



fluentd

The Open Source Data Collector

The Intelligent Data Collector: Acquire, Filter and Process Data Streams in Real-Time.

What's Fluentd?

Fluentd is the leading open-source tool to collect and process data streams. Fluentd will help you collect and manage the exploding volume and complexity of data we see today.

Fluentd is an open source data collector designed for processing high-volume data streams. Once installed on a server, it runs in the background to collect, parse, transform, analyze and store various types of data. Fluentd provides a cost-effective, lightweight solution to:

- Collect data from various data sources and store them into various storage systems.
- Build a flexible and adaptable data processing pipeline that can filter and aggregate data in real-time.

With Fluentd, you can create a lightweight yet powerful data processing platform for a wide range of data streams without the vendor lock-in and hefty cost of proprietary software. Since open sourced as an Apache 2.0 License project in 2011, Fluentd has become an indispensable piece of data infrastructure at many companies.

Featured Users



Technical Overview

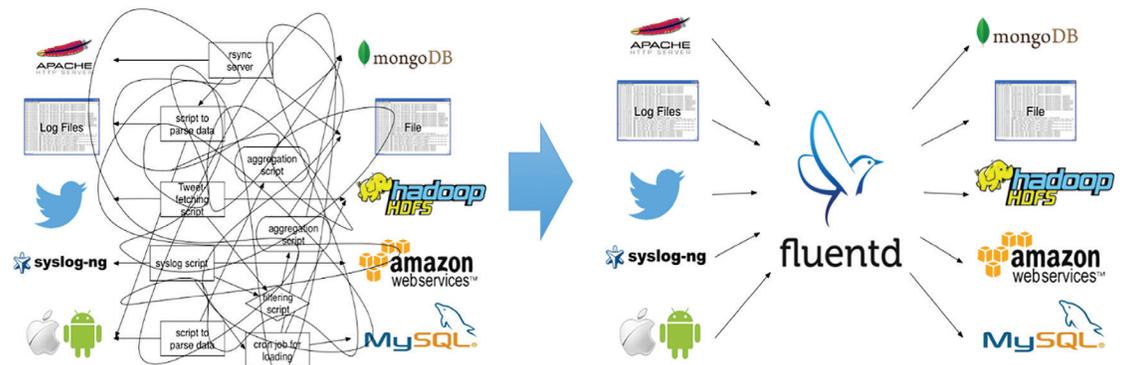
Fluentd's three key characteristics

1. **Easy installation:** small (< 10MB) and distributed as rubygems, rpm and deb packages.
2. **Tag-based routing:** simplify and scale data pipeline management
3. **Unlimited Connectivity:** take advantage of 200+ community-contributed plugins.

“Fluentd is easier to install and maintain and has better documentation and support than Flume and Scribe”

Best Practices for Amazon EMR AWS
Technical Whitepaper

Have more tech questions?
Check out our documentation
at <http://docs.fluentd.org>.



“Treasure Data’s data collector, **td-agent** (open-sourced as **Fluentd**), greatly simplified our real-time data collection.”

Wiki.com Huy Nguyen, Software Engineer

Flexible log transformation, forwarding and search

Unlike simple log forwarders, Fluentd is used for much more than simple data transfer or text search. For example, Fluentd is used to:

- **Store data in multiple systems** for many purposes, including flexible analytics, archiving and full-text search). Common targets include MongoDB, Amazon S3, Elasticsearch and many more.
- **Collect and correlate web server access logs** and application error logs to help understand whether server issues stem from the web server itself or an application-level bug.
- **Build** a simple complex event processing system to support exploratory data analysis across the organization.
- **Develop a real-time monitoring system** that, upon detecting a problem in the system, pages the on-call engineer via Twilio and sends out an incident report email to the team.
- **Create a cost-effective, more flexible alternative** for web analytics tools by turning Fluentd into a customizable REST API endpoint.

Treasure Agent vs. Fluentd

Treasure Agent is the commercially supported version of Fluentd maintained by Treasure Data. Treasure Agent is built on top of Fluentd with additional capabilities as below. Treasure Data sells and supports Treasure Agent as part of its cloud-based, managed service for data.

	Fluentd	Treasure Agent
Packaging	RubyGems	Ubuntu/RedHat
Technical Support	Community mailing list	Community mailing list Live chat and email support & Treasure Data Engineering Team (for paying customers)
OS-level Enhancement	-	Smarter memory usage via jemalloc
Deployment Tooling	-	init.d script & Chef recipe
Quality Assurance	-	Treasure Data Engineering Team
Monitoring	-	Treasure Agent Monitoring, with diagnostics dashboards in real-time with alerting.

Fluentd is Sponsored by Treasure Data

If you are interested in **Treasure Agent**, please visit TreasureData.com, contact us at info@treasuredata.com or call us 1-866-899-5386.

If you are interested in open source **Fluentd**, visit <http://fluentd.org>.

Treasure Data is one of the primary developers and maintainers of Fluentd, the open source project, and has its own commercially supported version, Treasure Agent. Treasure Data is the first end-to-end managed service in the cloud for big data analytics that is trusted and simple. By simplifying the process of data acquisition, storage and analysis, businesses get value from big data in days, not months. Easily and economically process enormous data volumes in near real-time with no infrastructure to manage. The service supports most data types, including multi-structured web, application, mobile and sensor data – and integrates seamlessly with existing data management and BI environments. Over 100 customers include Toyota, GREE, MobFox, Getjar and several F500 companies.