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?

ClearCode

✔ Use influent 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.*
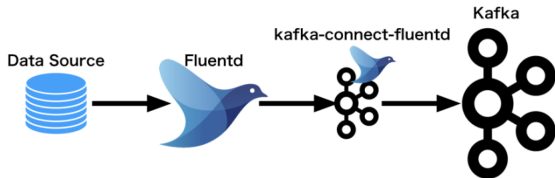>
> *[「Kafka Connect」 より引用]*

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
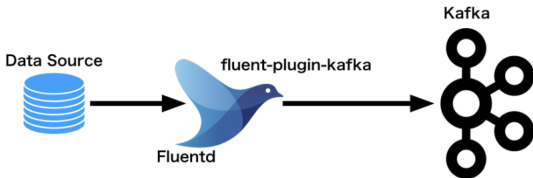
kafka-connect-fluentd

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

✓ 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

ClearCode

Fluentd

send events

send metrics

client

fluent-plugin-kafka
send events

metrics
influxdb

grafana

send events

send metrics
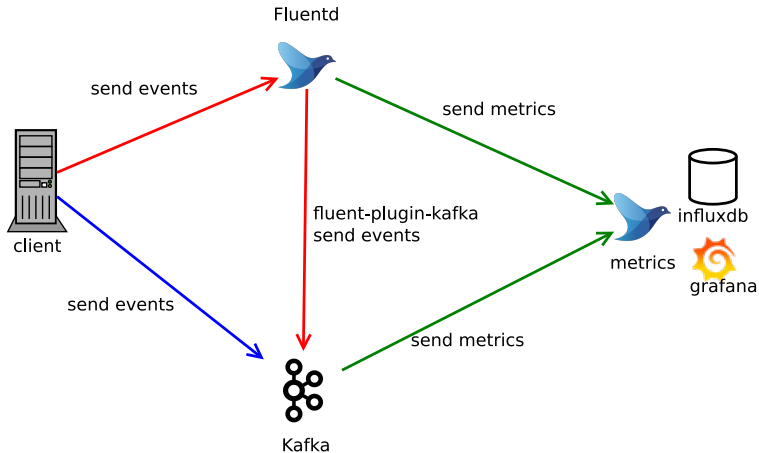
Kafka

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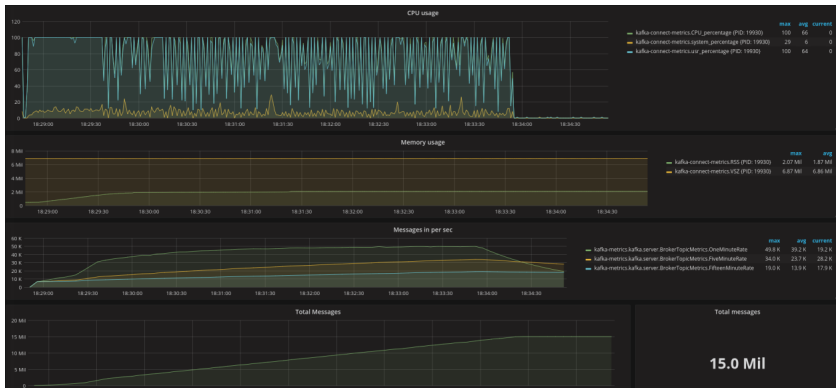
kafka-connect-fluentd can process about 50k events/sec

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

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

ClearCode

✓ 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

ClearCode

✓ Let's maintain fluentd plugins by community!