

# Cooperative works for Fluentd Community

Kenji Okimoto  
Hiroshi Hatake

*ClearCode Inc.*

*2018-02-15*

# Summary

- ✓ kafka-connect-fluentd and its benckmarking result
- ✓ maintaining fluentd plugins

# kafka-connect-fluentd?



- ✓ Fluentd Forward Protocol Server implemented with Kafka Connect API
- ✓ [kakfa-connect-fluentd](#)

# kafka-connect-fluentd?

- ✓ Use [inluent](#) for server implementation
  - ✓ Partial SSL/TLS support
- ✓ Aim to replace output plugins in [fluent-plugin-kafka](#)

# Kafka Connect?

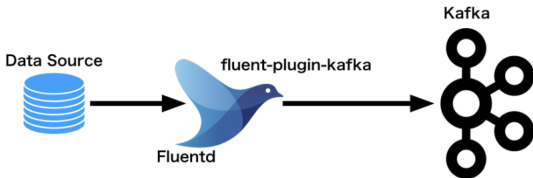
*Kafka Connect is a framework included in Apache Kafka that integrates Kafka with other systems. Its purpose is to make it easy to add new systems to your scalable and secure stream data pipelines.*

[「Kafka Connect」より引用]

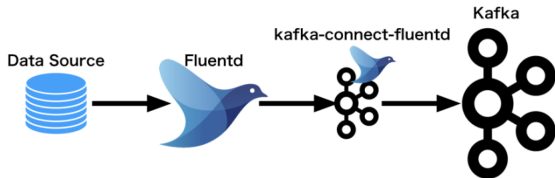
# Kafka Connect?

- ✓ Kafka Connector Source -> Kafka Producer
  - ✓ I'll talk about this implementation.
- ✓ Kafka Connector Sink -> Kafka Consumer
  - ✓ This implementation also exists.

# fluent-plugin-kafka



# kafka-connect-fluentd





# Difference

	kafka-connect-fluentd	fluent-plugin-kafka
language	Java	Ruby
Run on	Kafka side	Fluentd side

# Requirements of benchmarking

- ✓ Easy to use benchmark test tool
- ✓ Reproducible
- ✓ Compare performance with same basis
  - ✓ Easy to confirm results

# Easy to use benchmarking test tool

- ✓ There is no tool which is easy to use from command line.
  - ✓ Dummer + in\_tail exist, but it is hard to apply high load.

# Crated benchmark test tool

- ✓ [fluent-benchmark-client](#)
  - ✓ Implemented by Kotlin language
  - ✓ Sending log relies on [fluency](#)
    - ✓ SSL/TLS is not supported for now.

# Reproducible

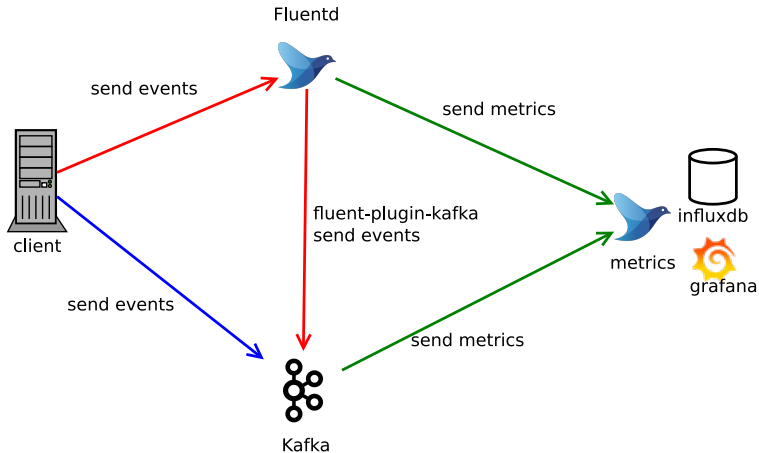
- ✓ Built with Terraform + Ansible
- ✓ Developing specific branch on [fluentd-benchmark](#)
  - ✓ Need to tidy up implementation

# Compare performance with same basis

## ✓ [kafka-fluent-metrics-reporter](#)

- ✓ Kafka plugin which sends Kafka metrics into Fluentd
- ✓ Write a script which parses result of pidstat and sends into Fluentd
  - ✓ To measure CPU and memory usage

# Benchmark environment



GCP n1-standard-2 (vCPUx2, memory 7.5GB)

# out\_kafka CPU usage 10000 events/sec



out\_kafka CPU usage is 40-60%



# kafka-connect-fluentd

## CPU usage

### 10000 events/sec



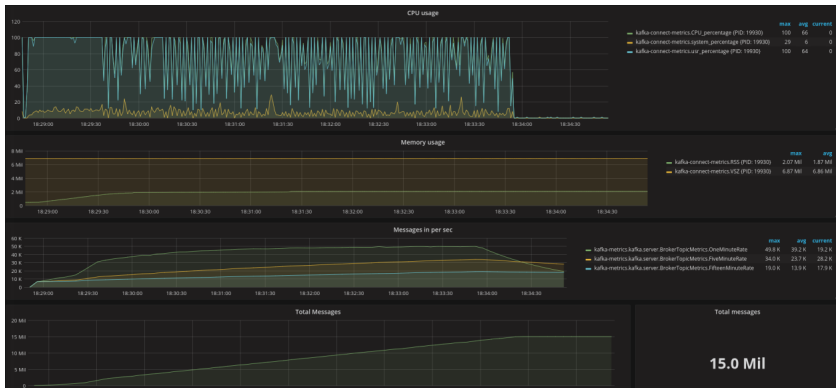
kafka-connect-fluentd CPU usage is less than 20%

# out\_kafka\_buffered 30000 events/sec



out\_kafka\_buffered cannot process 30k events/sec

# kafka-connect-fluentd 50000 events/sec



kafka-connect-fluentd can process about 50k events/sec

# Benchmark results

- ✓ kafka-connect-fluentd can handle more 50,000 events/sec
- ✓ output plugin of fluent-plugin-kafka can handle 10,000 events/sec
  - ✓ cannot handle 30,000 events/sec in this environment

# Fluentd Community
















## List of All Plugins

This page gets updated periodically to tabulate all the Fluentd plugins listed on Rubygems. [Go here](#) to browse the plugins by category.

[Input/Output plugin](#) | [Filter plugin](#) | [Parser plugin](#) | [Formatter plugin](#) | [Obscured plugin](#)

### Input / Output plugins:

Collect events from sources or send events to destinations

Certified	Download	Name	Author	About	Version
	1862841	<a href="#">elasticsearch</a>	dega_jr	Elasticsearch output plugin for Fluent event collector	2.0.0
	1253480	<a href="#">secure-forward</a>	TAGOMORI Sasaki	Message forwarding over SSL, with authentication	0.4.5
	952710	<a href="#">fluent</a>	TAGOMORI Sasaki	create sub-plugin dynamically per tag, with template configuration and parameters	0.3.3
	1000000	<a href="#">fluent-event-receiver</a>	fluent-event-receiver	Fluent event record receiver	0.9.1
	1000000	<a href="#">fluent-kafka</a>	fluent-kafka	Fluent Kafka collector	1.1.1
	1000000	<a href="#">fluent-telegram</a>	fluent-telegram	Fluent telegram collector	0.3.4
	1000000	<a href="#">fluent-journald</a>	fluent-journald	This is a fluent input plugin, it reads logs from the systemd journal.	0.3.1
	466301	<a href="#">kinesis</a>	Arc42/Kit Services	Fluentd output plugin that sends events to Amazon Kinesis.	2.1.0
	373795	<a href="#">bigquery</a>	Naya M. pear307	Fluentd plugin to store data on Google BigQuery, by load, or by stream inserts	1.2.0
	330992	<a href="#">kafka</a>	Hidenori Taguchi, Masahiro Nakagawa	Fluentd plugin for Apache Kafka > 0.8	0.6.6
	330992	<a href="#">kafka</a>	Hidenori Taguchi, Masahiro Nakagawa	Fluentd plugin for Apache Kafka > 0.8	0.6.6
	302612	<a href="#">ec2-metadata</a>	SANAMOTO Takuro	Fluentd output plugin to add ec2 metadata fields to a event record	0.1.1
	261896	<a href="#">flowcounter</a>	TAGOMORI Sasaki	Plugin to counts messages/bytes that matches, per minutes/hours/days	1.0.0
	249617	<a href="#">dstat</a>	Shinichi Miura	Dstat input plugin for Fluent event collector	0.3.9
	197300	<a href="#">gelf-ss</a>	Alex Yamauchi, Eric Beatty	Buffered fluentd output plugin to GELF (Deylog!).	1.0.4

Fluentd community has lots of plugins  
Over 700+ plugins

# Fluentd Community

- ✓ Fluentd community has...
  - ✓ lots of up-to-date plugins
  - ✓ lots of outdated plugins

# Fluentd Community

- ✓ Fluentd community has...
  - ✓ lots of up-to-date plugins
  - ✓ lots of **outdated** plugins

# How does handle outdated plugins?

- ✓ Send PRs
- ✓ Report issues
- ✓ Become a maintainer



# What means **outdated**?

- ✓ out-of-date dependencies
  - ✓ c.f. client libraries
- ✓ out-of-date class inheritance
  - ✓ `Fluent::Input`, `Fluent::HttpInput` etc.

# What means **outdated**?

- ✓ Using old API
  - ✓ Gap between v0.12 and v1.0 API
- ✓ **Fluent::Engine.emit** directly
- ✓ Lack of requires etc.

# Maintaining Plugins

- ✓ Normally, same as other rubygems

# Maintaining Plugins

- ✓ Report issues
- ✓ Send PRs

# Maintaining Plugins

- ✓ 130+ PRs still opened...
- ✓ 1400+ PRs merged
  - ✓ in 2.5 years

# Maintaining Plugins

- ✓ But, sometimes plugin authors are busy....

# Let's take over Fluentd plugins project

- ✓ Ask plugin authors
  - ✓ to add project collaborator
  - ✓ to add gem owner
- ✓ Create a cooperative working organization

# Become project collaborator

- ✓ Maintaining only
  - ✓ takus/fluent-plugin-ec2-metadata
  - ✓ y-ken/fluent-plugin-anonymizer



# Become project collaborator and gem owner

- ✓ Still exists in original place, but maintaining ourselves
  - ✓ uken/fluent-plugin-elasticsearch
  - ✓ y-ken/fluent-plugin-geoip
  - ✓ fluent/fluent-plugin-rewrite-tag-filter

# Create cooperative working organization

- ✓ Take over outdated fluentd plugin project
- ✓ New plugins which will be widely used
  - ✓ c.f. fluent-plugin-concat
- ✓ And some collaborations

# Create cooperative working organization

✓ [fluent-plugins-nursery](#)

# fluent-plugins-nursery

- ✓ For Fluentd plugins that are not maintained by original authors
- ✓ Aim to provide a sustainable maintenance system for Fluentd community

# fluent-plugins-nursery

- ✓ Almost plugins are taken over maintaining from original author.
  - ✓ fluent-plugin-map
  - ✓ fluent-plugin-redis
  - ✓ fluent-plugin-irc etc.

# fluent-plugins-nursery

- ✓ Let's maintain fluentd plugins by community!