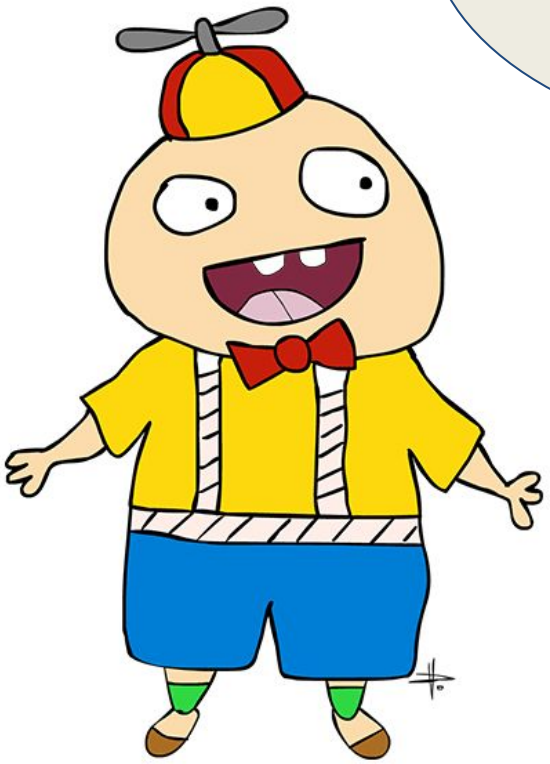
Пример


kibana
Discover
Visualize
Dashboard
Timelion
Dev Tools
Management
Dashboard / Example Dashboard
Share
Edit
Last 15 minutes
Filter...
Markdown Example
Pie Chart Example
Area Chart Example
This is a tutorial dashboard
The markdown widget uses **markdown** syntax
Blockquotes in Markdown use the > character
Count
@timestamp per 30 seconds
Search Example
Time
_source
March 30th 2017, 10:55:03.582
index: logstash-0 @timestamp: March 30th 2017, 10:55:03.582 ip: 27.72.124.209 extension: jpg response: 200 geo.coordinates: { "lat": 33.89177944, "lon": -89.02367194 } geo.src: US geo.dest: MX geo.srcdest: US:MX @tags: success, security utc_time: March 30th 2017, 10:55:03.582 referer: http://facebook.com/success/george-nelson agent: Mozilla/5.0 (X11; Linux i686) AppleWebKit/534.24 (KHTML, like Gecko) Chrome/11.0.696.50 Safari/534.24 clientip: 27.72.124.209 bytes: 2,895 host: media-for-the-masses.theacademyofperformingartsandscience.org request: /uploads/zhai-zhigang.jpg url: https://media-f

Шаги доставки

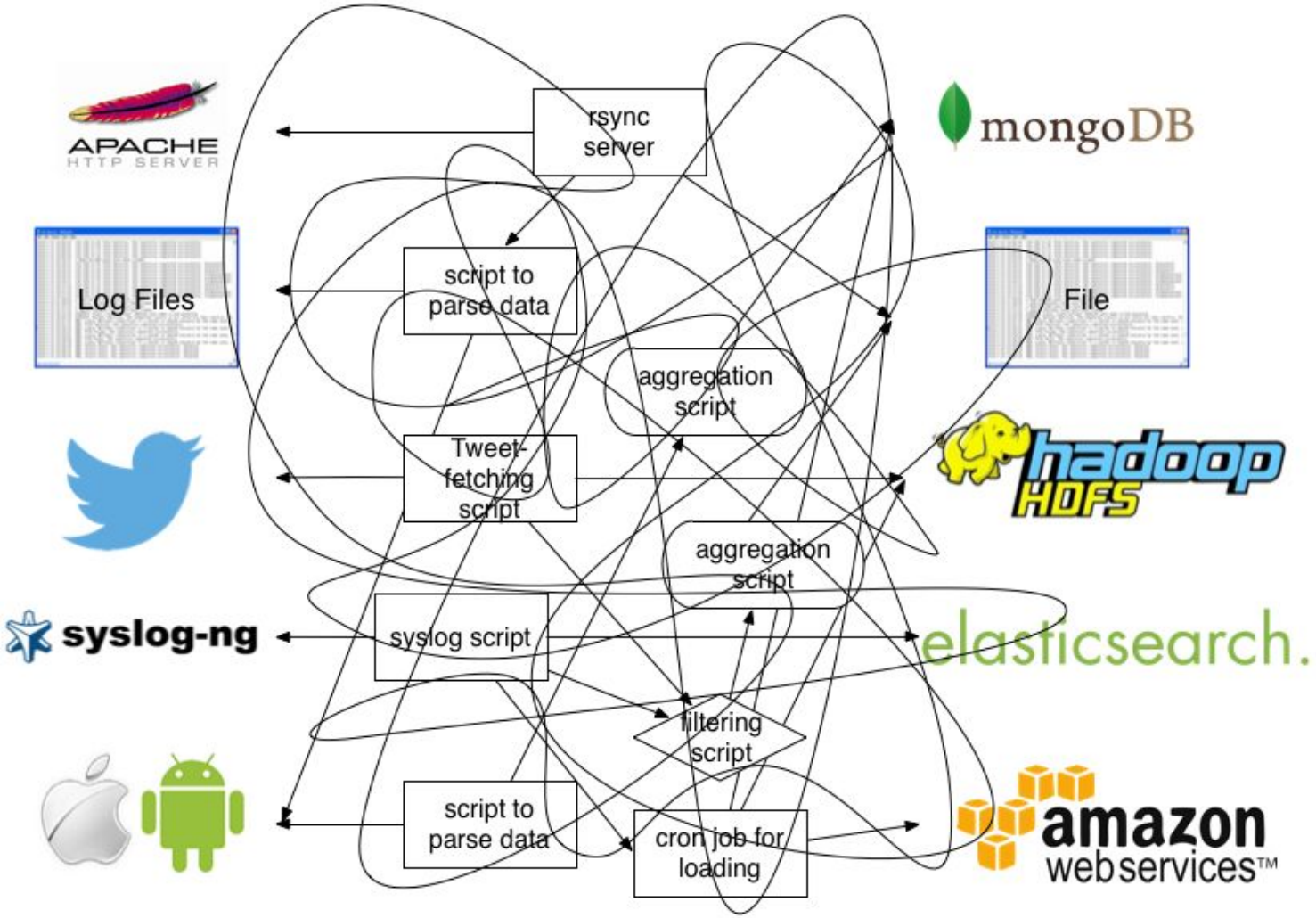


Сейчас запилю



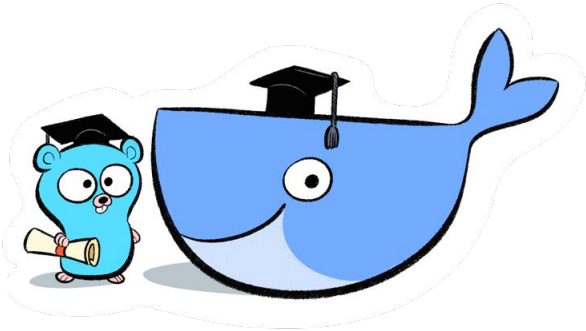
Пару скриптиков и норм!

Ой...

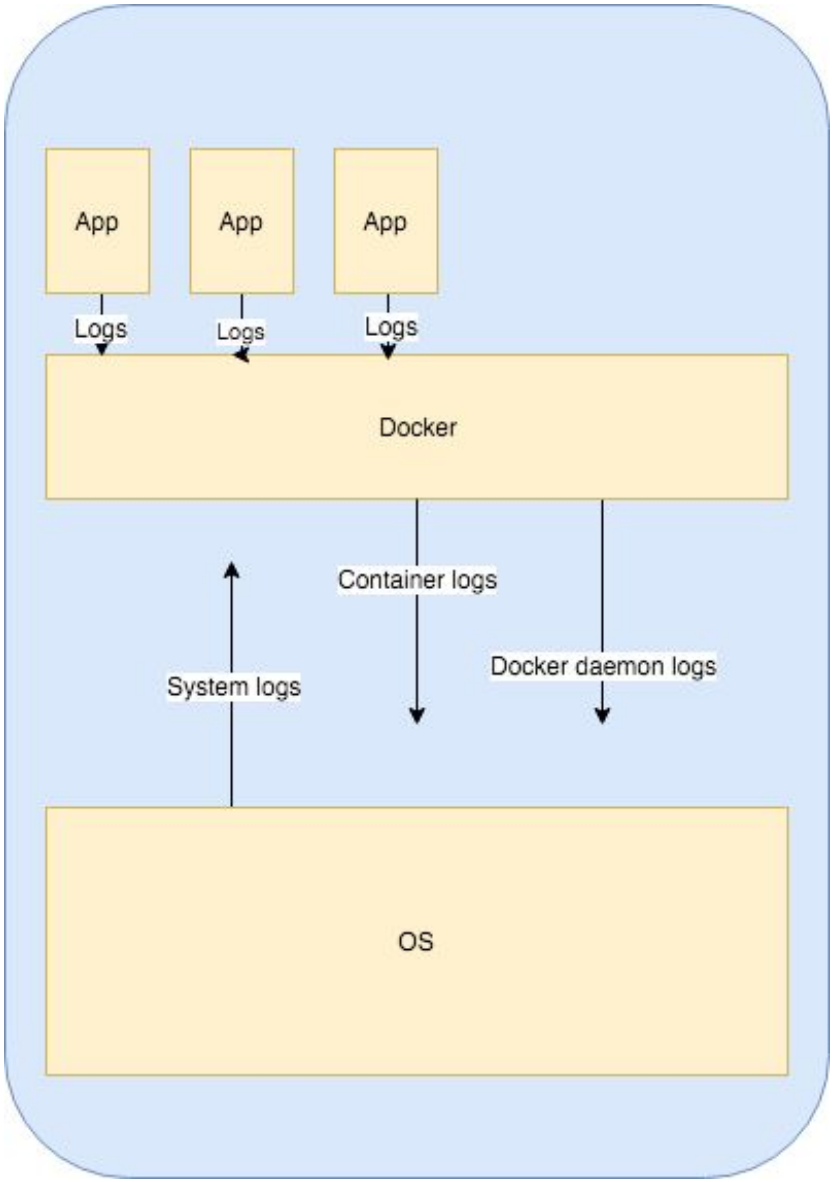

APACHE
HTTP SERVER
rsync server
mongoDB
Log Files
script to parse data
File
aggregation script
Tweet-fetching script
hadoop HDFS
aggregation script
syslog-ng
syslog script
elasticsearch.
filtering script
script to parse data
cron job for loading
amazon web services™


14:48
$00700911
ПОТРАЧЕНО
CJ
OG
N

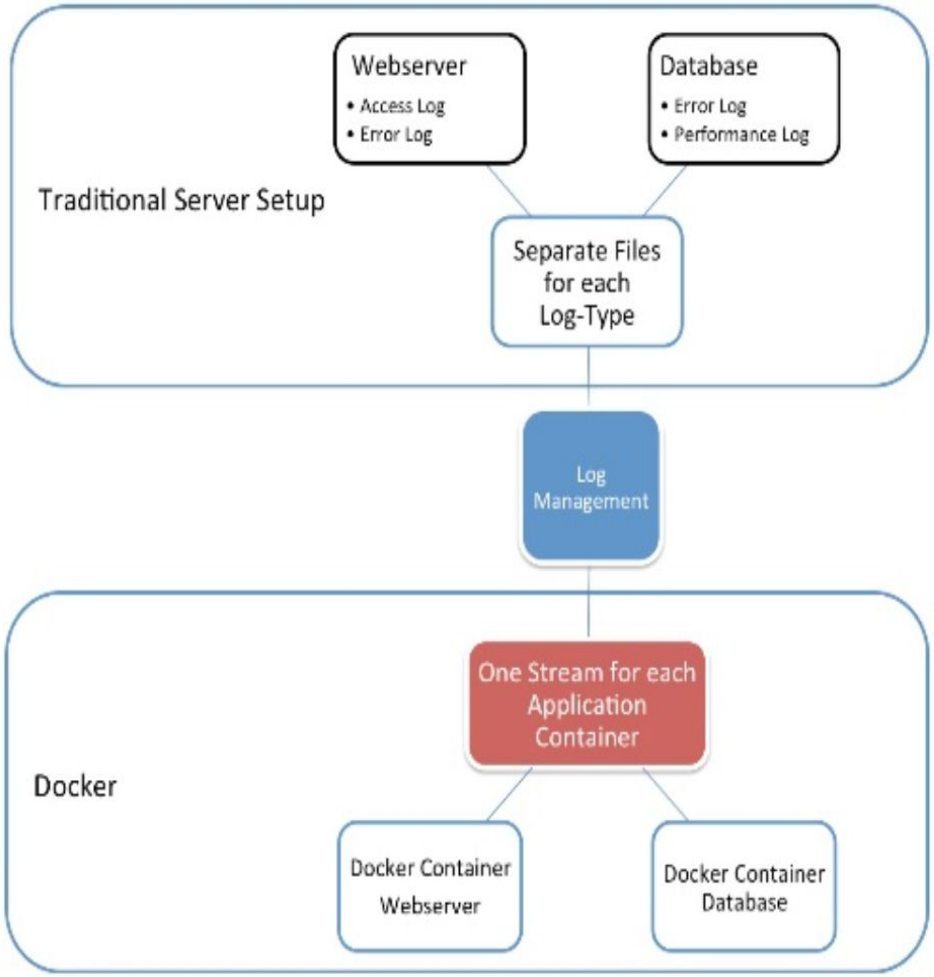
Docker логи



Логи на ноде


App
App
App
Logs
Logs
Logs
Docker
Container logs
Docker daemon logs
System logs
OS

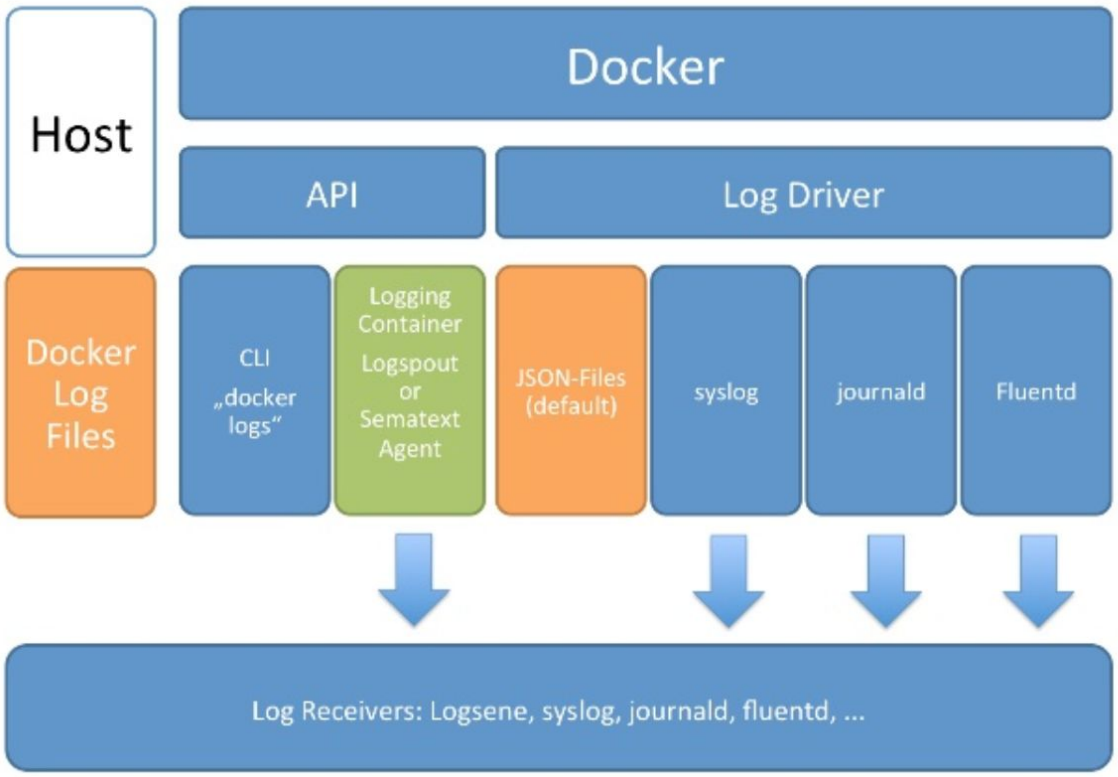
Логи приложений в docker



- Traditionally separate files for each Application and Log-Type
  - error.log
  - access.log

- Docker Logs are stdout / stderr of processes running in a container
- Most official images log to console

Docker



- **Docker Log Drivers**
  - json-file, syslog, fluentd, journald, gelf
- **Docker API based Logging Containers**
  - Logspout
  - Sematext Docker Container
- **Custom images with installed log shipper (syslog)**

Docker log driver

+ Простой путь доставки логов в local/remote destination
+ Настраивается per container или глобально

- Нет настройки парсинга
- Контейнер падает если удаленный приемник логов недоступен
- Будет плохо если диски медленные

Docker logging drivers

- none
- json-file
- syslog
- journald
- gelf
- fluentd
- awslogs
- splunk
- etwlogs
- gcplogs

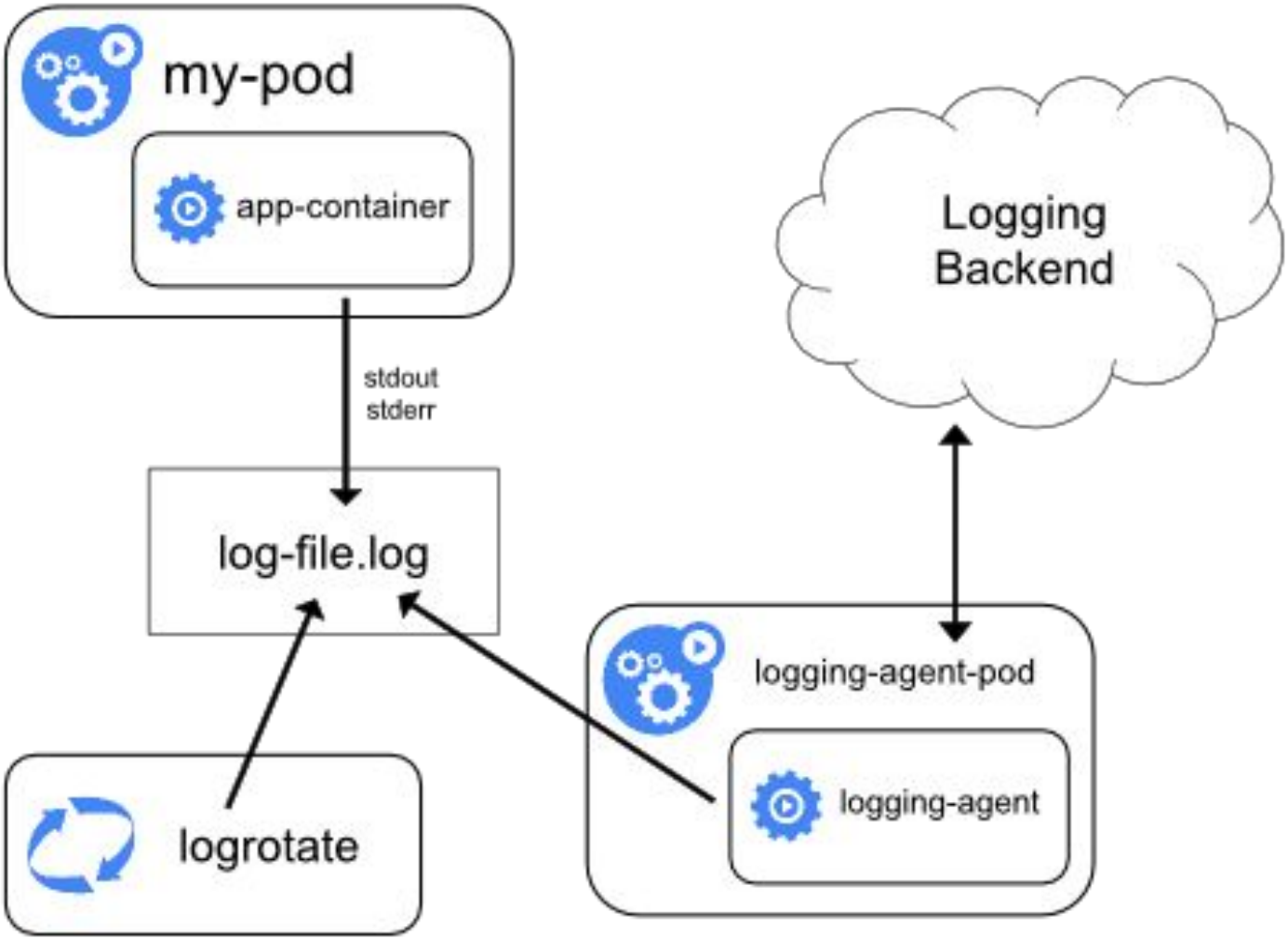
Benchmarks

Запускаем приложение, которое выдает на stdout 30000 сообщений (в формате json) длиной ~530 байт.

Результаты:

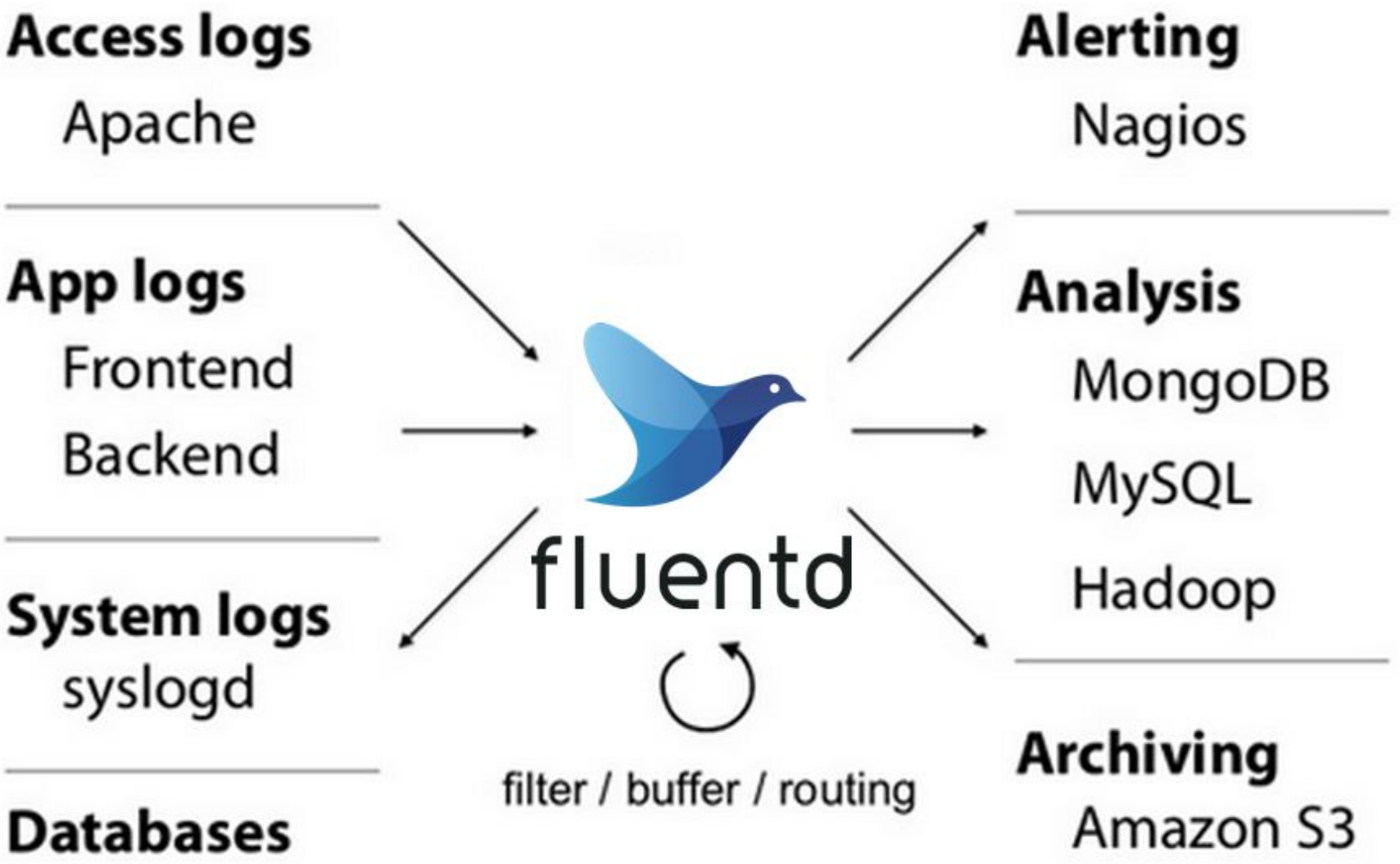
- json-file 0.2s
- journald 0.9s
- syslog 0.75s

Logging agent

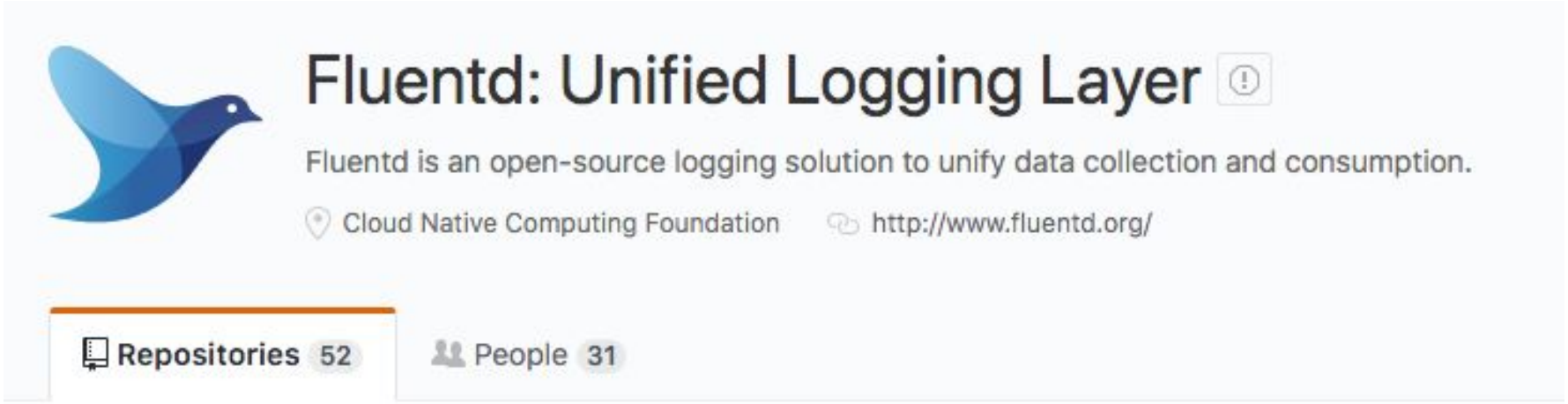


Решение

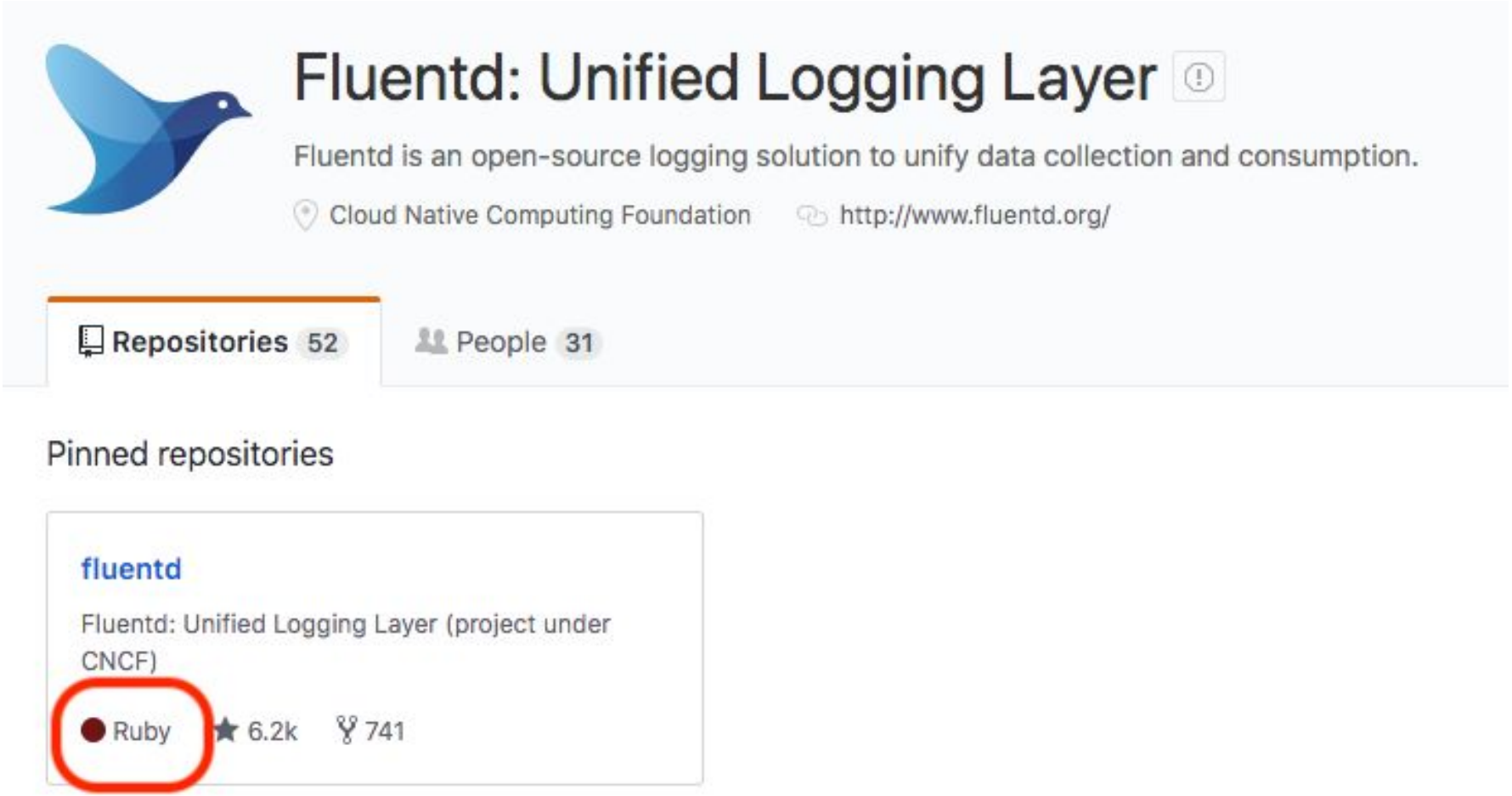
Готовое решение



Fluentd


Fluentd: Unified Logging Layer
Fluentd is an open-source logging solution to unify data collection and consumption.
Cloud Native Computing Foundation
http://www.fluentd.org/
Repositories 52
People 31
Pinned repositories
fluentd
Fluentd: Unified Logging Layer (project under CNCF)
Ruby
6.2k
741

Fluentd

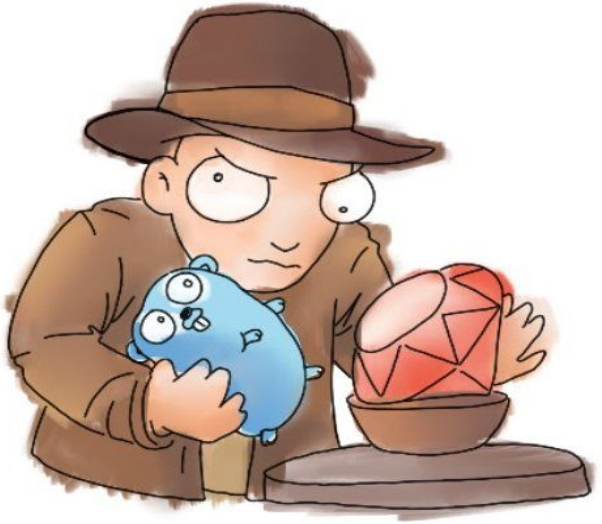

Fluentd: Unified Logging Layer
Fluentd is an open-source logging solution to unify data collection and consumption.
Cloud Native Computing Foundation
http://www.fluentd.org/
Repositories 52
People 31
Pinned repositories
fluentd
Fluentd: Unified Logging Layer (project under CNCF)
Ruby
6.2k
741

Fluentd

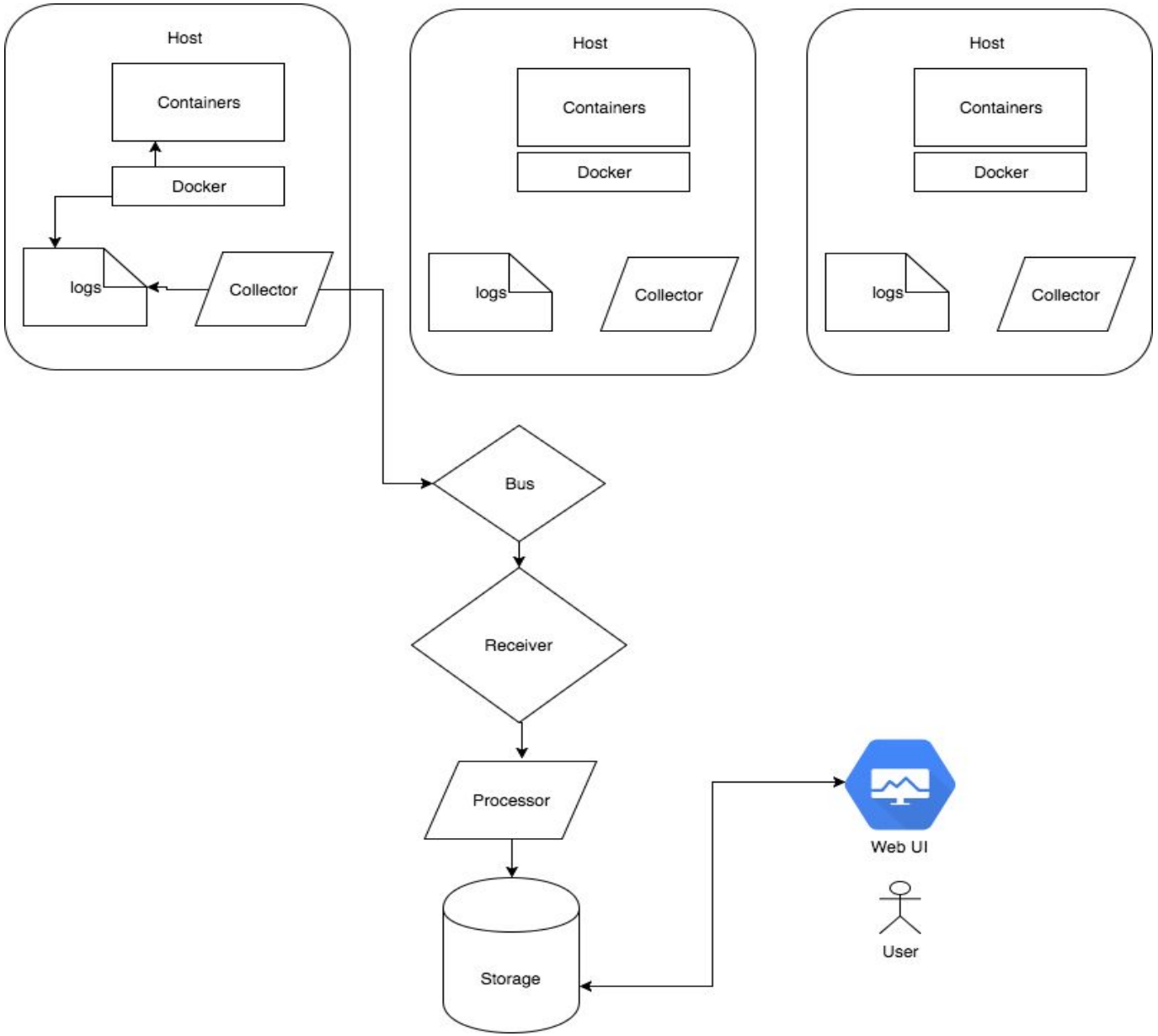

ты говорила, что
пишешь бекенд
А на каком языке?
На руби
Смотри,
берешь гем

Если серьезно

1. Из команды никто не знает ruby
2. Fluentd есть память больше чем наши приложения
3. Flexibility менее важно чем system footprint



Схема

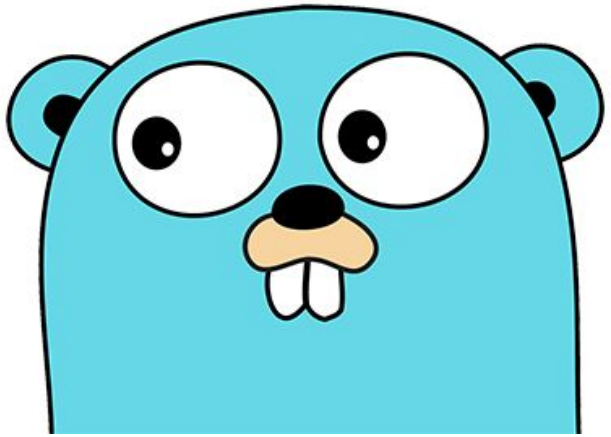

Host
Containers
Docker
logs
Collector
Host
Containers
Docker
logs
Collector
Host
Containers
Docker
logs
Collector
Bus
Receiver
Processor
Storage
Web UI
User

Вопросы

1. Что использовать для снятия логов с машин и сервисов?
2. Какой транспорт использовать для передачи логов от коллектора до хранилища?
3. Что использовать для хранилища?
4. Нужен ли компонент обработчик (фильтр например)?
5. Где разворачивать хранилище?
6. Какой инфраструктурой пользоваться?
7. Как мониторить состояние системы?

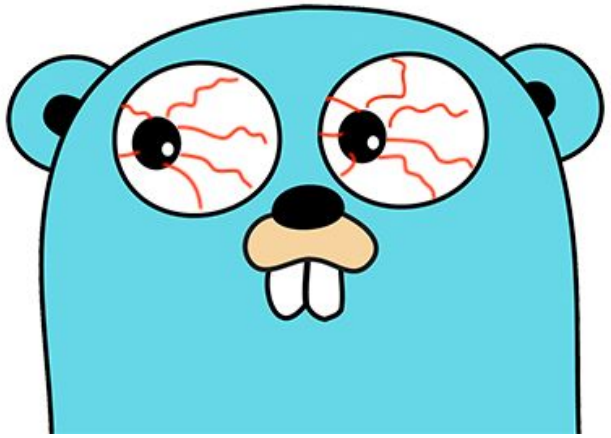
Collector

- logspout
- rsyslog
- syslog-ng
- filebeat
- fluent-bit
- nxlog
- hecka
- systemd journald remote
- custom
- ...



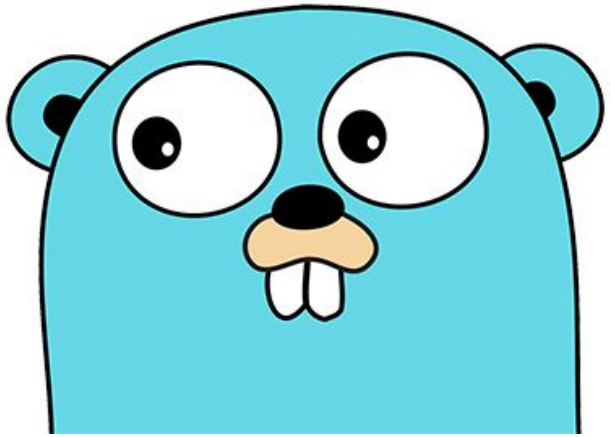
Bus

- Kafka
- Scribe
- Flume
- Kestrel
- Fluentd
- RabbitMQ
- BookKeeper
- Calligraphus
- ActiveMQ
- ...



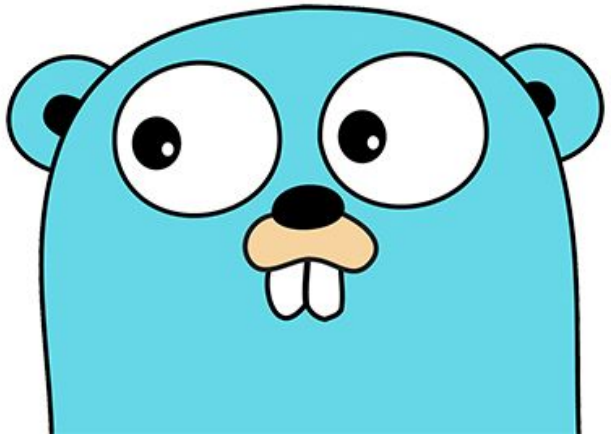
Receiver

- logstash
- fluent-bit
- syslog-ng
- custom
- ...



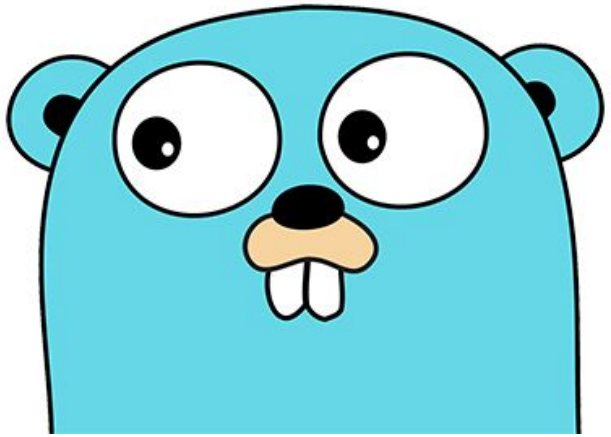
Storage

- Elasticsearch
- Mongo
- RDBMS
- ClickHouse
- ...

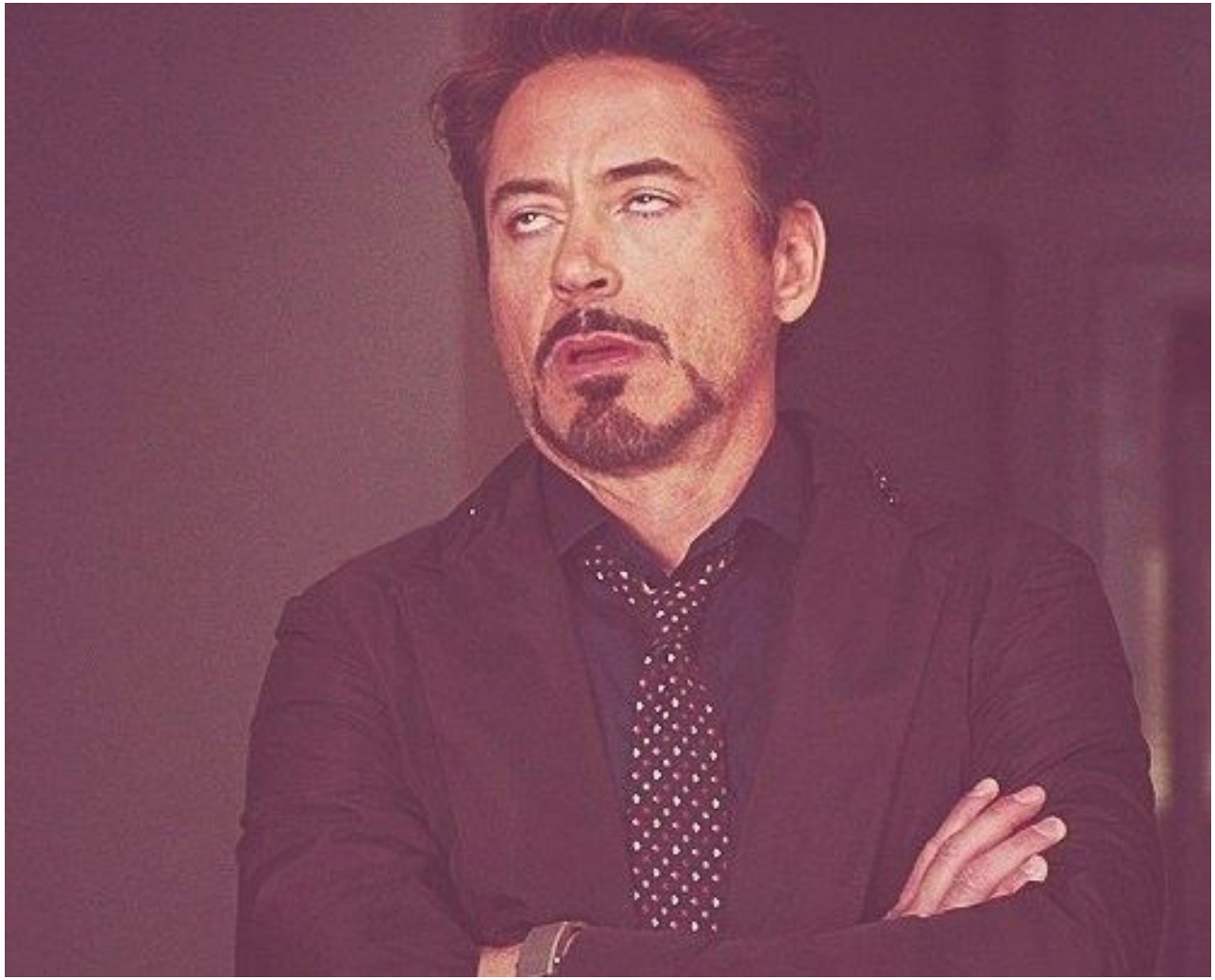


Web UI

- Kibana
- Grafana

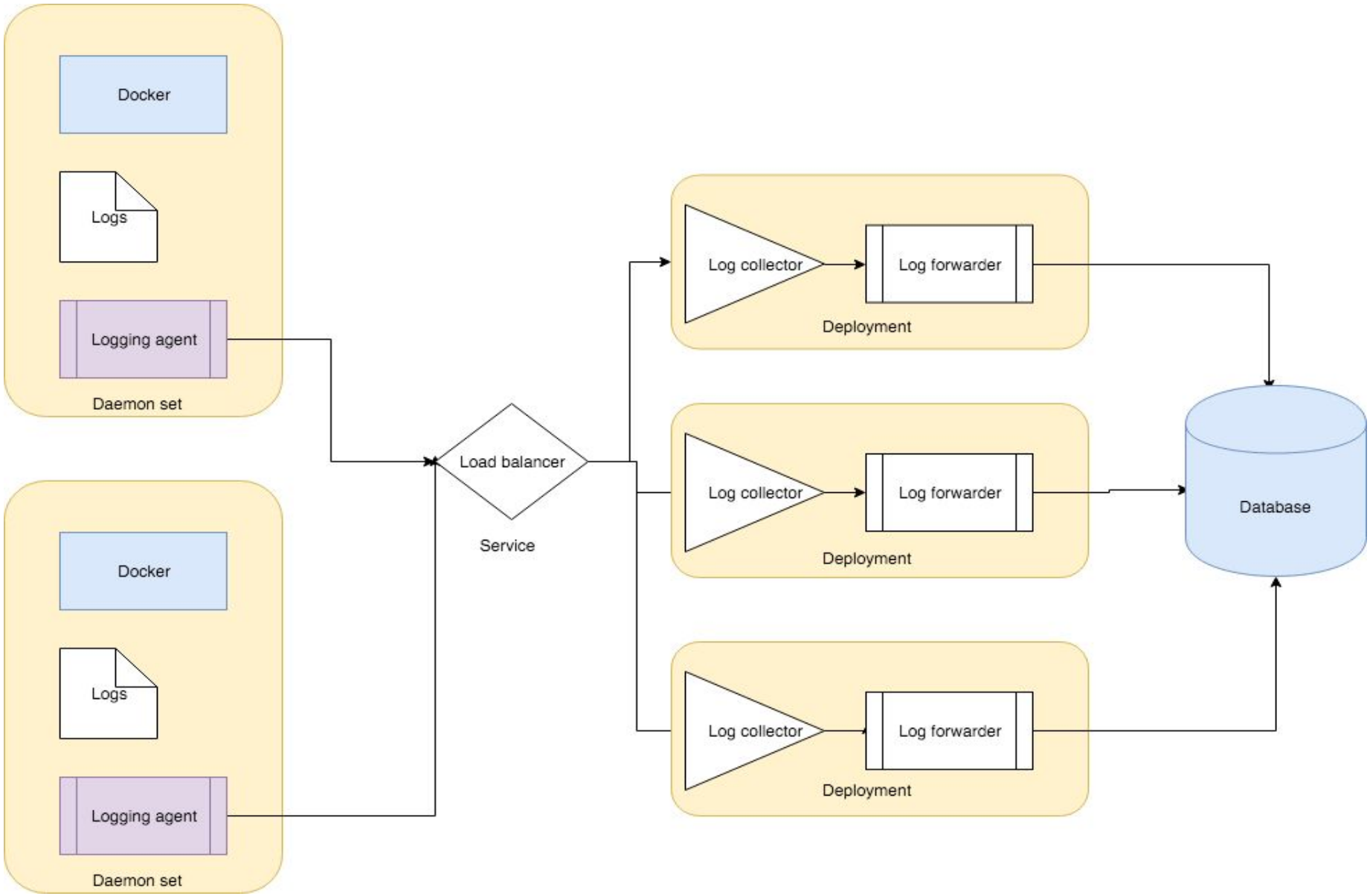


OMG



С начала

Снова схема


Docker
Logs
Logging agent
Daemon set
Docker
Logs
Logging agent
Daemon set
Load balancer
Service
Log collector
Log forwarder
Deployment
Log collector
Log forwarder
Deployment
Log collector
Log forwarder
Deployment
Database

Log agent: syslog-ng vs fluent-bit

syslog-ng (си)

+ Позволяет описать много input, output и parser для одного экземпляра демона
- Нет никакой интеграции с docker или k8s

fluent-bit (си)

+ Есть плагин для работы с k8s api
- Умеет только один input и output

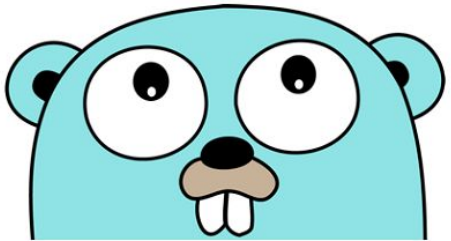
Log agent: syslog-ng vs fluent-bit

syslog-ng

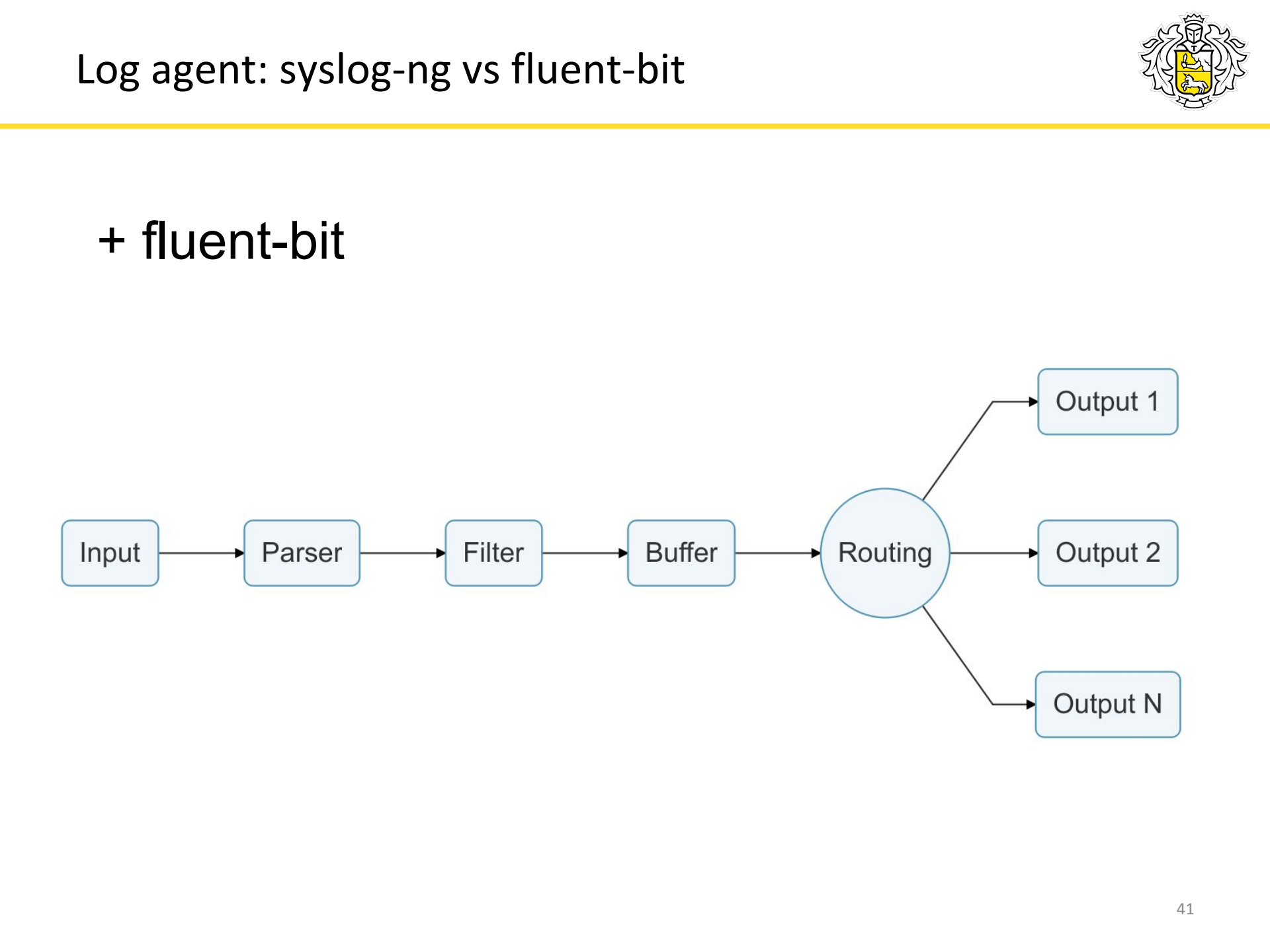
- На нагрузке 10000 в сек падает и ломает указатели на позицию файла журналов
- После перезапуска падал в segfault

fluent-bit

- Проблем не обнаружено



Log agent: syslog-ng vs fluent-bit



+ fluent-bit

Storage: elasticsearch vs clickhouse

Search engine + storage - Elasticsearch (Java)

- Популярное решение, используется в Graylog, ELK
- Есть UI - Kibana
- Позволяет хранить без схемы
- Query language - lucene

Column based - ClickHouse (C++)

- Строгая схема
- UI нет, только сторонний
- Query language - SQL

RDBMS

- Сложно масштабировать, нет кластеризации, performance

Storage: elasticsearch vs clickhouse

Нода database - Ubuntu 16.04.3 x64

- Intel(R) Xeon(R) Platinum 8168 CPU @ 2.70GHz 4 core
- 6 GB Memory
- SSD

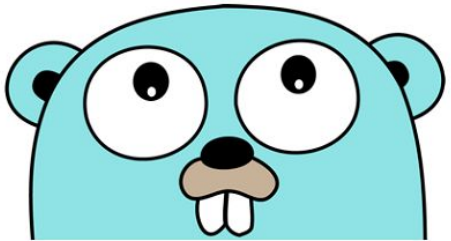
Нода load generator - Ubuntu 16.04.3 x64

- Intel(R) Xeon(R) Platinum 8168 CPU @ 2.70GHz 2 core
- 3 GB Memory
- SSD

Storage: elasticsearch vs clickhouse

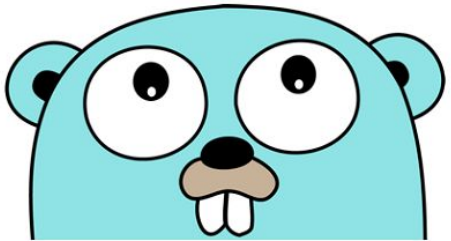
Elasticsearch 100 000 000 записей

Съел всю память в хипе и 12 Gb места на диске



Storage: elasticsearch vs clickhouse

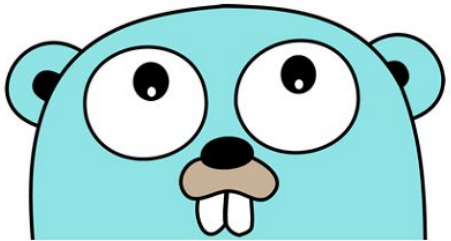
```
1000000 msg, bulk 1000000

real  0m23.633s
user  0m22.000s
sys 0m0.580s

# du -csh elasticsearch/nodes

390M
```



Storage: elasticsearch vs clickhouse

Clickhouse 100 000 000 записей

Съел 400mb памяти и 3.3 Gb места на диске



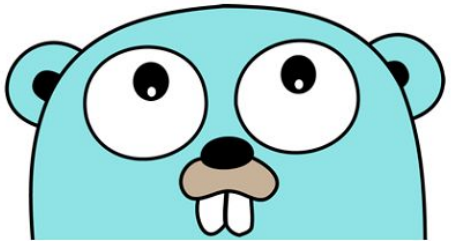
Storage: elasticsearch vs clickhouse

```
1000000 msg, bulk 1000000

real 0m4.134s
user 0m6.972s
sys 0m0.180s

# du -csh clickhouse/data

30M
```

Storage: elasticsearch vs clickhouse

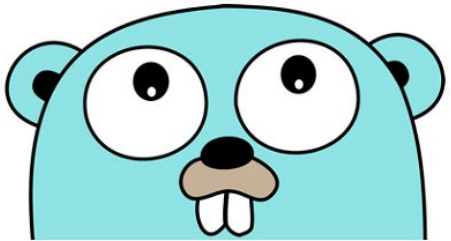
\+ Clickhouse

Оптимально по performance и месту на диске



Проблемы

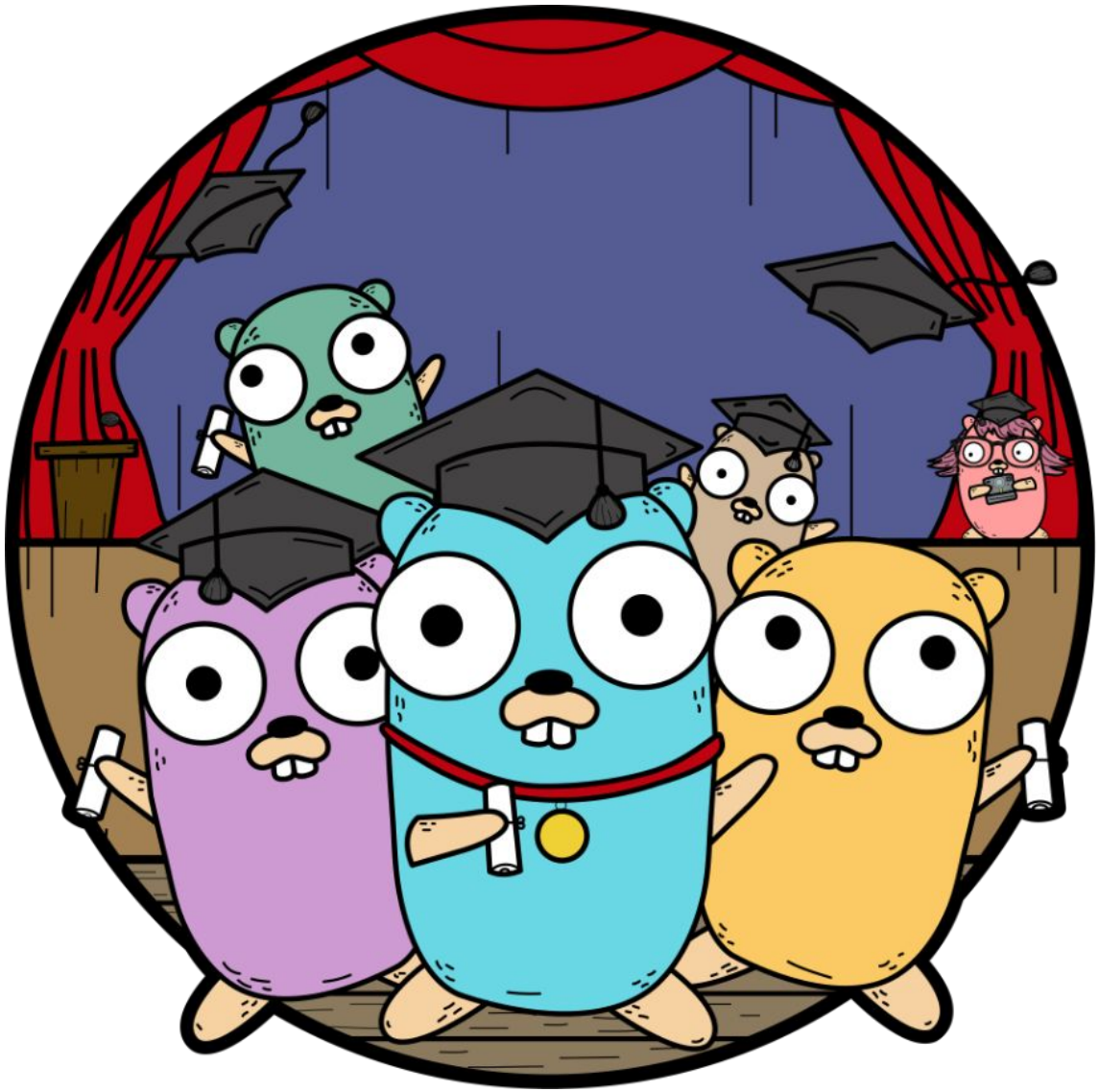
1. fluent-bit не умеет писать в clickhouse
2. нужно парсить, чтобы разложить в типизированную схему



Проблема разных логов

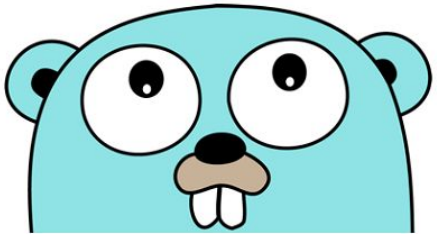
```
> docker run -d -p 80:80 --rm --name apache-app -v "$PWD":/usr/local/apache2/htdocs/ httpd:2.4
> docker logs apache-app

AH00558: httpd: Could not reliably determine the server's fully qualified domain name, using 172.17.0.2. Set the 'ServerName'
    directive globally to suppress this message
AH00558: httpd: Could not reliably determine the server's fully qualified domain name, using 172.17.0.2. Set the 'ServerName'
    directive globally to suppress this message
[Sat Jul 25 11:52:13.381905 2015] [mpm_event:notice] [pid 1:tid 140020361725824] AH00489: Apache/2.4.16 (Unix) configured --
    resuming normal operations
[Sat Jul 25 11:52:13.382133 2015] [core:notice] [pid 1:tid 140020361725824] AH00094: Command line: 'httpd -D FOREGROUND'
192.168.59.3 - - [25/Jul/2015:11:52:22 +0000] "GET / HTTP/1.1" 200 429
192.168.59.3 - - [25/Jul/2015:11:52:22 +0000] "GET /favicon.ico HTTP/1.1" 404 209
192.168.59.3 - - [25/Jul/2015:11:52:22 +0000] "GET /favicon.ico HTTP/1.1" 404 209
192.168.59.3 - - [25/Jul/2015:11:52:29 +0000] "GET /index.js HTTP/1.1" 200 1258
192.168.59.3 - - [25/Jul/2015:11:53:12 +0000] "GET /index.js HTTP/1.1" 304 -
192.168.59.3 - - [25/Jul/2015:11:53:18 +0000] "GET /index.js HTTP/1.1" 200 1258
```

Дипломированные Go специалисты

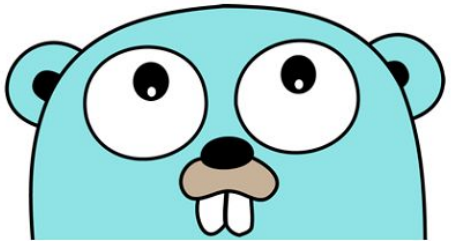


Clickhouse forwarder

- Демон на Go
- Работает через pipe с fluent-bit
- Парсит разные типы логов
- Собирает batch для записи в Clickhouse

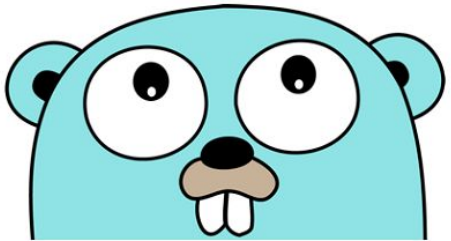


Clickhouse forwarder



Clickhouse forwarder

1. github.com/kshvakov/clickhouse
2. github.com/mailru/easyjson
3. github.com/hashicorp/logutils
4. github.com/facebookgo/flagenv



Deploy - Helm chart

Chart - пакет для kubernetes

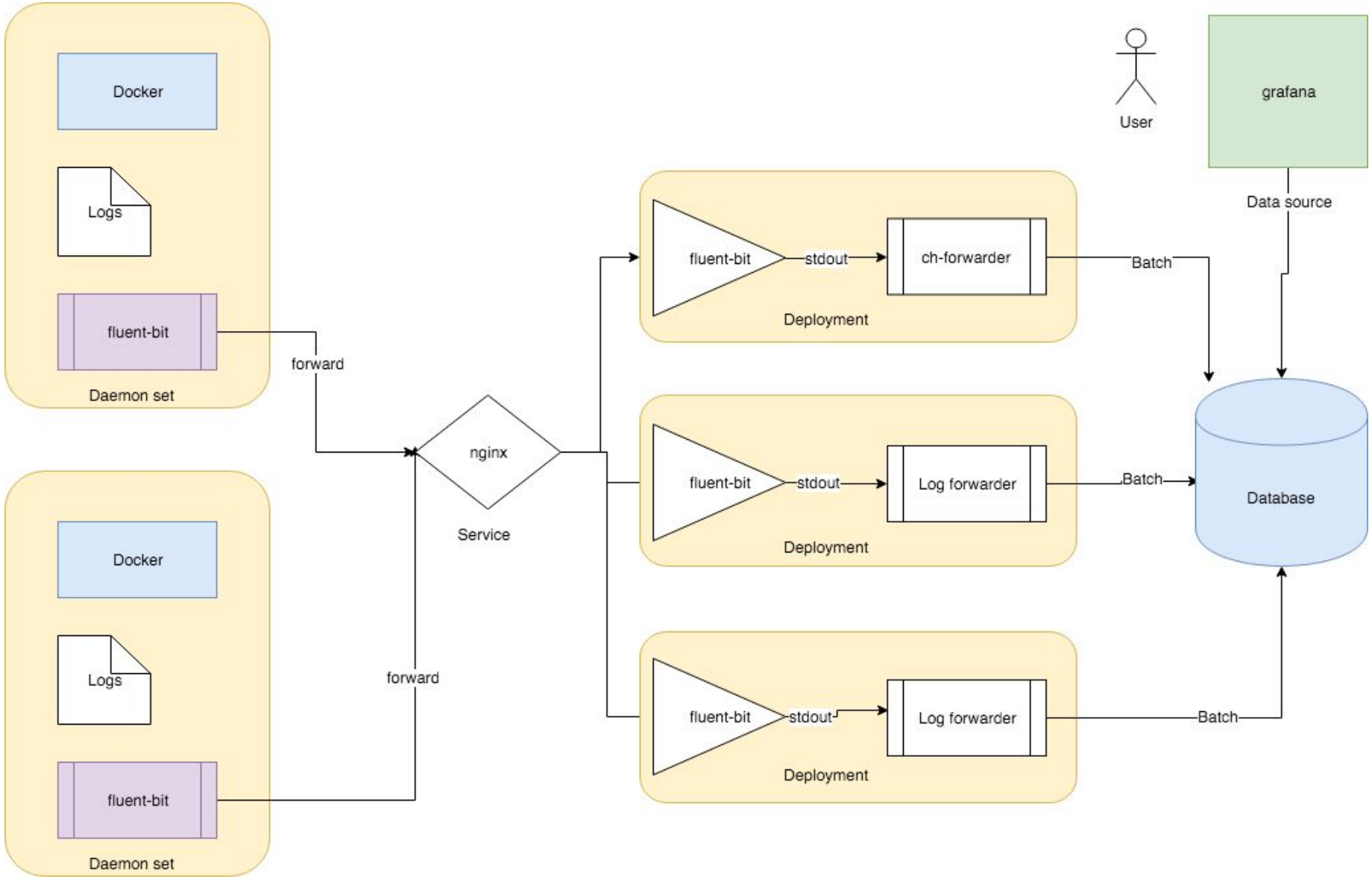
Create, Update, Delete операции

содержит шаблоны ресурсов, переменные, скрипты

Deploy - Helm chart

```
logging
├── Chart.yaml
├── files
│   ├── create_merges.sh
│   └── create_partitions.sh
├── templates
│   ├── _helpers.tpl
│   ├── db-pre-install.job.yaml
│   ├── db-rotate.cronjob.yaml
│   ├── db-scripts.configmap.yaml
│   ├── db.secret.yaml
│   ├── log-agent.clusterrole.yaml
│   ├── log-agent.clusterrolebinding.yaml
│   ├── log-agent.configmap.yaml
│   ├── log-agent.daemonset.yaml
│   ├── log-agent.serviceaccount.yaml
│   ├── log-collector.configmap.yaml
│   ├── log-collector.deployment.yaml
│   └── log-collector.service.yaml
└── values.yaml
```

Схема

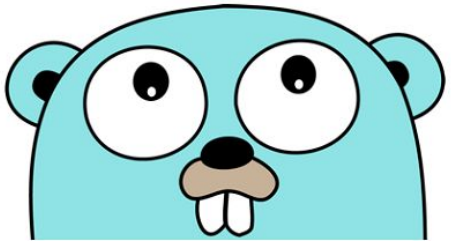

Docker
Logs
fluent-bit
Daemon set
forward
nginx
Service
Docker
Logs
fluent-bit
Daemon set
forward
fluent-bit
stdout
ch-forwarder
Deployment
Batch
fluent-bit
stdout
Log forwarder
Deployment
Batch
fluent-bit
stdout
Log forwarder
Deployment
Batch
User
grafana
Data source
Database

Итоги проекта

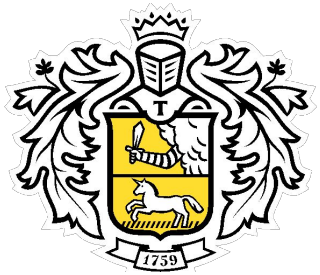
Срок разработки - 2 мес
Пропускная способность - 10 000 msg/sec с машины

Что не доделали:

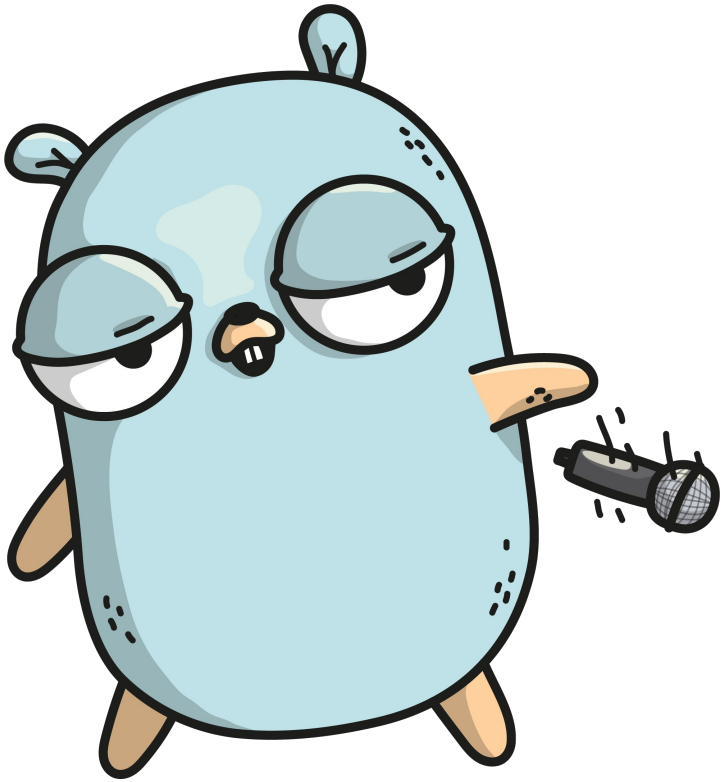
- UI еще в разработке, нужен plugin для grafana



Когда рассказываешь об этом проекте



Тинькофф



Tinkoff.ru