CNCF FLUENTD PROJECT JOURNEY REPORT

Project Description

Fluentd is an open source data collector that unifies data collection and consumption for better use and understanding. Developed by Treasure Data, Fluentd was created to solve log/data collection and distribution needs at scale, offering a comprehensive and reliable service that can be implemented in conjunction with microservices and generic cloud monitoring tools. With 650+ plugins connecting it to many data sources and data outputs, it is no wonder Fluentd was the 2016 Bossie Awards winner for the best open source datacenter and cloud software. Additionally, the project graduated in April 2019.

This report assesses the state of the Fluentd project and how CNCF has impacted its progress and growth. Without access to a multiverse to play out alternative scenarios, it is impossible to sort out causation. However, we can document correlations. This report is part of a series of project journey reports published by CNCF focused on graduated projects.

Project Snapshot

Fluentd's first commit was made on June 18, 2011. Between joining CNCF on November 9, 2016, and today, Fluentd has added:

7,452	**52.2K**	**6K**	**43,161**	**1,046**
CONTRIBUTORS	CODE COMMITS	PULL REQUESTS	CONTRIBUTIONS	CONTRIBUTING COMPANIES

*Note: These statistics were collected with the DevStats tool, which CNCF built in collaboration with CNCF project communities. "Contributor" is defined as somebody who made a review, comment, or commit, or created a PR or issue.

CNCF Premise of Open Source Software Development

A basic premise behind CNCF conferences, (including KubeCon + CloudNativeCon and PromCon), and open source in general, is that most interactions are positive-sum. There is no fixed amount of investment, mindshare, or development contributions allocated to specific projects. Just as open source development is based on the idea that, collectively, we are smarter together than any one of us alone, open source foundations work to make the entire community better. Equally important, a neutral home for a project and community fosters this type of positive-sum thinking, and drives growth and diversity that we believe are core elements of a successful open source project.

CODE DIVERSITY

## Diversity Across Company Size and Type (End User, Vendor, Foundation)

From its roots at Treasure Data, the Fluentd project has grown to incorporate meaningful code contributions from more than 1,000 organizations. High-velocity open source projects like Fluentd garner wide adoption and contribution from both vendor and end-user communities. As such, Fluentd code contributions come from a wide range of companies, fostering user-driven innovation.

Contributors to Fluentd include many of the world's largest tech companies, such as Google, Microsoft, ARM, Red Hat, and Amazon, as well as fast-growing mid-size companies like Splunk. Contributions also come from dozens of small businesses and startups, such as Clearcode and ShopKeep. The diversity of vendor contributors is also expanding; Google has grown to become the third-largest contributor to Fluentd since project inception. Contributing organizations to Fluentd are well distributed between vendors and end users, demonstrating that end-user innovation can foster and sustain fast-growing, successful projects.

By the Numbers

The top two contributing companies to Fluentd as of the end of the December 2019 reporting period were Treasure Data and Clearcode, with 31% and 13% of contributions, respectively. Treasure Data and Clearcode provided the majority of initial code contributions to the project over the first two years, but the Fluentd project has diversified to include many additional companies. The total number of companies contributing code has increased by 144% since Fluentd joined CNCF, from 429 to 1,046. As Treasure Data's percentage of all contributions has decreased, the company has continued to contribute a high volume of code even as Google and ARM have dramatically expanded their contributions. This indicates a healthy dynamic in which the project originators continue to contribute high volumes of code but encourage other organizations to contribute a greater percentage of code over time, sharing stewardship and growing the community. Another key project health indicator is the number of contributors. Fluentd has enjoyed 237% expansion of individual contributors over the four years since the project joined CNCF. During the year before joining CNCF, Fluentd accumulated 2,214 contributors. In the two years since, Fluentd has added 7,452 contributors.

Fluentd contributions by company (cumulative)
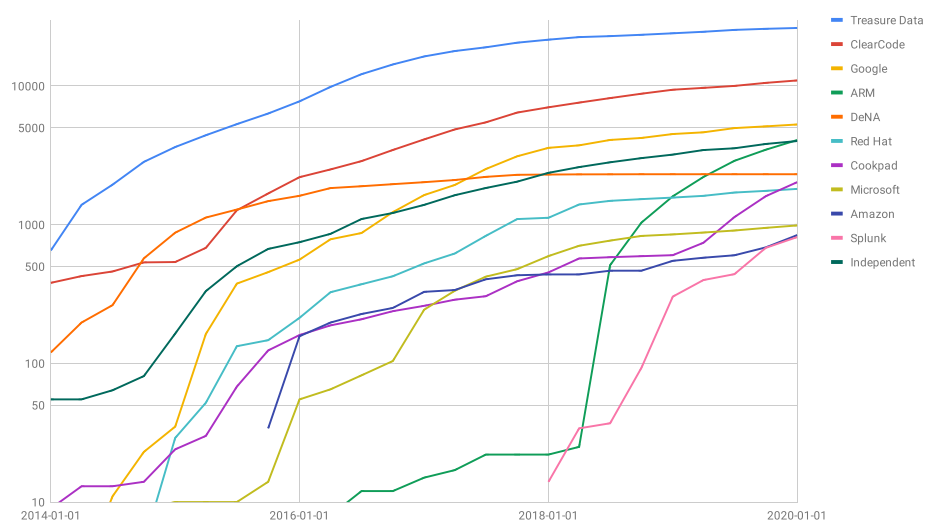
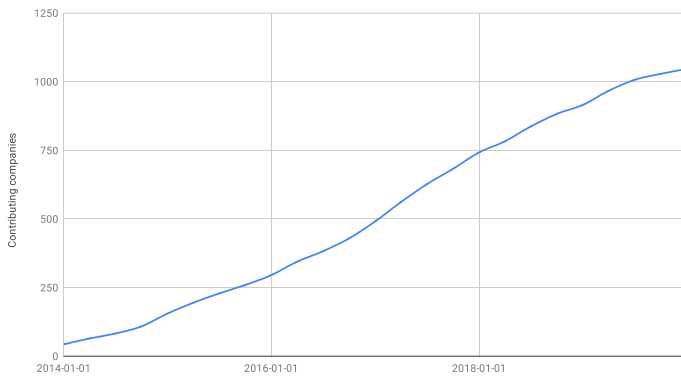Cumulative growth of contributions by company since Fluentd project launch (Q1 2014-Q1 2020)
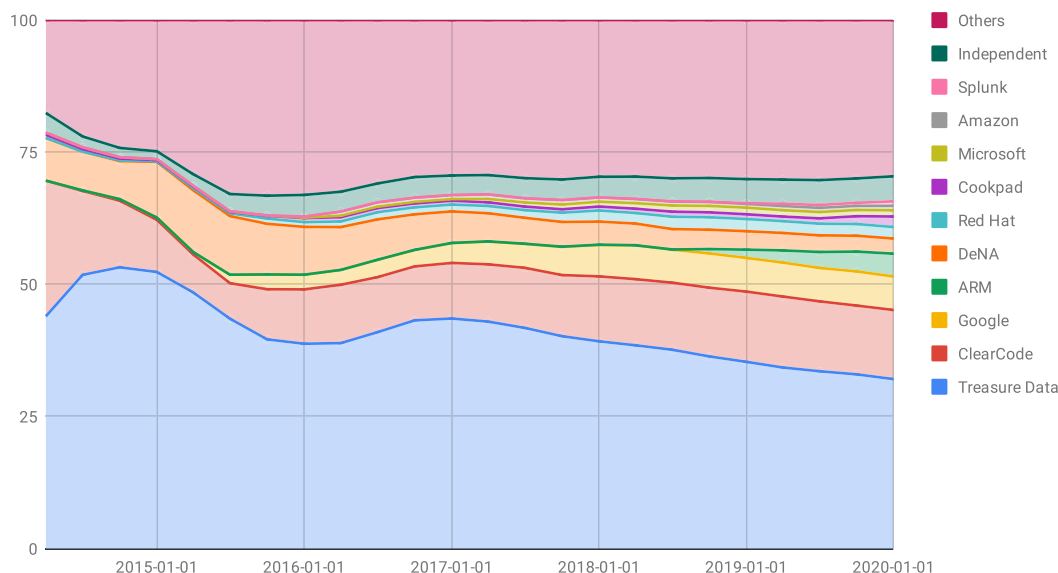
Percentage of Fluentd code contributions by company

Percentage breakdown of contributions by company since Fluentd project launch (Q1 2014-Q1 2020)
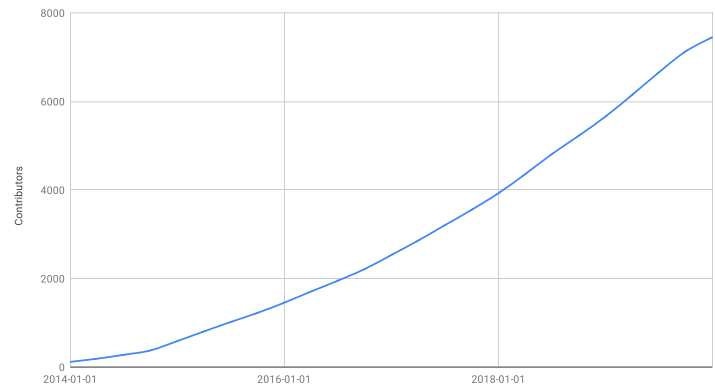
Total number of companies contributing Fluentd code

Cumulative number of companies contributing by quarter (Q1 2014-Q1 2020)

Number of contributors to Fluentd project

Cumulative growth in contributors by quarter (Q1 2014-Q1 2020)

GEOGRAPHIC DIVERSITY OF CONTRIBUTORS

Contributors to Fluentd have come from more than two dozen countries spread across five continents. The geographic diversity of contributions expanded quickly from eight countries in the first year of the project to twelve during the second year. The chart shows the percentage of contributors over time, broken down by country (based on self-reported location on GitHub).
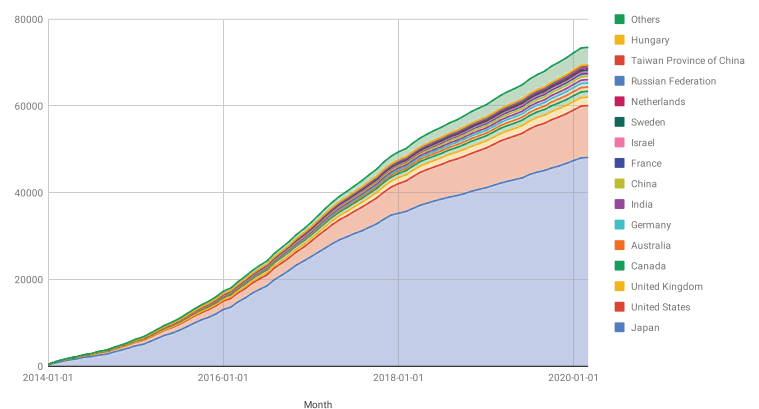

Map of project contributors by country since project launch (Q1 2014-Q1 2020)

Percentage contributors to Fluentd by country
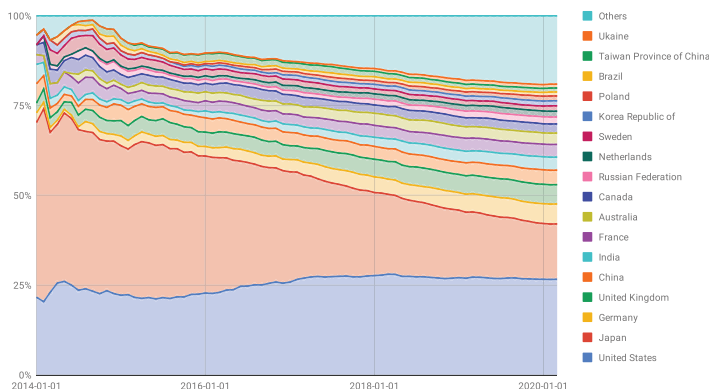
Change in number of monthly contributors by country as a percentage of total (Q1 2014-Q1 2020)

Change in number of contributors by country (Q1 2014-Q1 2020)

DEVELOPMENT VELOCITY

Among the top projects in terms of velocity, Fluentd is thriving. One way we track developer velocity is with the following formula: velocity = commits + PRs + issues+ authors. We also look at the growth of PRs, code commits, and issues filed as separate line charts. A third way to examine velocity is by looking at the cumulative number of contributors over time. These charts illustrate sharply rising velocity for Fluentd.
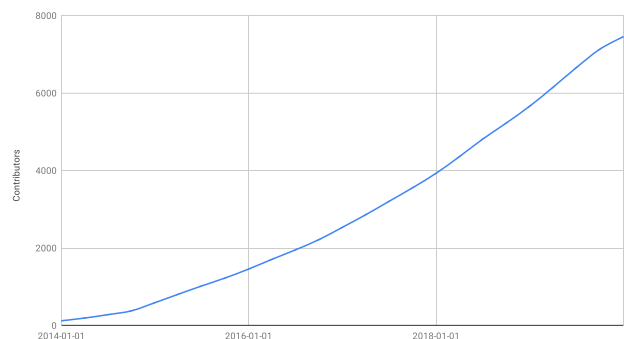
Monthly Velocity of Fluentd
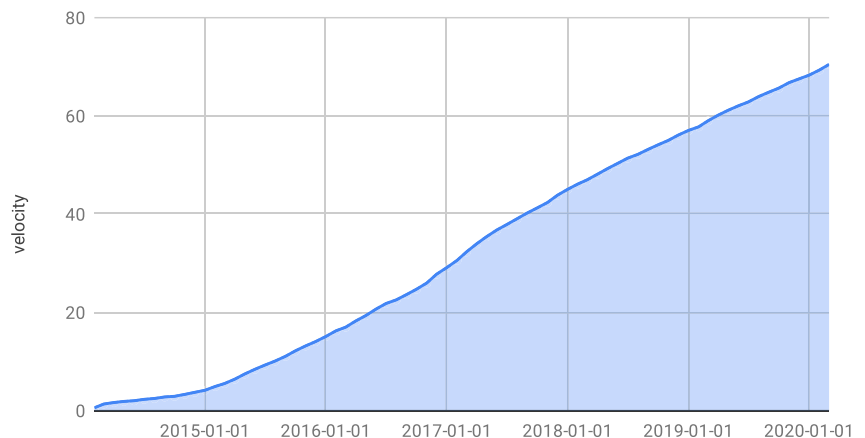
Fluentd commits, PRs, and issues
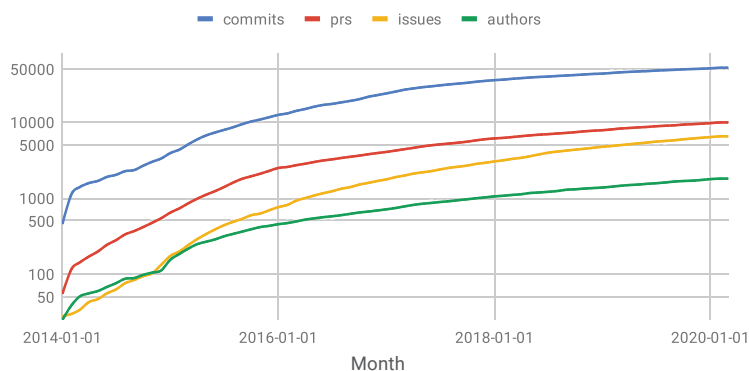
Growth of Fluentd pull requests, code commits, issues, and authors over time (Q1 2014-Q1 2020)

Number of contributors to Fluentd project

Cumulative growth of Fluentd contributors over time (Q1 2014-Q1 2020)

EDUCATION, EVENTS, AND SPONSORSHIP

Growth of community participation in education, events, and sponsorship are reliable proxies for the health of a project.

The Fluentd project actively participates in KubeCon + CloudNativeCon North America, China, and Europe through a variety of presentations and talks from community leaders. In 2019, the project hosted 12 presentations across all CNCF flagship events and one Mini Summit.

MARKETING GROWTH AND PROGRAMS

Website and Social Media

When Fluentd joined CNCF in November 2016, the foundation started promotional efforts to help sustain, nurture, and expand the Fluentd community. This includes blog posts, email newsletter mentions, and social media support. Thanks in part to these marketing efforts, public awareness of and interest in Fluentd has grown quickly. Google Analytics data for Fluentd shows an increase in pageviews since the project was contributed to CNCF, totaling more than 7M to date.

The project has nearly 4,000 followers on Twitter, having increased its following since joining CNCF.
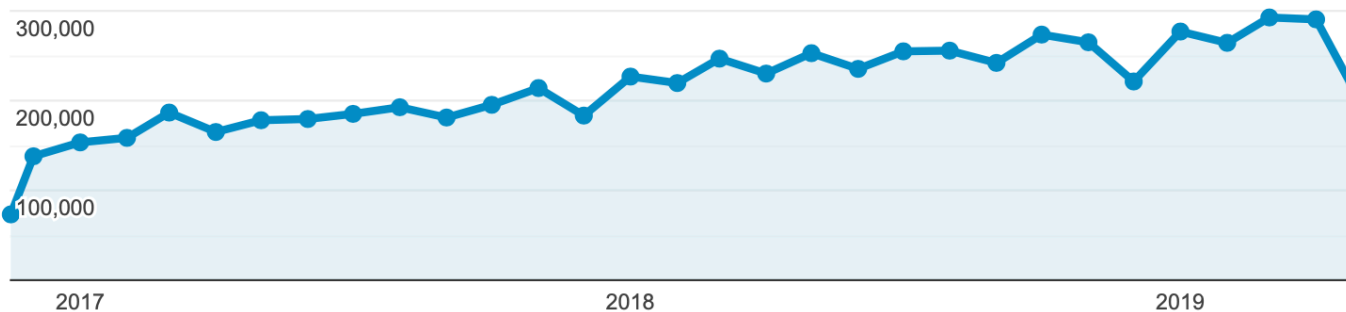
Growth in monthly pageviews from November 2016 to January 2020

PROJECT DOCUMENTATION*

Continuous additions to and improvements of project documentation are essential for the growth of any open source project. Robust documentation is critical to educating new users, and to helping existing users resolve problems and understand a project's capabilities. Fluentd documentation has rapidly expanded. Since joining CNCF, the number of authors and companies committing documentation to Fluentd has grown by 199.6% and 149.5%, respectively. As of this report, 752 authors have committed, and 302 companies are involved in committing documentation. The number of documentation commits has increased by 226% since Fluentd joined CNCF (as of the end of February 2020).
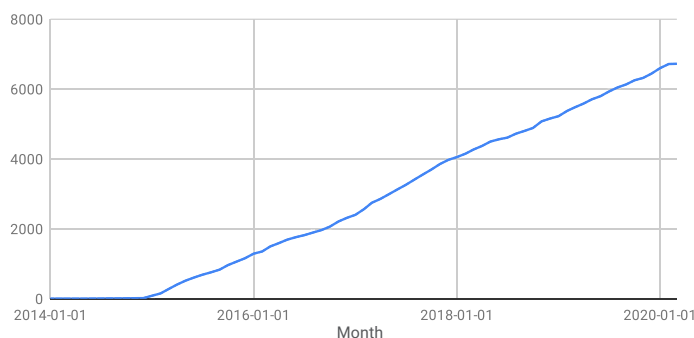
Documentation for Fluentd is collected in .md files. CNCF uses the DevStats tool to automatically collect and count statistics of all relevant .md files in the Fluentd repositories in GitHub.

Fluentd documentation committers
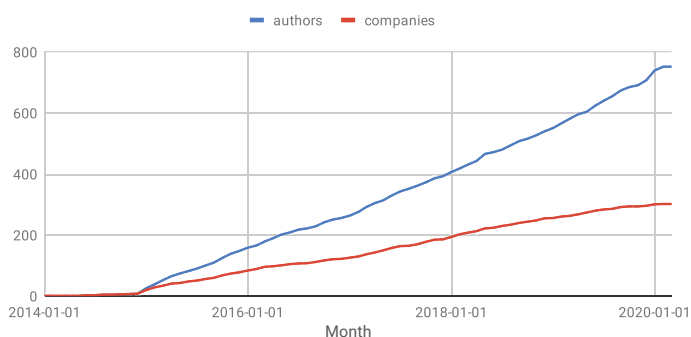
Growth in participation in Fluentd project documentation
(Q1 2014-Q1 2020)

Documentation commits

Cumulative growth of Fluentd project documentation
commits (Q1 2014-Q1 2020)

CONCLUSION

CNCF is committed to **fostering and sustaining an ecosystem** of open source, vendor-neutral projects by democratizing state-of-the-art software development and deployment patterns to make technology accessible for everyone. We hope this report provides a useful portrait of how CNCF is fostering and sustaining the growth of Fluentd.

Quotes:

"The design and dedication of contributors and maintainers to Fluentd have proven time and again the value of a dedicated open community in creating a long-lasting project that welcomes new contributors. Fluentd is the oldest project within CNCF and it first caught my attention back in 2012. Each company I've worked with since has found value in using it to move, transform, and store data anywhere. Its early adoption of message queues, a package manager (RubyGems) for plugins, multi language SDKs and taking a logs-as-streams approach was ahead of its time. Contributors continue to create more integrations than ever with every cloud platform and database from the well-known to the obscure having a plugin. You'd be hard pressed to find a larger plugin ecosystem to integrate into for logging and use for both technical and business use cases of users new and old to the project, which is a hallmark of the continued success of Fluentd. The project maintainers continue to innovate and maintain relevance for cloud native container-based workloads and as a bridge for those moving workloads to the cloud." - Jordan Hamel, Principal PM of Microsoft Azure

"Fluent Bit is the future of the Fluentd project. It took the ideas that made Fluentd great and re-implemented them in C for high performance, and low memory overhead. As it grows in popularity, I expect it to be seen as not just a log collector, but as a generic lightweight data collection agent. In the cloud, data collectors are deployed to millions of instances and Fluent Bit's low resource usage can lead to significant savings. I think developers should standardize on Fluent Bit for all telemetry data collection; I am excited by opportunities to work with the OpenTelemetry community as they expand their focus to logs." - Wesley Pettit, Maintainer of Fluent Bit, a project within Fluentd

"Banzai Cloud is passionate about Cloud Native technology and observability. Fluentd provides the flexibility and reliability that is mandatory to operate in a highly dynamic environment like Kubernetes. Logging Operator, our open source project, relies on these capabilities to provide the easiest way to configure logging on a cluster." -Sándor Guba, Lead Engineer and Co-Founder at Banzai Cloud