Fluentd

Open Source Data Collector

Eduardo Silva
eduardo@treasuredata.com
@edsiper

Jan 23, 2016
Scale14x, Pasadena!

spread the word!

#scale14x #fluentd

@edsiper

About Me

TREASURE DATA

Eduardo Silva

- Github & Twitter @edsiper
- Personal Blog http://edsiper.linuxchile.cl

Treasure Data

- Open Source Engineer
- Fluentd / Fluent Bit http://github.com/fluent

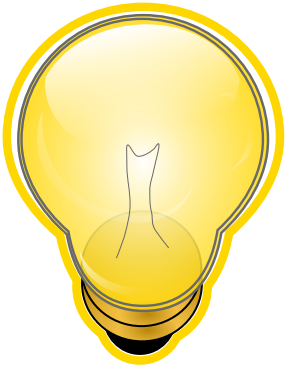
Projects

- Monkey HTTP Server http://monkey-project.com
- Duda I/O http://duda.io

Logging

Logging Matters

Pros



- Application status
- Debugging
- General information about anomalies: errors
- Troubleshooting / Support
- Local or Remote (network)

Logging Matters

From a business point of view

- Input data → Analytics
- User interaction / behaviors
- Improvements

Assumptions

Logging Matters

Assumptions

- I have enough disk **space**
- I/O operations will **not** block
- Log messages are **human readable**
- My logging mechanism **scale**

Logging Matters

Assumptions

> *Basically, yeah.. it **should** work.*

Concerns

Logging Matters

Concerns

- Logs **increase** = data **increase**
- Message format get more **complex**
- Did the Kernel flush the buffers ? (sync(2))
- Multi-thread application ?, **locking** ?
- **M**ultiple Applications = **M**ultiple Logs

Logging Matters

Concerns

> *If **M**ultiple Applications = **M**ultiple logs*
>
> ***M**ultiple Hosts **x** **M**ultiple Applications **= ???***

OK, so:

1. Logging matters
2. It's really beneficial
3. but...

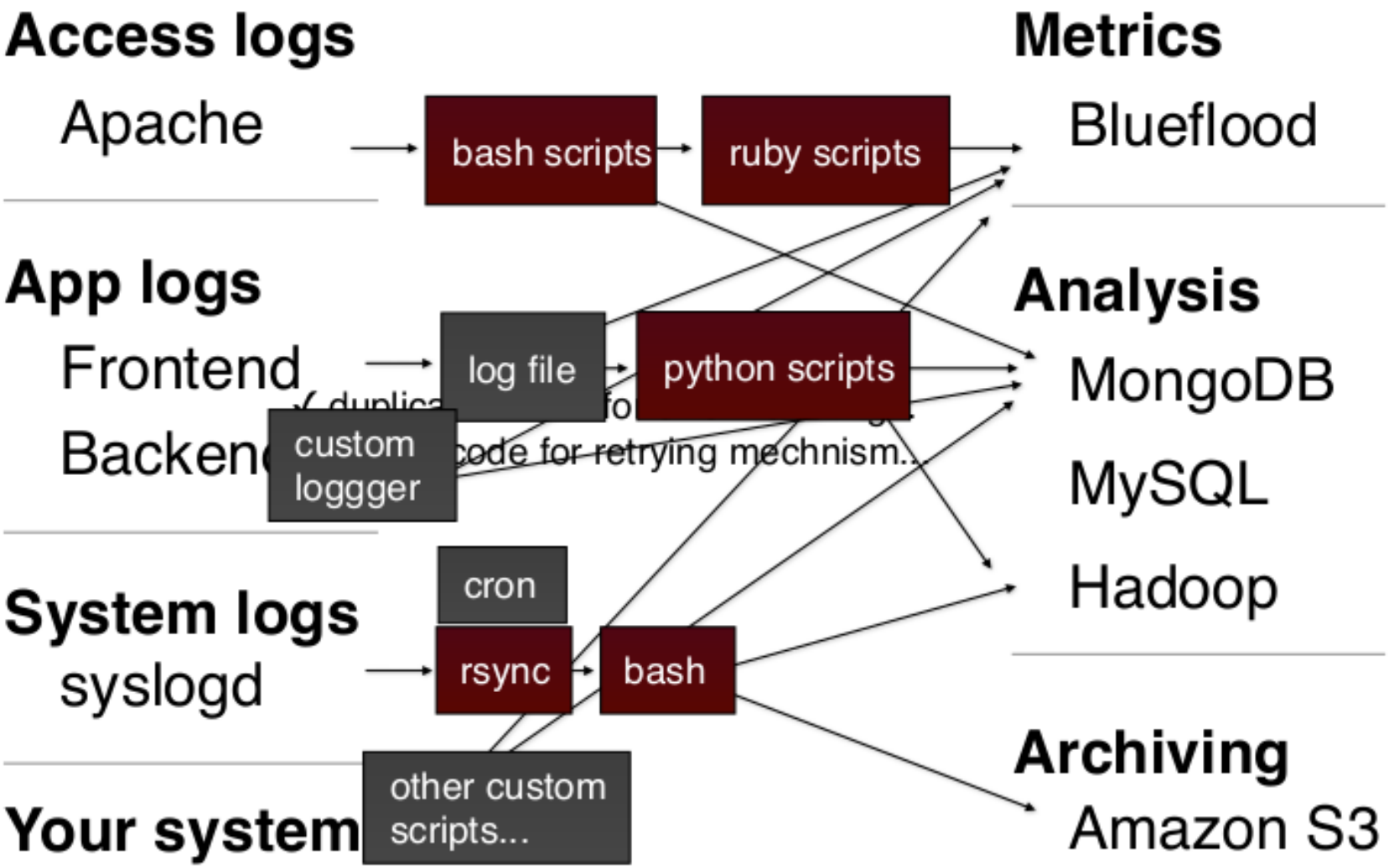
- It needs to be done right.

Logging

Common sources & inputs

- **Application** Logs
  - Apache
  - NginX
  - Syslog (-ng)
- **Custom** applications / Languages
  - C, Ruby, Python, PHP, Perl, NodeJS, Java, etc.

In a galaxy
not so far away...


Access logs
Apache
bash scripts
ruby scripts
Metrics
Blueflood
App logs
Frontend
log file
python scripts
Backen
custom loggger
code for retrying mechnism...
Analysis
MongoDB
MySQL
Hadoop
System logs
syslogd
cron
rsync
bash
Archiving
Amazon S3
Your system
other custom scripts...

How to parse/store multiple data sources ?

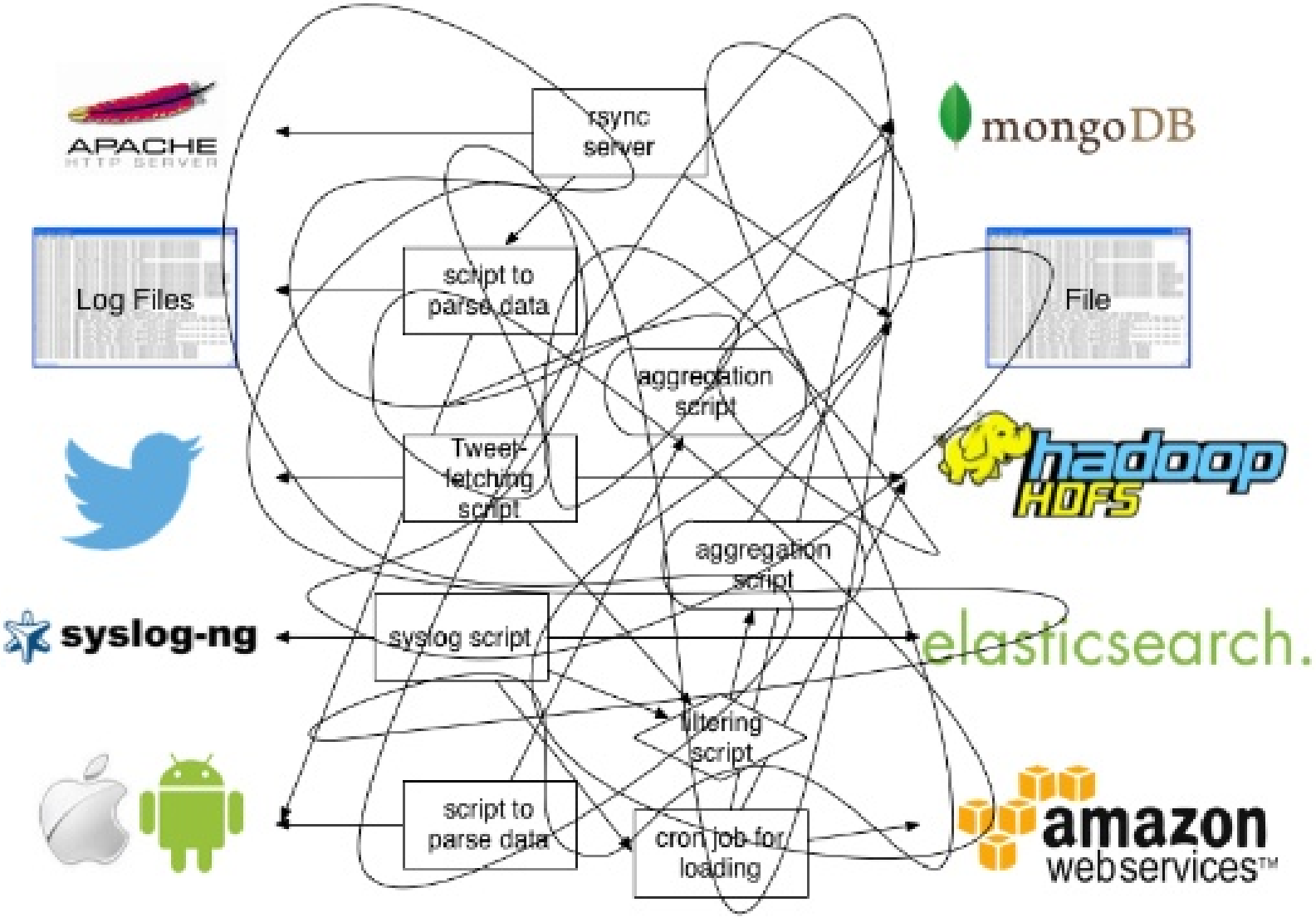
> note: performance matters!

fluentd

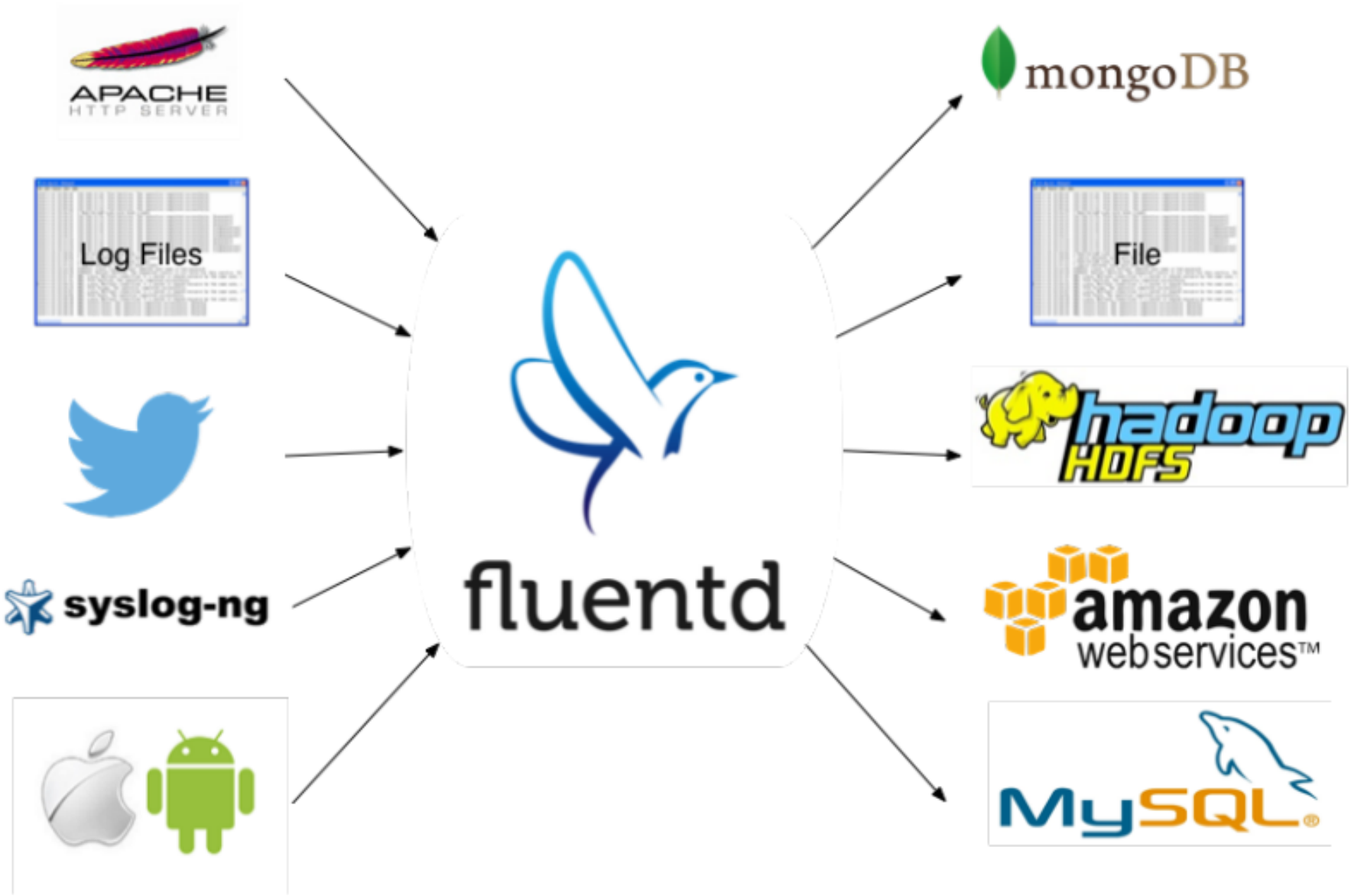
Fluentd is an open source data collector

> It let's you unify the data collection for a better use and understanding of data.

before


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
webservices™

after


APACHE
HTTP SERVER
Log Files
syslog-ng
fluentd
mongoDB
File
hadoop
HDFS
amazon
webservices™
MySQL

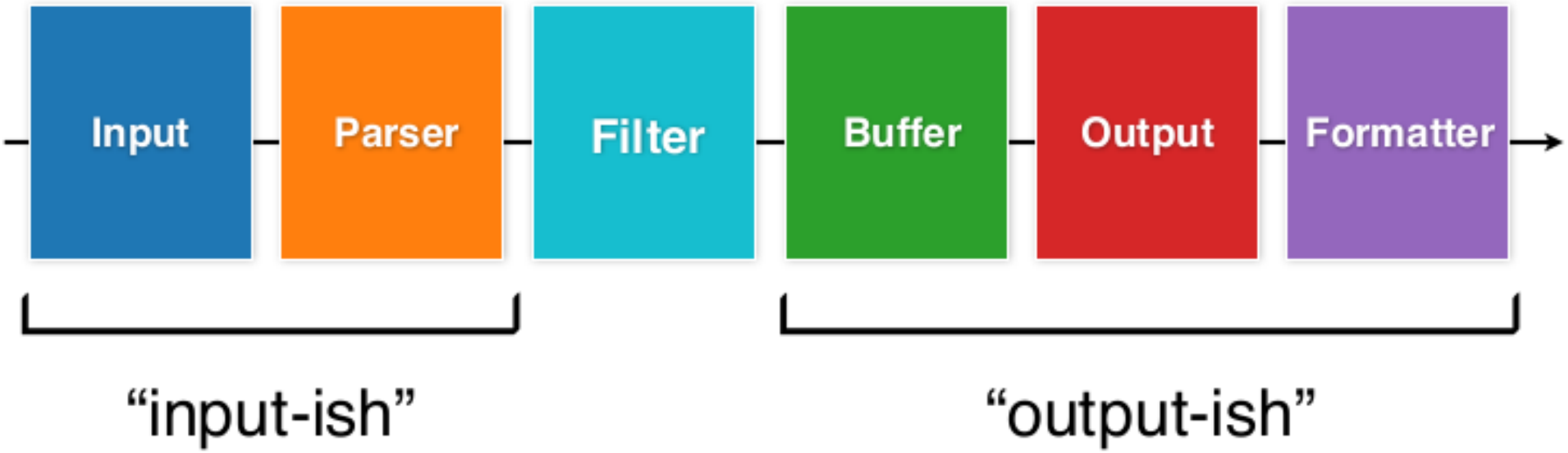
Fluentd

Highlights

- High **Performance**
- Built-in **Reliability**
- **Structured** Logs
- **Plug**gable Architecture
- More than 300 plugins! (input/filtering/output)

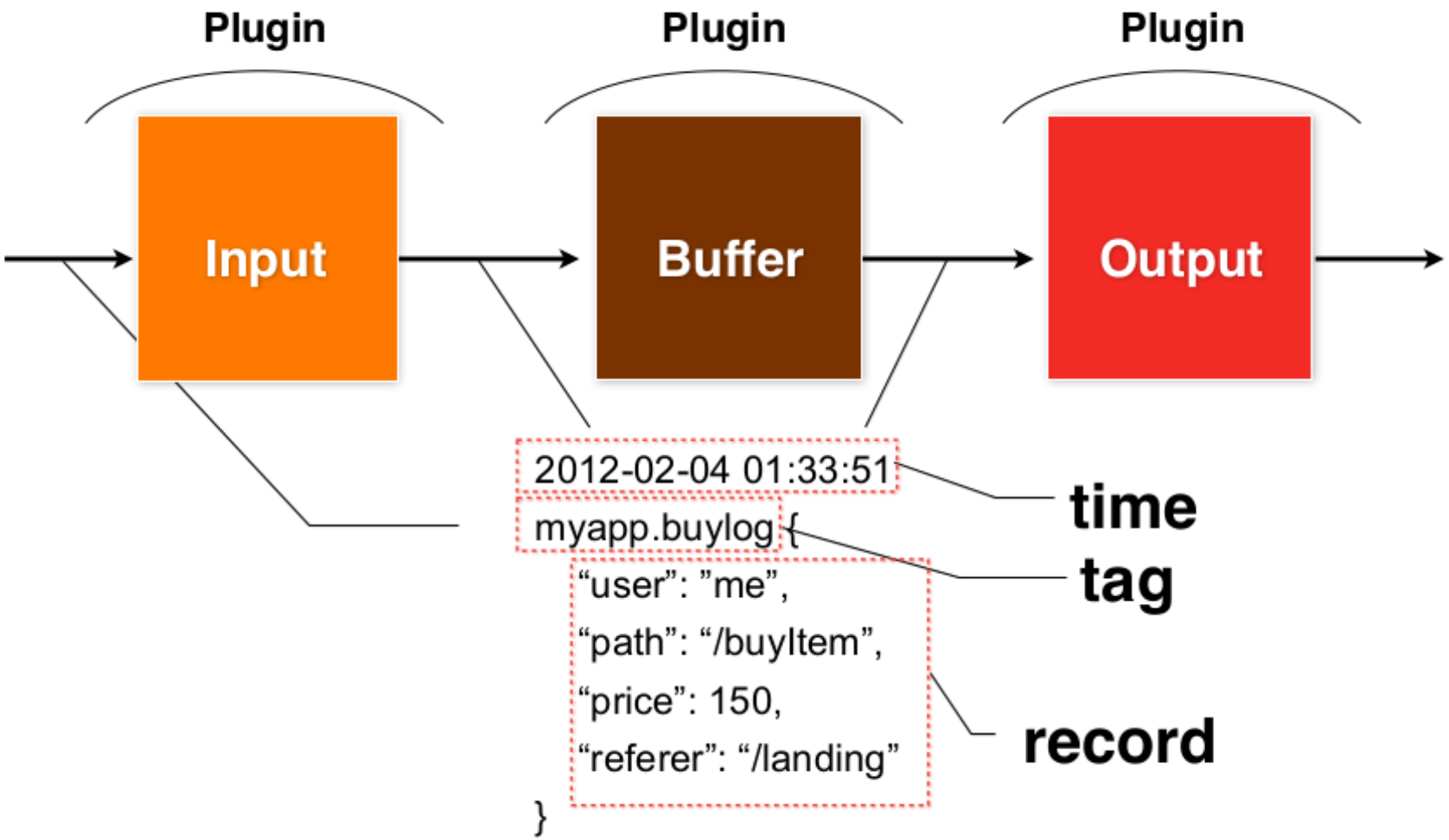
Fluentd

Architecture



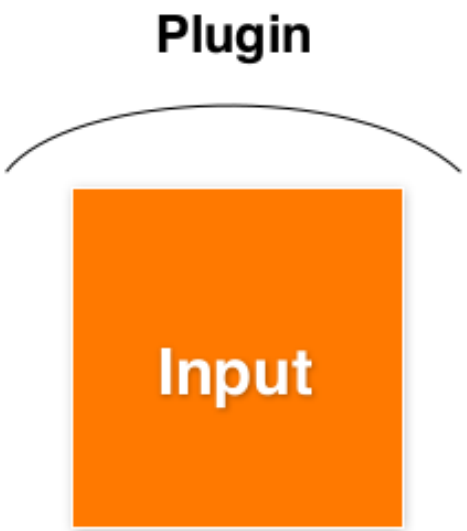
Fluentd

Internals simplified



Fluentd

Input plugins



HTTP+JSON (in_http)
File tail (in_tail)
Syslog (in_syslog)
...

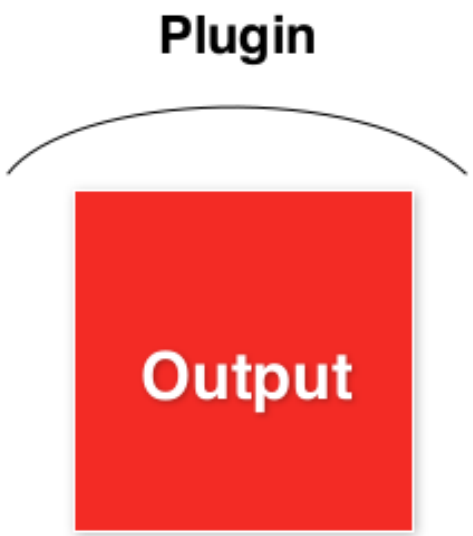
✓ Receive logs

✓ Or pull logs from data sources

✓ in non-blocking manner

Fluentd

Output plugins

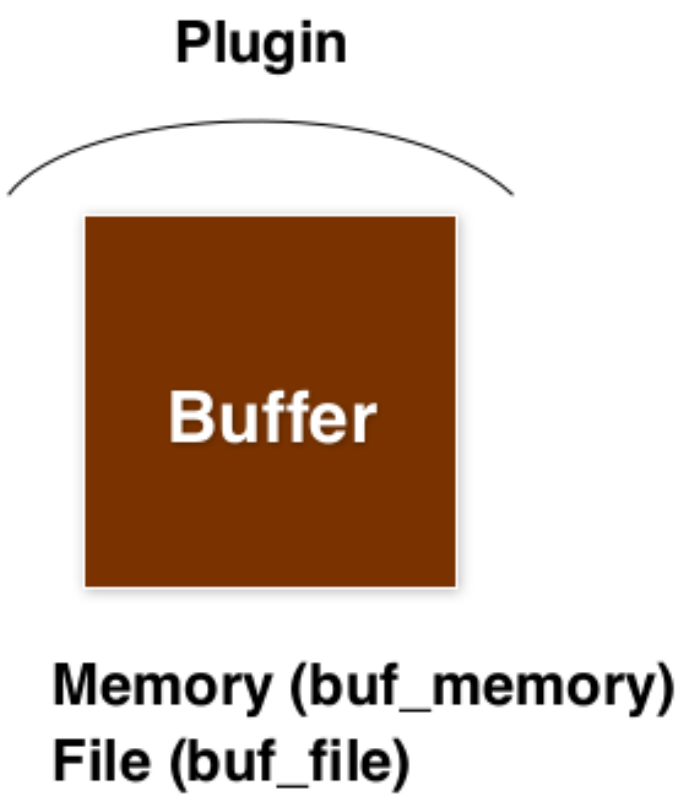


File (out_file)
Amazon S3 (out_s3)
MongoDB (out_mongo)
...

✓ Write or send event logs

Fluentd

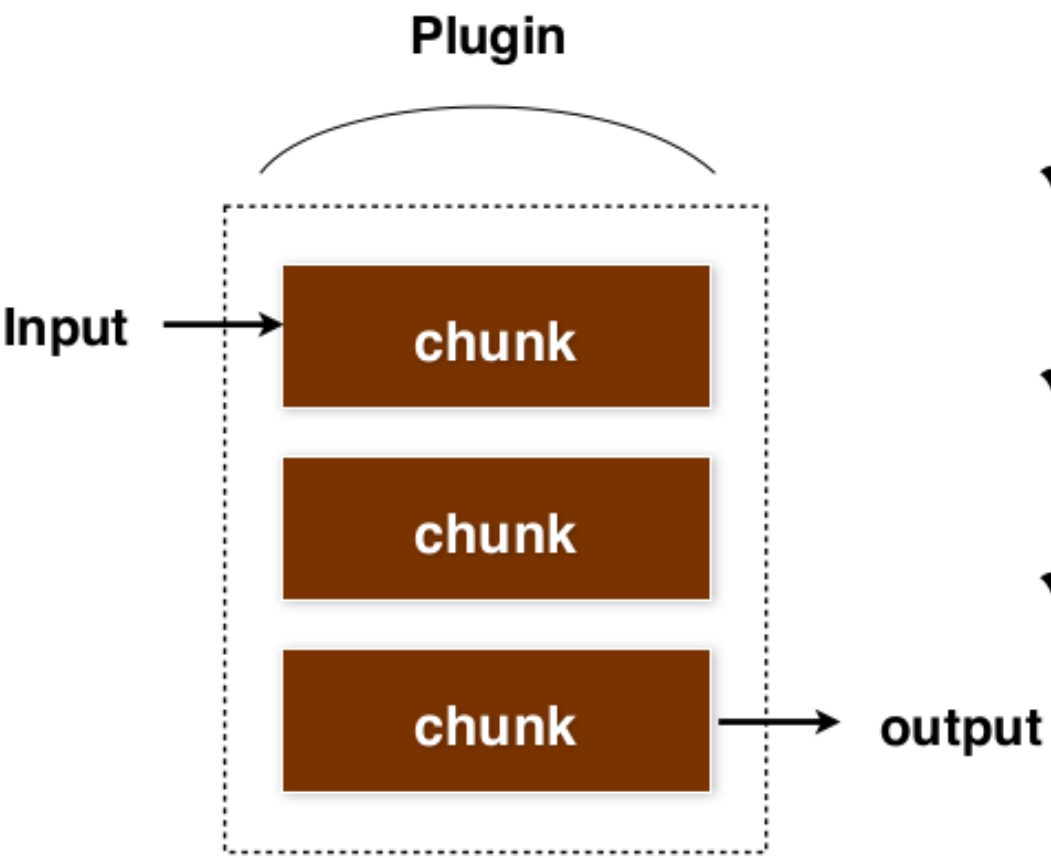
Buffer plugins



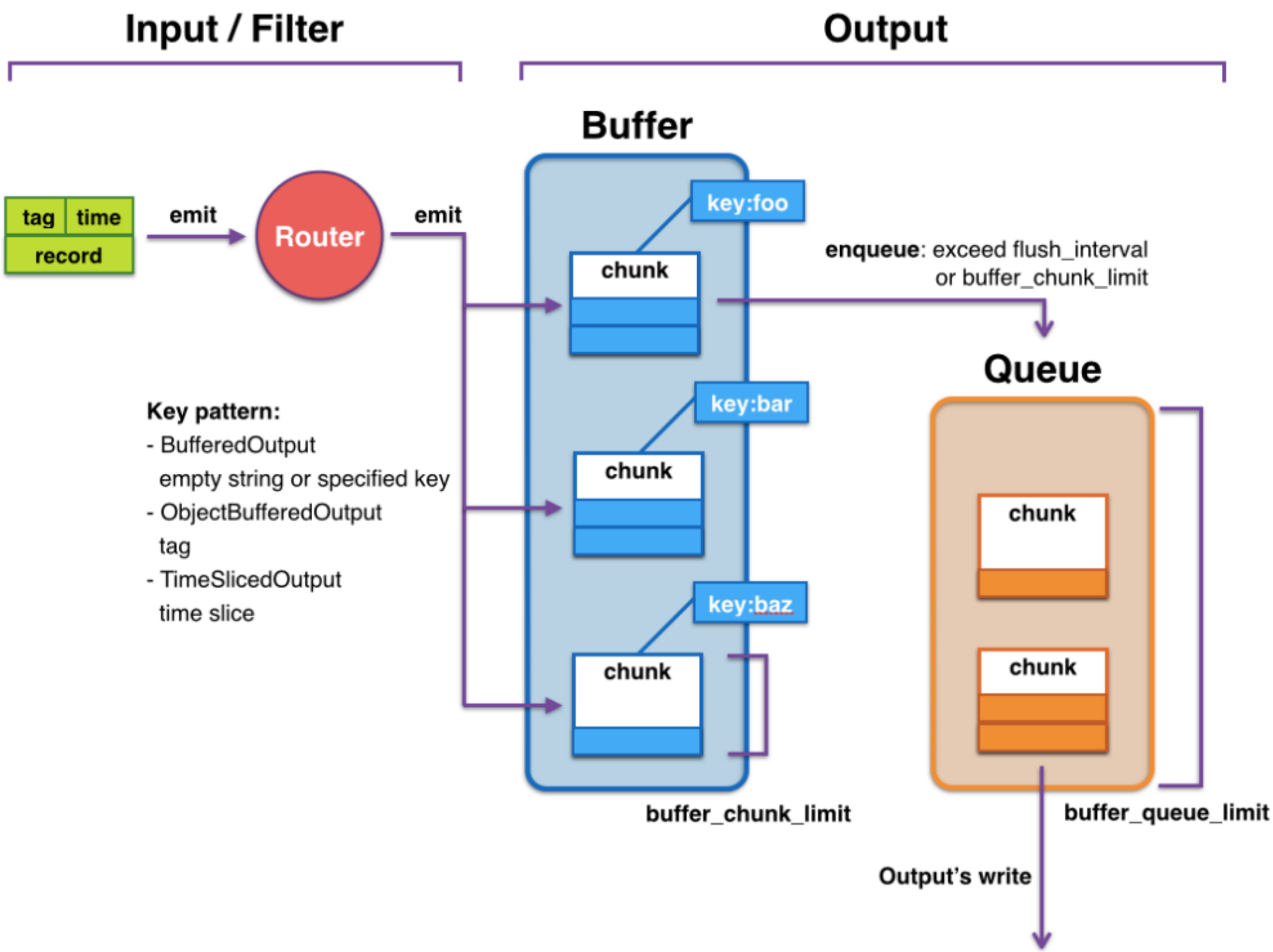
- ✓ Improve performance
- ✓ Provide reliability
- ✓ Provide thread-safety

Fluentd

Buffer plugins



- ✓ Improve performance
- ✓ Provide reliability
- ✓ Provide thread-safety


Input / Filter
Output
Buffer
tag
time
record
emit
Router
emit
key:foo
chunk
enqueue: exceed flush_interval
or buffer_chunk_limit
Queue
key:bar
chunk
Key pattern:
- BufferedOutput
empty string or specified key
- ObjectBufferedOutput
tag
- TimeSlicedOutput
time slice
chunk
key:baz
chunk
chunk
chunk
buffer_chunk_limit
buffer_queue_limit
Output's write

M x N → M + N

Access logs

Apache

App logs

Frontend

Backend

System logs

syslogd

Databases

fluentd

buffer/filter/route

Alerting

Nagios

Analysis

MongoDB

MySQL

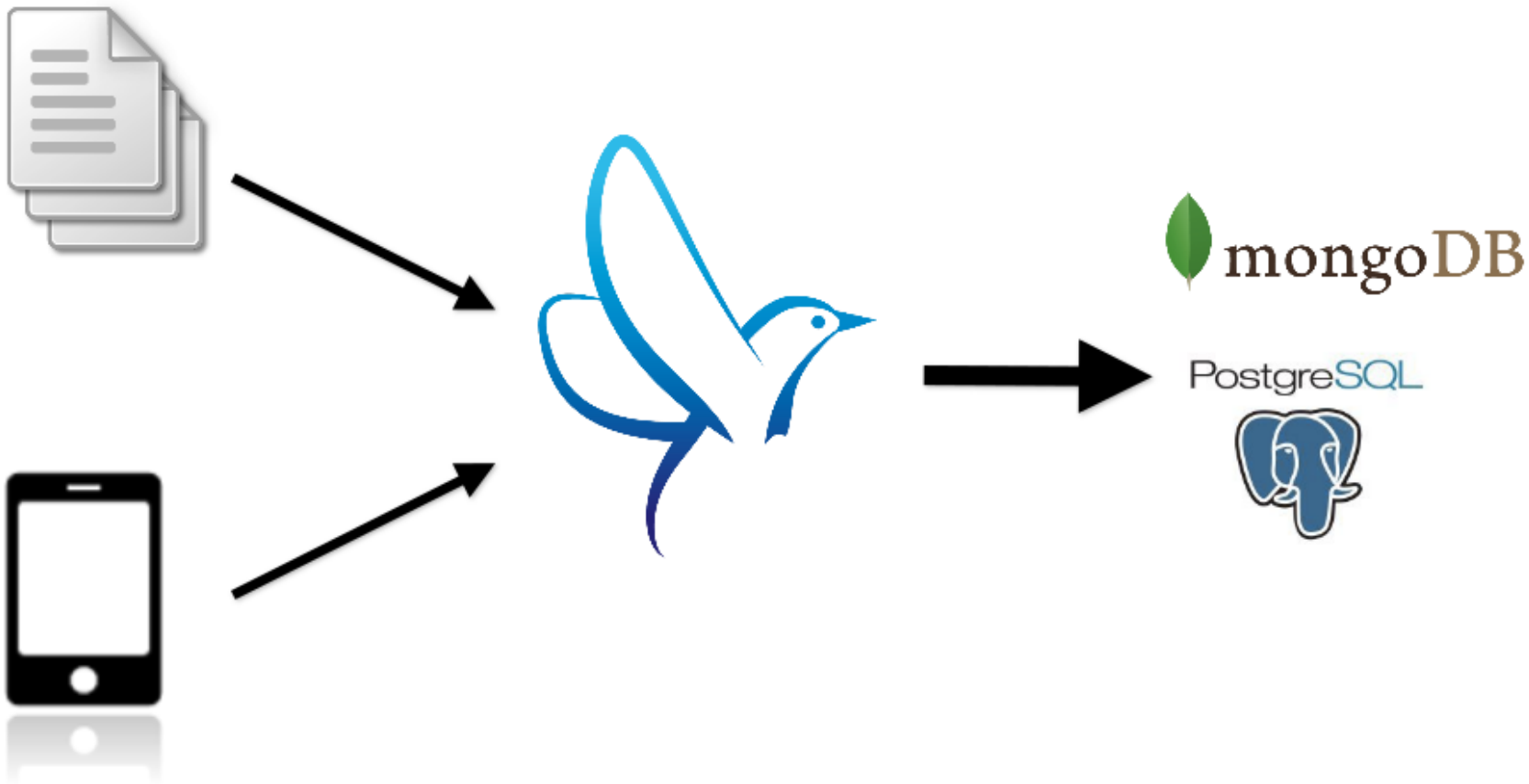
Hadoop

Archiving

Amazon S3

Fluentd

Simple Forwarding



Fluentd

Simple Forwarding: configuration

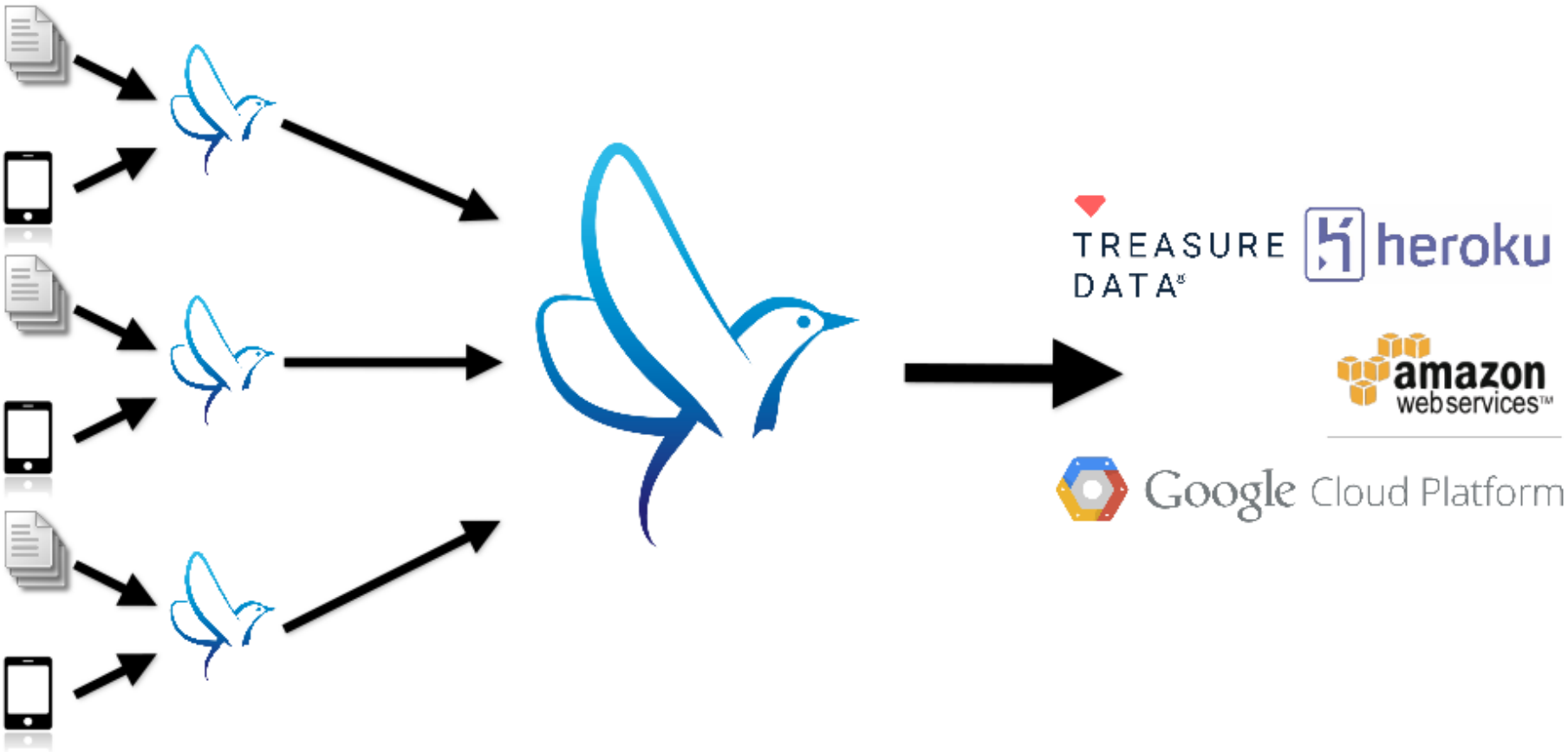
```
# logs from a file
<source>
  type      tail
  path      /var/log/httpd.log
  format    apache2
  tag       backend.apache
</source>

# logs from client libraries
<source>
  type    forward
  port    24224
</source>
```

```
# store logs to MongoDB
<match backend.*>
  type        mongo
  database    fluent
  collection  test
</match>
```

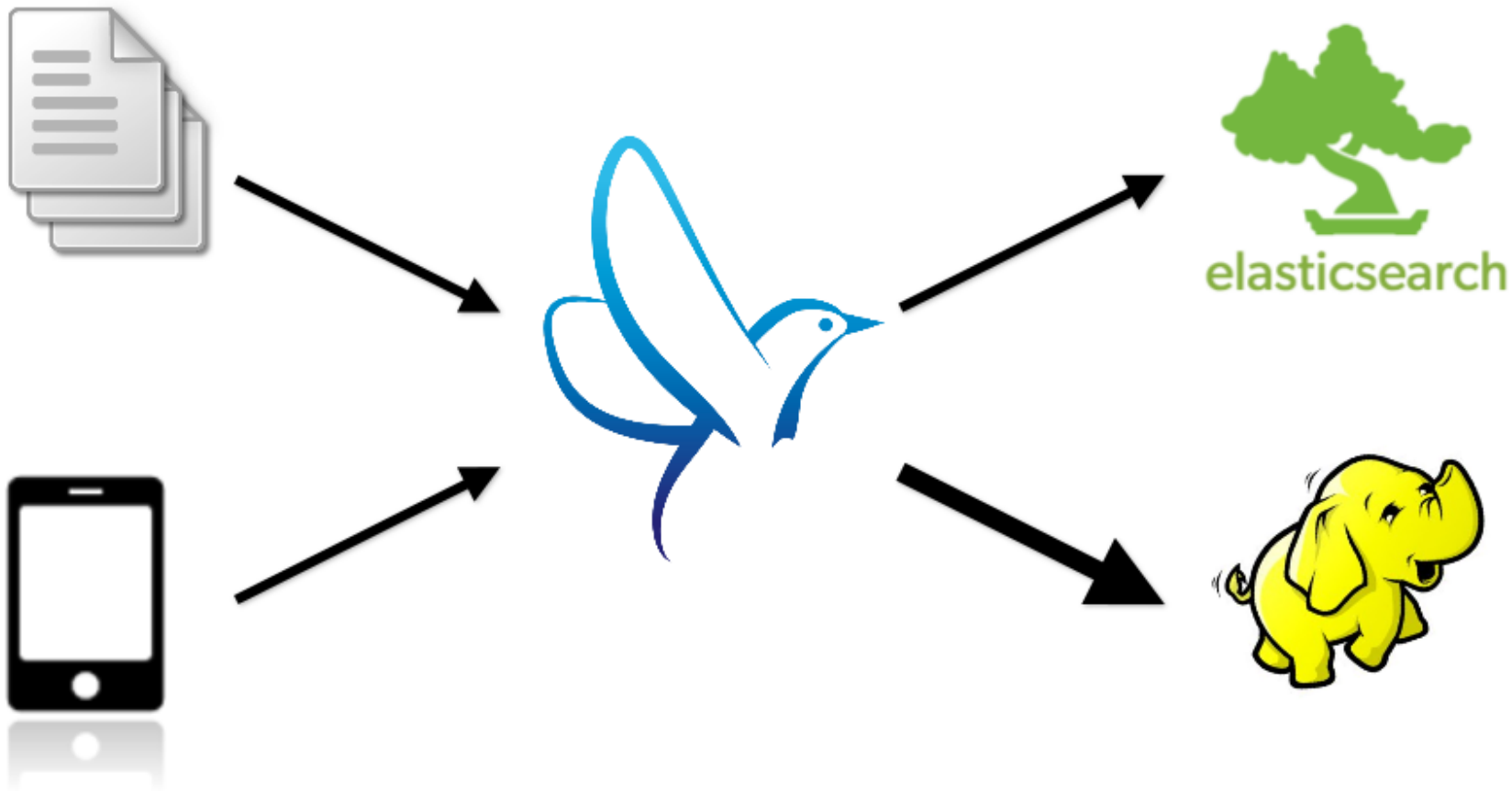
Fluentd

Less Simple Forwarding



Fluentd

Lambda Architecture



Fluentd

```
# logs from a file
<source>
  type      tail
  path      /var/log/httpd.log
  format  apache2
  tag       backend.apache
</source>

# logs from client libraries
<source>
  type     forward
  port     24224
</source>
```

```
# store logs to MongoDB
<match *.*>
  type          copy
  <store>
    type         elasticsearch
    logstash_format  true
  </store>

  <store>
    type         webhdfs
    host         192.x.y.z
    port         50070
    path         /path/to/hdfs
  </store>
</match>
```

Who uses Fluentd in production ?


LINE

PPTV
每个人的网络电视

slideshare
Present Yourself

Nintendo

GREE

BANDAI
NAMCO
BANDAI NAMCO Studios

viki

amazon
webservices™

BACKPLANE

DeNA

DRECOM
with entertainment

ContextLogic

kubernetes

TREASURE
DATA

We collect

1M events per second !

Internet of Things

Internet of Things

Facts

- IoT will grow to many **billions** of devices over the next decade.
- Now it's about **device** to **device** connectivity.
- Different **frameworks** and **protocols** are emerging.
- It needs **Logging**.

Internet of Things

Alliances

> Vendors formed alliances to join forces and develop generic software layers for their products:

OPEN INTERCONNECT CONSORTIUM℠

CISCO · GE Software · SIEMENS · SAMSUNG · IBM · hp · intel

ALLSEEN ALLIANCE

Canon · LG Life's Good · Panasonic · SONY · SHARP · Microsoft

Internet of Things

Solutions provided

Alliance		Framework
OPEN INTERCONNECT CONSORTIUM℠	→	IoTivity
ALLSEEN ALLIANCE	→	AllJoyn™

IoT & Big Data

Analytics

IoT requires a **generic solution** to collect events and data from different sources for further analysis.

Data can come from a specific framework, radio device, sensor or other. How do we collect and **unify** data **properly** ?

@fluentbit

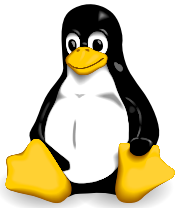
Fluent Bit is an open source data collector

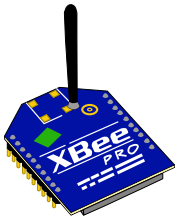
> It let's you collect data from IoT/Embedded devices and transport It to third party services.

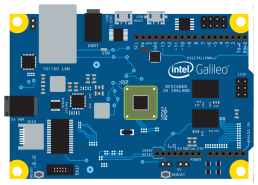
Fluent Bit

Targets

- Services
- Sensors / Signals / Radios
- Operating System information
- Automotive / Telematics







Fluent Bit

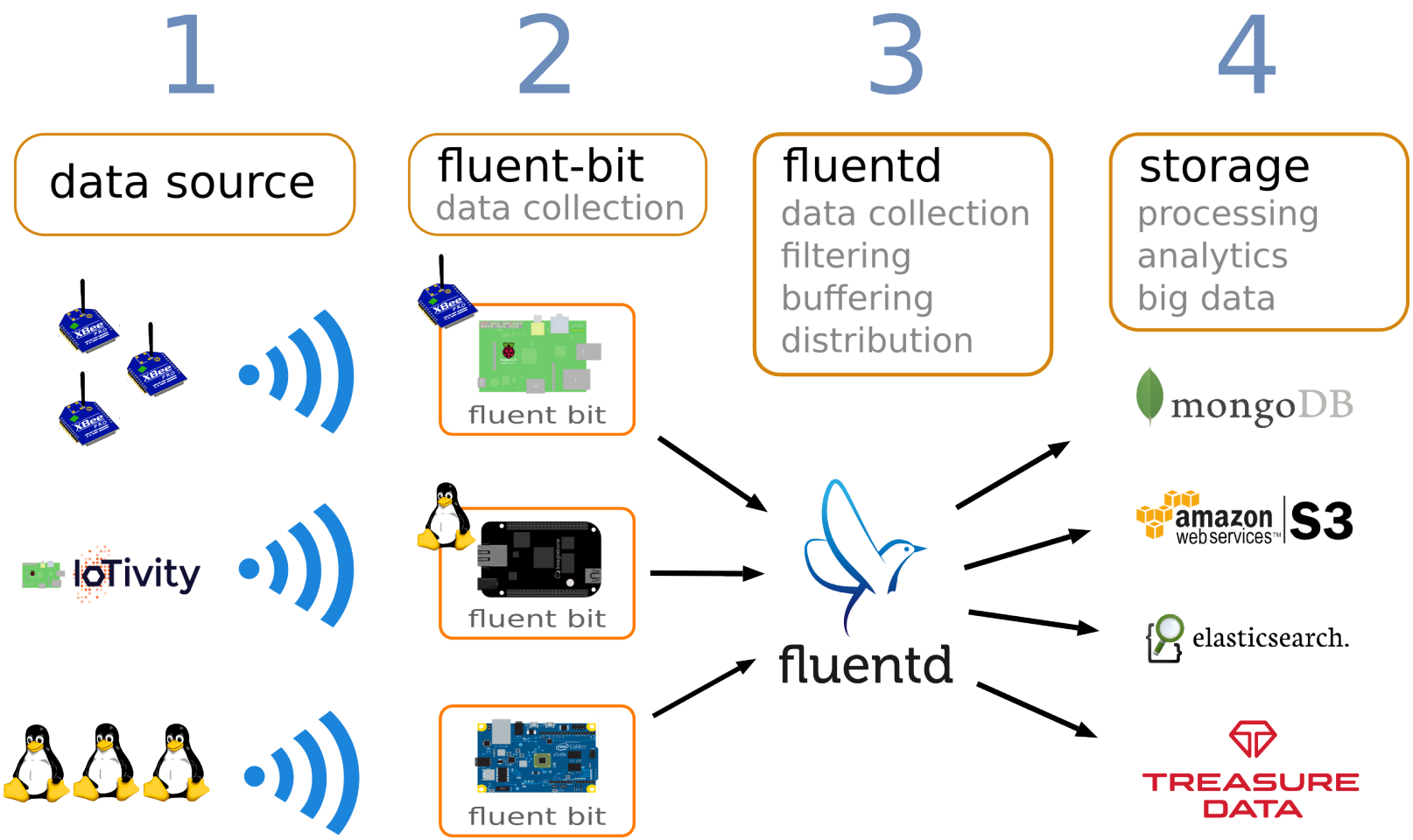
Requirements

IoT and Embedded environment requires special handling, specifically on performance and resource utilization:

- Lightweight
- Written in **C** Language
- Customizable, **pluggable** architecture
- Full integration with **Fluentd**

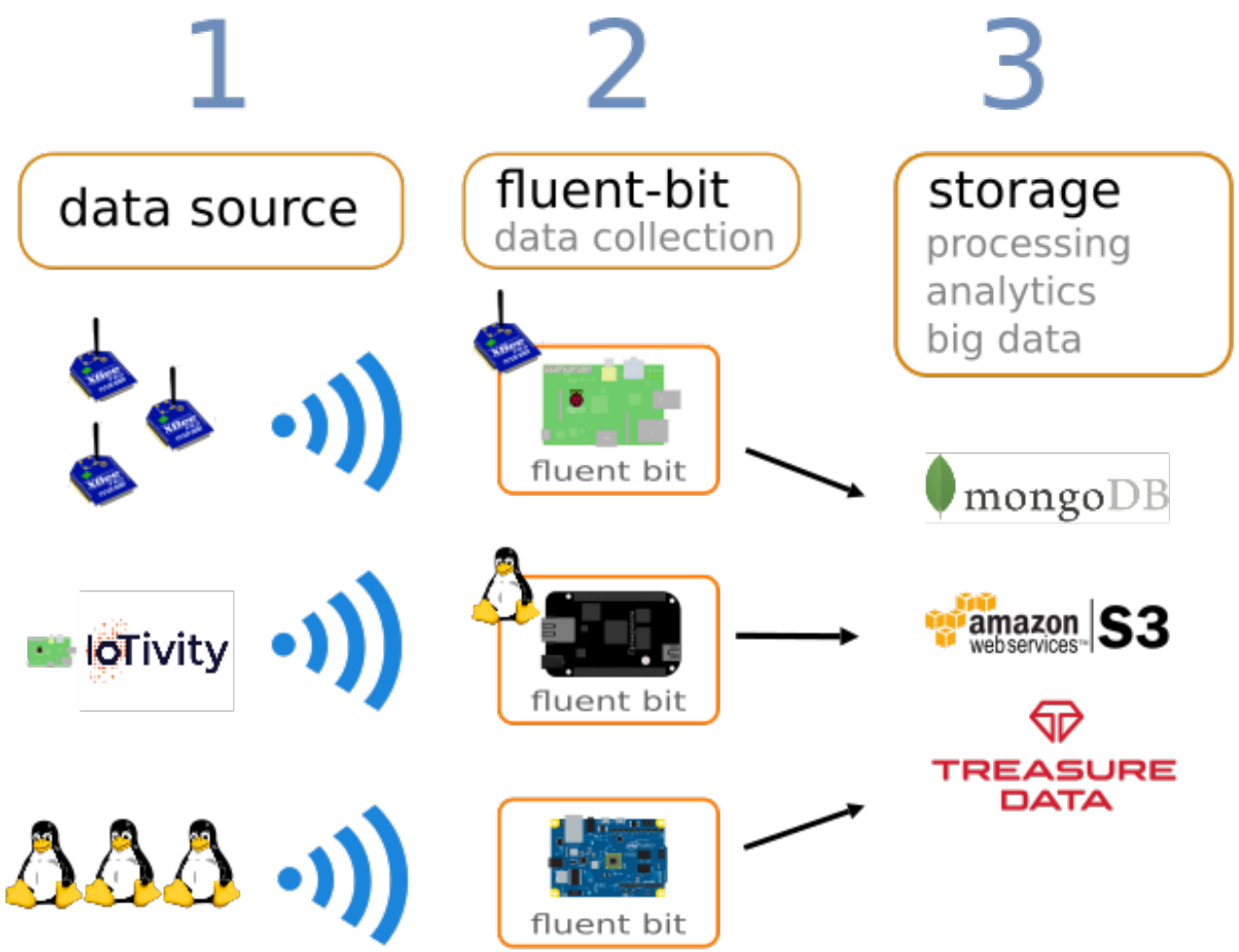
Fluent Bit

Integration



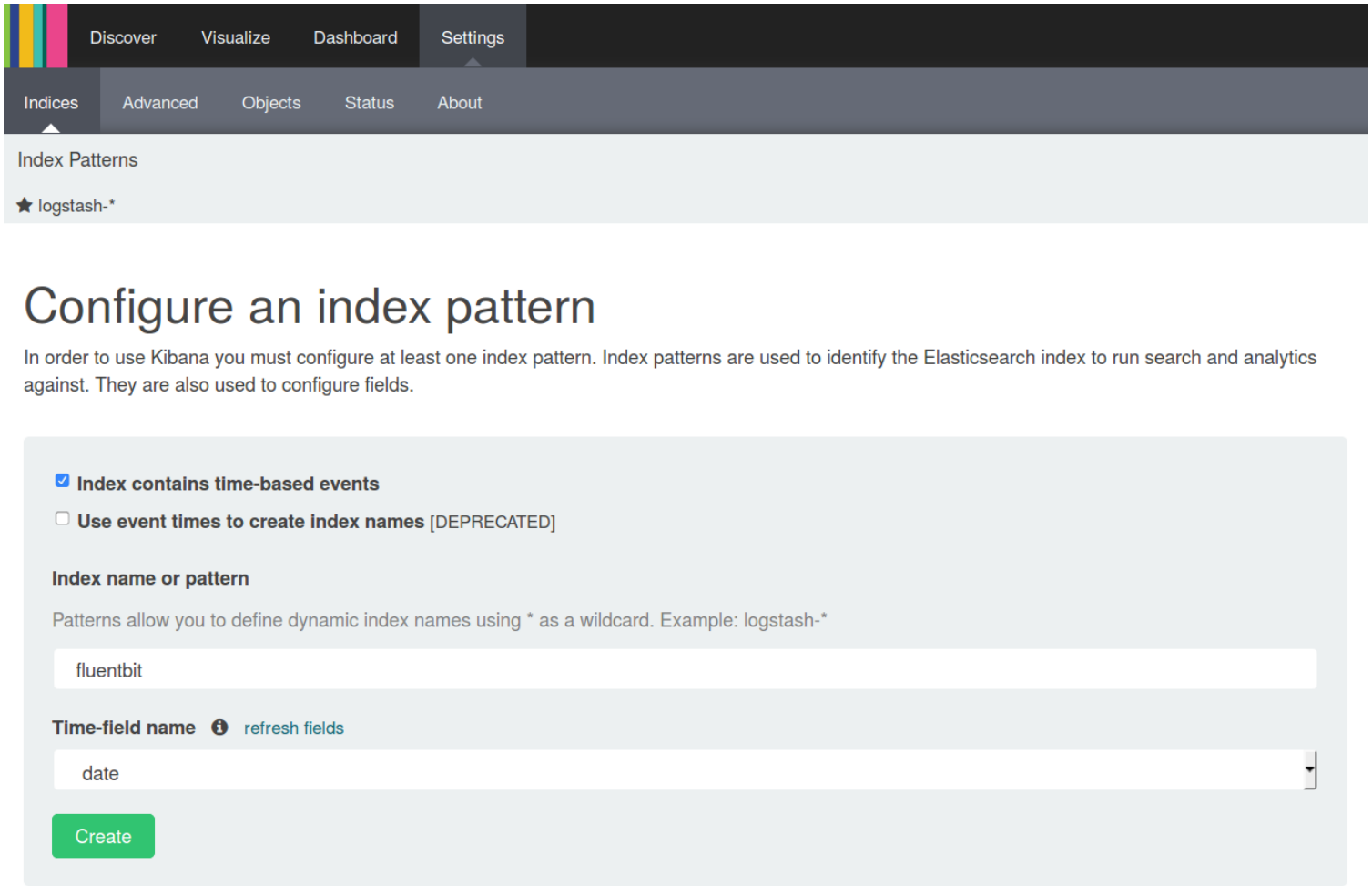
Fluent Bit

Direct Output



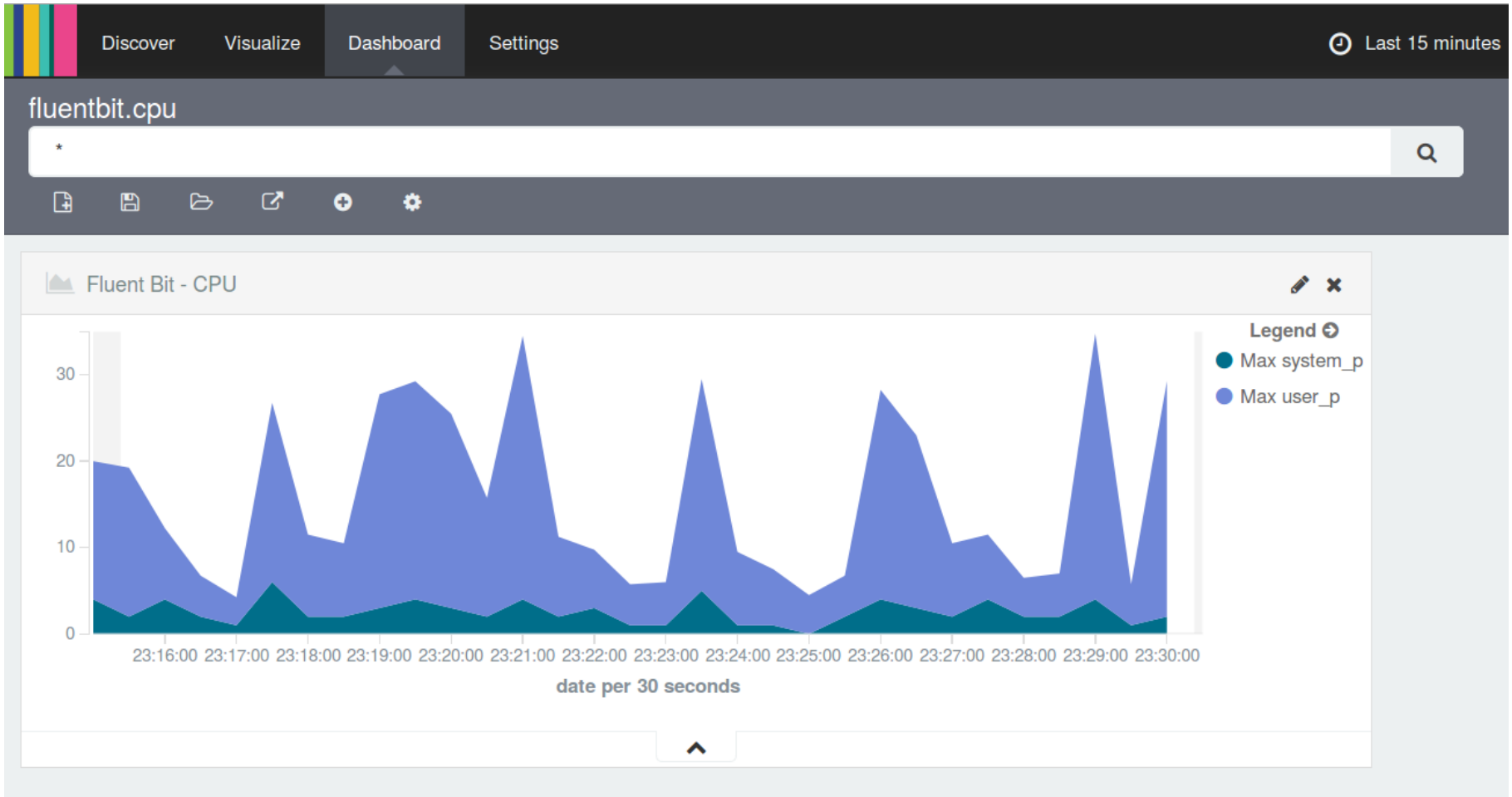
Fluent Bit

Elastic Search support

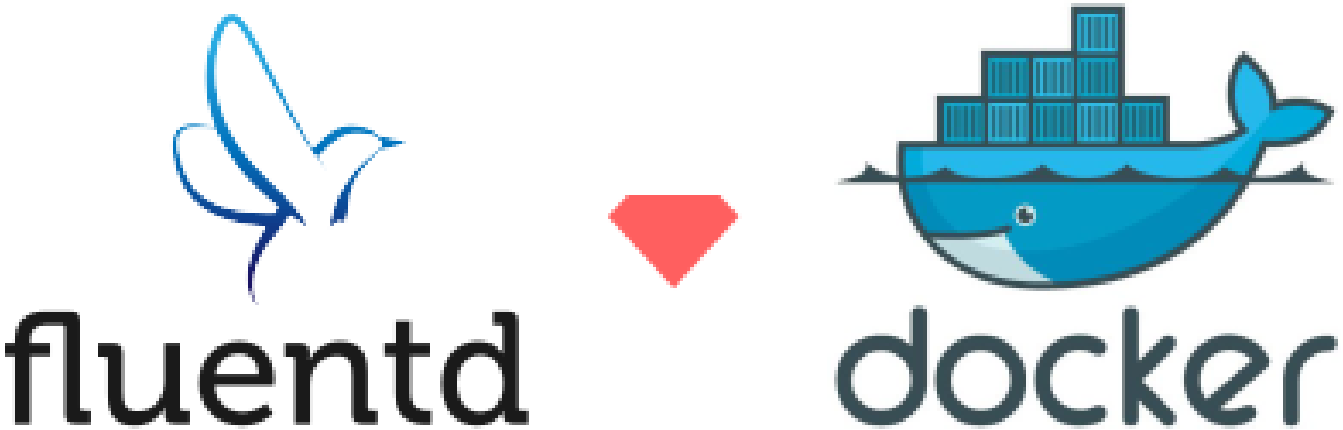


Fluent Bit

Elastic Search: Dashboard



Containers


fluentd
docker

Docker

Logging driver

- Docker v1.6 released the concept of logging drivers
- Route container output
- Fluentd ?

Docker

docker / docker

Watch 2,081 | Star 23,470 | Fork 5,853

Add new Logging driver "fluentd" #12876

Merged **LK4D4** merged 2 commits into `docker:master` from `tagomoris:logger-driver-fluentd` 26 days ago

Conversation 167 | Commits 2 | Files changed 33 | +7,763 −3

tagomoris commented on Apr 29

This patch provides `fluentd` logging plugin for docker, which is mentioned at #12540.

How to confirm behavior of this patch/new logging driver.

1. Setup dev docker container
2. Build repository with this patch by `make BINDDIR=. binary`
3. Copy binary and execute it in debug mode `docker -dD`
4. Attach from external environment by `docker exec -it kickass_heisenberg /bin/bash`
5. Build/execute another docker container from other console, built on Dockerfile and configuration file below:

```
# sh
```

Labels

status/4-merge

Milestone

No milestone

Assignee

No one assigned

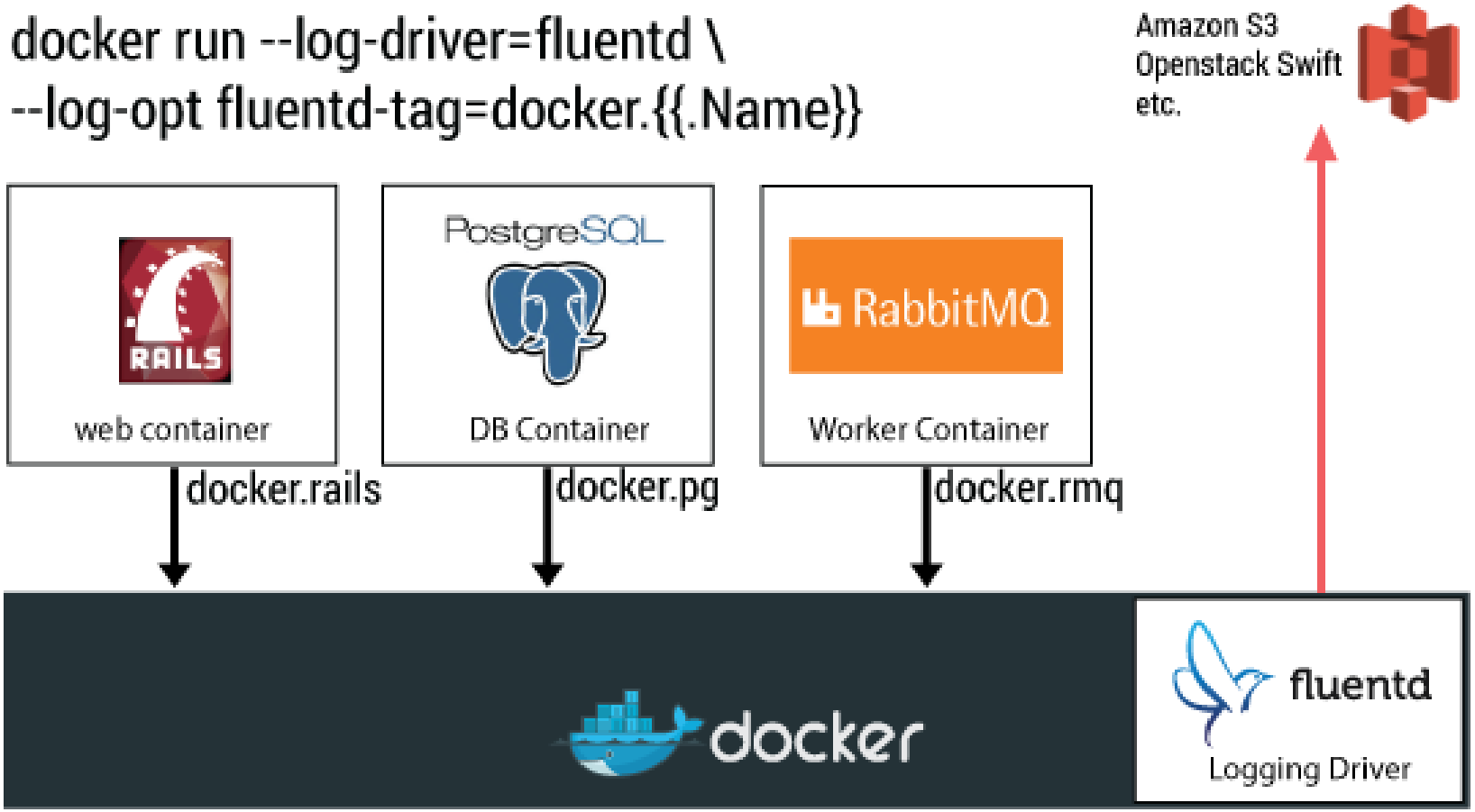
Notifications

Subscribe

Docker v1.8

Fluentd Logging driver!

```
docker run --log-driver=fluentd \
--log-opt fluentd-tag=docker.{{.Name}}
```



Docker

Data Stream

tag	time	source	container_id	container_name	log message
docker.3fd...	1441402468	stdout	3fd8678d487e...	/angry_kamam	Hello world!

docker.3fd...	1441402468	stdout
3fd8678d4...	/angry_kam	message

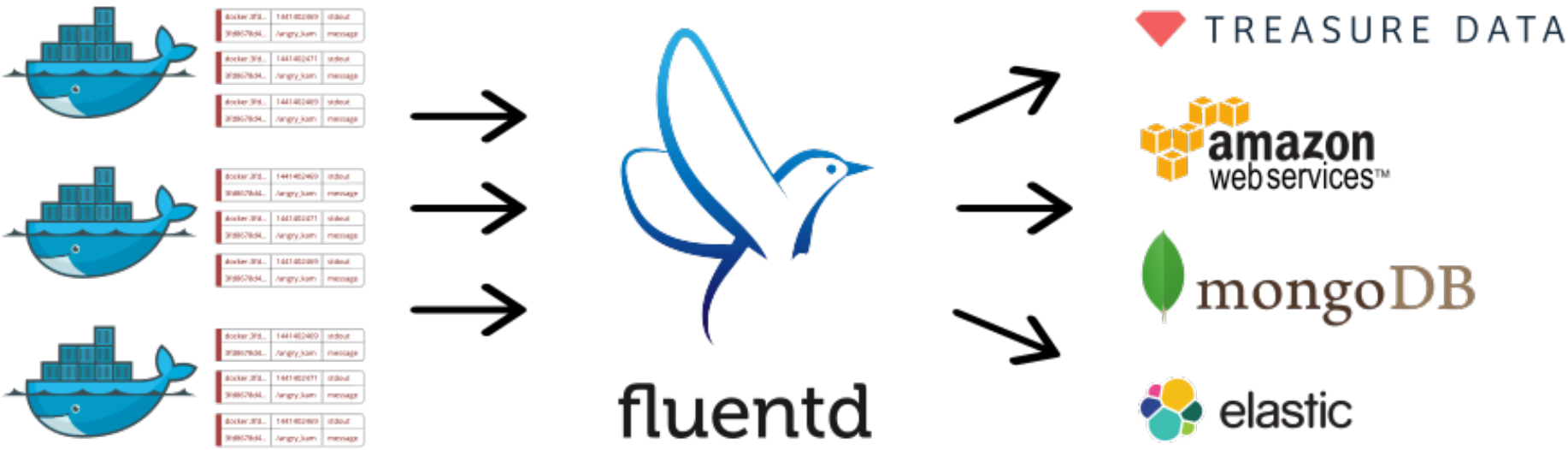
docker.3fd...	1441402469	stdout
3fd8678d4...	/angry_kam	message

docker.3fd...	1441402470	stdout
3fd8678d4...	/angry_kam	message

docker.3fd...	1441402471	stdout
3fd8678d4...	/angry_kam	message

Docker

Data Stream




fluentd


node
JS

NodeJS

Fluent-Logger (NPM)

```
var logger = require('fluent-logger')

logger.configure('tag', {
  host: 'localhost',
  port: 24224,
  timeout: 3.0
});

logger.emit('label', {record: 'this is a log'});
```

We Love Data!

TREASURE DATA

- http://fluentd.org
- http://fluentbit.io
- https://docs.docker.com/reference/logging/fluentd/
- http://github.com/fluent/fluentd

Thank you!