

PulsarProposal

Pulsar Proposal

Abstract

Pulsar is a highly scalable, low latency messaging platform running on commodity hardware. It provides simple pub-sub semantics over topics, guaranteed at-least-once delivery of messages, automatic cursor management for subscribers, and cross-datacenter replication.

Proposal

Pub-sub messaging is a very common design pattern that is increasingly found in distributed systems powering Internet applications. These applications provide real-time services, and need publish-latencies of 5ms on average and no more than 15ms at the 99th percentile. At Internet scale, these applications require a messaging system with ordering, strong durability, and delivery guarantees. In order to handle the “five 9’s” durability requirements of a production environment, the messages have to be committed on multiple disks or nodes.

Pulsar has been developed at Yahoo to address these specific requirements by providing a hosted service supporting millions of topics for multiple tenants. The current incarnation of Pulsar has been open-sourced under Apache license in September 2016 and it is the direct evolution of systems that were developed at Yahoo since 2011.

We believe there is currently no other system that provides a multi-tenant hosted messaging platform capable of supporting a huge number of topics while maintaining strict guarantees for durability, ordering and low latency. Current solutions would require to run multiple individual clusters with additional operational work and capacity overhead.

Since the open sourcing of Pulsar, the development has been done exclusively on the public Github repository and two major releases were shipped (1.15 and 1.16), along with multiple minor ones. Several other companies have expressed interest in the project and its future direction.

Rationale

Pulsar is a platform that is built on top of several other Apache projects. In particular, Apache [BookKeeper](#) is used to store the data and Apache [ZooKeeper](#) is used for coordination and metadata storage. Pulsar is also interoperable out of the box with Apache Storm, to provide an easy to use stream processing solution.

We want to establish a community outside the scope of initial core developers at Yahoo and we believe that the Apache Foundation is a great fit and long-term home for Pulsar, as it provides an established process for community-driven development and decision making by consensus. This is exactly the model we want to adopt for future Pulsar development.

Initial Goals

The initial goals will be to move the existing codebase to Apache and integrate with the Apache development process. Furthermore, we plan for incremental development, and releases along with the Apache guidelines.

Current Status

Pulsar has been in service at large scale for more than 2 years at Yahoo. In this time around 60 different applications were integrated with Pulsar. Other companies are evaluating it as well and have been contributing code to the project.

Meritocracy

We value meritocracy and we understand that it is the basis to form an open community that encourages multiple companies and individuals to contribute and get invested in the project future. We will encourage and monitor participation and make sure to extend privileges and responsibilities to all contributors.

Community

We have validated, through the interest demonstrated by Pulsar users at Yahoo, that a reliable hosted pub-sub messaging platform represent a very important building block for web-scale distributed applications. We believe that many companies can benefit by applying the same model and that bringing Pulsar to Apache will get the community to grow stronger.

Core Developers

Pulsar has been initially developed at Yahoo and received significant contributions from Yahoo Japan. After having open-sourced the project there have been contribution from developers from several external companies.

Alignment

Pulsar builds upon other Apache projects such as [ZooKeeper](#) and [BookKeeper](#), along with a number of other Apache libraries. We have already integrated with Storm and we envision to integrate with multiple other systems in the streaming and big data space.

Known Risks

Orphaned Products

Yahoo has been doing most of the development and, given that many internal platforms depends on Pulsar, it is heavily invested in the long term success of the the project. Yahoo has a long history participating in open-source projects, and has been also a long time contributor to the Apache community.

Inexperience with Open Source

Many Pulsar contributors are already familiar with the open source process and several of them are committers on other Apache projects. We will be actively working with experienced Apache community members to improve our project.

Homogenous Developers

The initial committers are employed by large companies including Yahoo, Yahoo! Japan, Salesforce and [MercadoLibre](#). We hope to grow the community and to include additional committers based on their contributions to the project.

Reliance on Salaried Developers

It is expected that Pulsar development will occur on both salaried time and on volunteer time, after hours. The majority of initial committers are paid by their employer to contribute to this project. However, they are all passionate about the project, and we are confident that the project will continue even if no salaried developers contribute to the project.

Relationships with Other Apache Products

As mentioned in the Rationale section, Pulsar is closely dependent and integrated with [BookKeeper](#) and [ZooKeeper](#) and Storm. There are ongoing to integrate with other projects such Apache Spark. We look forward to collaborating with those communities, as well as other Apache communities.

An Excessive Fascination with the Apache Brand

We are applying to the Incubator process because we think it is the next logical step for the Pulsar project after open-sourcing the code in 2016. This proposal is not for the purpose of generating publicity. Rather, we want to make sure to create a very inclusive and meritocratic community, outside the umbrella of a single company. Yahoo has a long standing history of contributing to Apache projects and the Pulsar developers and contributors understand the implication of making it an Apache project.

Documentation

- Pulsar code base: <https://github.com/yahoo/pulsar>
- Pulsar documentation: <https://github.com/yahoo/pulsar/blob/master/docs/Documentation.md>
- Blog post: [Open-sourcing Pulsar](#), [Pub-sub Messaging at Scale](#)

Initial Source

The Pulsar codebase is currently hosted on Github: <https://github.com/yahoo/pulsar>. This is the exact codebase that we would migrate to the Apache Software Foundation.

Source and Intellectual Property Submission Plan

The Pulsar source code in Github is currently licensed under Apache License v2.0 and the copyright is assigned to Yahoo. All the contributions from external parties have been received under Apache style CLA. If Pulsar fulfills and passes the conditions for being an Incubator project in the ASF, Yahoo will transition the source code ownership to the Apache Software Foundation via the Software Grant Agreement.

External Dependencies

To the best of our knowledge, all of Pulsar dependencies are distributed under Apache compatible licenses.

External dependencies licensed under Apache License 2.0:

Athenz, JCommander, HPPC - High Performance Primitive Collections for Java, FasterXML Jackson, Caffeine Async Cache, GSON, Guava, Netty, [DataSketches](#), Joda-time, Jna Java Native Access, Lz4-java, [AsyncHttpClient](#), Jetty, SnakeYAML

ASF Projects:

[BookKeeper](#), [ZooKeeper](#), Storm, Log4J, Commons (BeanUtils, CLI, Codec, Collections, Configuration, Digester, IO, Lang, Lang3, Logging)

Others:

- Protobuf (3-clause BSD)
- JLine (BSD License)
- Jersey (CDDL - Version 1.1)
- [HdrHistogram](#) (BSD License)
- RocksDB-JNI (3-clause BSD)
- SLF4J API (MIT)

Required Resources

Mailing lists

- users@pulsar.incubator.apache.org
- dev@pulsar.incubator.apache.org
- commits@pulsar.incubator.apache.org
- private@pulsar.incubator.apache.org (with moderated subscriptions)

Git Repository

- <https://git-wip-us.apache.org/repos/asf/incubator-pulsar>

Issue Tracking

- JIRA Pulsar (PULSAR)

Initial Committers

- Matteo Merli - <mmerli@apache.org>
- Joe Francis - <joe@yahoo-inc.com>
- Rajan Dhabalia - <rdhabalia@yahoo-inc.com>
- Sahaya Andrews Albert - <sandrews@yahoo-inc.com>
- Maurice Barnum - <msb@yahoo-inc.com>
- Ludwig Pummer - <ludwig@yahoo-inc.com>
- Jai Asher - <jai1@yahoo-inc.com>
- Siddharth Boobna - <sboobna@apache.org>
- Nozomi Kurihara - <nkurihar@yahoo-corp.jp>
- Yuki Shiga - <yushiga@yahoo-corp.jp>
- Masakazu Kitajo - <maskit@apache.org>
- Sebastián Schepens - <sebastian.schepens@mercadolibre.com>
- Brad [McMillen](#) - <bradtm@yahoo-inc.com>
- Bobbey Reese - <breesee@yahoo-inc.com>
- Masahiro Sakamoto <massakam@yahoo-corp.jp>
- Hiroyuki Sakai <hsakai@yahoo-corp.jp>

Affiliations

- Matteo Merli - Streamlio
- Joe Francis - Yahoo
- Rajan Dhabalia - Yahoo
- Sahaya Andrews Albert - Yahoo
- Maurice Barnum - Yahoo
- Ludwig Pummer - Yahoo
- Jai Asher - Yahoo
- Siddharth Boobna - Salesforce
- Nozomi Kurihara - Yahoo! Japan
- Yuki Shiga - Yahoo! Japan
- Masakazu Kitajo - Apple
- Sebastián Schepens - Mercado Libre
- Brad [McMillen](#) - Yahoo
- Bobbey Reese - Yahoo

Sponsors

Champion

- Bryan Call

Nominated Mentors

- Dave Fisher
- Jim Jagielski
- P. Taylor Goetz
- Francis Liu

Sponsoring Entity

- The Apache Incubator PMC