PegasusProposal

Abstract

Pegasus is a distributed key-value storage system that is designed to be horizontally scalable, strongly consistent and high-performance.

- Pegasus codebase: https://github.com/XiaoMi/pegasus
- Website: https://pegasus-kv.github.io

Proposal

Pegasus is a key-value database that delivers low-latency data access together with horizontal scalability, using hash-based partitioning. Pegasus uses PacificA protocol for strong consistency and RocksDB as the underlying storage engine.

We propose to contribute the Pegasus codebase and associated artefacts (e.g., documentation, website content, etc.) to the Apache Software Foundation, and aim to build an open community around Pegasus's continued development in the ‘Apache Way’.

Background

Apache HBase was recognized as mostly the only large-scale KV store solution in XiaoMi Corp until Pegasus came out in 2015. The original purpose of Pegasus was to solve the problems caused by HBase’s two-level architecture and implementation, including high latency because of Java GC and RPC overhead of the underlying distributed filesystem, and long failover time because of single point of RegionServer and recovery overhead of splitting and replaying the HLog files.

Pegasus aims to fill the gap between Redis and HBase. As the former is in-memory, low latency, but does not provide a strong-consistency guarantee. And unlike the latter, Pegasus server is entirely written in C++ and its read-write path relies merely on the local filesystem.

Apart from performance requirements, we also need a storage system to ensure multiple-level data safety and support fast data migration among data centers, automatic load balancing, and online partition splitting.

After investigating lots of existing storage systems in the open-source world, we could hardly find a suitable solution to satisfy all the requirements. So the journey of Pegasus begins.

Rationale

Pegasus is a mature and active project which has been widely adopted in XiaoMi. After the initial release of open source project in 2017, we have seen a great amount of interest across a diverse set of users and companies.

Our experiences at committers and PMC members on other Apache projects have convinced us that having a long-term home at Apache foundation would be a great fit for the project, to ensure that processes and procedures are in place to keep project and community ‘healthy’ and free of any commercial, political or legal faults.

Initial Goal

Move the existing codebase, website, documentation, and mailing lists to Apache-hosted infrastructure.

Work with the infrastructure team to implement and approve our code review, build, and testing workflows in the context of the ASF.

Incremental development and releases along with Apache guidelines.

Current Status

Pegasus has been an open-source project on GitHub https://github.com/XiaoMi/pegasus since October 2017.

Meritocracy

The intent of this proposal is to start building a diverse developer and user community around Pegasus following the ASF meritocracy model. We plan to invite more people as committers if they contribute to this project.

Releases

Pegasus has undergone multiple public releases, listed here: https://github.com/XiaoMi/pegasus/releases.

These old releases were not performed in the typical ASF fashion. We will adopt the ASF source release process upon joining the incubator.

Code Reviews

Pegasus’s code reviews are currently public on Github https://github.com/XiaoMi/pegasus/pulls.
Community

Pegasus seeks to develop developer and user communities during incubation.

Core Developers

Currently, most of the core developers of Pegasus are working in the KV-Storage Team of Xiaomi. Yingchun Lai is one of the Apache Kudu PMC members. Zuoyan Qin is an experienced open-source developer who created sofa-pbrpc in his last job in Baidu. Wei Huang is also an active contributor of Apache Doris (Incubating).

- Zuoyan Qin (https://github.com/qinzuoyan)
- Yuchen He (https://github.com/hycdong)
- Tao Wu (https://github.com/neverchanje)
- Yingchun Lai (https://github.com/acelyc111)
- Wei Huang (https://github.com/vagetablechicken)
- Shuo Jia (https://github.com/Shuo-Jia)
- Liwei Zhao (https://github.com/levy5307)

Alignment

Pegasus is aligned with several other ASF projects.

We are working on a new feature to load data from the HDFS filesystem. Pegasus can also generate and store checkpoints to HDFS, for both backup and analysis purpose. We currently support offline analysis on checkpoints powered by Apache Spark.

Known Risks

Orphaned Products

The core developers of XiaoMi’s Pegasus team work full time on this project. There is very little risk of Pegasus getting orphaned since at least one large company (XiaoMi) is extensively using it in production, with currently a scale of 70+ clusters, 800+ tables, and more than 70TB data. Furthermore, since Pegasus was open-sourced at the beginning of October 2017, it has received more than 1200 stars and been forked more than 200 times, and also received some issues and pull requests from developers and users outside XiaoMi. We plan to extend and diversify this community further through Apache.

Inexperience with Open Source

The core developers are all active users and followers of open source. They are already committers and contributors to the Pegasus Github project. All have been involved with the source code that has been released under an open-source license, and several of them also have experience developing code in an open-source environment.

Several of the developers in XiaoMi’s storage team are committers and/or PMC members on other ASF projects (Kudu, HBase, Doris, etc.). They will guide others to practice the Apache Way together along with other incubator mentors.

Homogenous Developers

The project has received some contributions from developers outside of XiaoMi, and is starting to attract a user community as well. We hope to continue to encourage contributions from these developers and community members, and grow them into committers as they have time to continue their contributions.

Reliance on Salaried Developers

XiaoMi invested in Pegasus as general key-value storage used in company widely. The core developers have been dedicated to this project for nearly five years.

Besides, we look forward to attracting more people outside XiaoMi to contribute to this project, either paid engineers working on storage area, or individual volunteers, as long as they have enthusiasm for the Pegasus project.

An Excessive Fascination with the Apache Brand

Pegasus is proposing to enter incubation at Apache in order to help efforts to diversify the committer-base, not so much to capitalize on the Apache brand. The Pegasus project is in production use already inside XiaoMi, but is not expected to be a XiaoMi product for external customers. As such, the Pegasus project is not seeking to use the Apache brand as a marketing tool.

Documentation

Information about Pegasus can be found at https://github.com/XiaoMi/pegasus. The following links provide more information about Pegasus in open source:
Initial Source

Besides the core codebase, Pegasus also hosts its side projects on Github under XiaoMi Group. Specifically, the initial source includes:

Client libraries with different languages:

- Java-Client: https://github.com/XiaoMi/pegasus-java-client
- Scala-Client: https://github.com/XiaoMi/pegasus-scala-client
- NodeJS-Client: https://github.com/XiaoMi/pegasus-nodejs-client
- Go-Client: https://github.com/XiaoMi/pegasus-go-client
- Python-Client: https://github.com/XiaoMi/pegasus-python-client

Components of Pegasus:

- rDSN: https://github.com/XiaoMi/rdsn
- RocksDB: https://github.com/XiaoMi/pegasus-rocksdb

rDSN was initially a distributed framework developed by Zhenyu Guo from Microsoft, and we have heavily refactored and improved it to make it more fit for Pegasus. rDSN is MIT & Apache-2.0 dual-licensed. The code licensed Apache-2.0 belongs to XiaoMi and the copyright of MIT-licensed code is assigned to Microsoft. It's in our plan to merge Pegasus and rDSN as one project.

RocksDB is a Facebook-developed storage engine. Pegasus added some enhancements and modifications that may be incompatible with the original implementation. RocksDB is licensed under Apache 2.0 License.

External Dependencies

Pegasus has the following external dependencies.

- RocksDB (Apache 2.0)
- Apache Thrift (Apache 2.0)
- Boost (Boost Software License)
- Apache Zookeeper (Apache 2.0)
- Google s2geometry (BSD)
- Google gflags (BSD)
- fmtlib (BSD)
- POCO (Boost Software License)
- rapidjson (Tencent)
- libevent (BSD)
- Google gperftools (BSD)
- cameron314/concurrentqueue (BSD)
- cameron314/readerwriterqueue (BSD)
- XiaoMi/galaxy-fds-sdk-cpp (Apache 2.0)
- jupp0r/prometheus-cpp (MIT)
- curl (The curl license)
- nlohmann/json (MIT)
- abseil-cpp (Apache 2.0)
- antirez/linenoise (BSD-2)
- antirez/sds (BSD-2)

Build and test dependencies:

- Apache Maven (Apache Software License v2.0)
- cmake (BSD)
- Google gtest (Apache Software License v2.0)

Required Resources

Mailing List

There are currently no mailing lists. The usual mailing lists are expected to be set up when entering incubation:

- private@pegasus.incubator.apache.org
- dev@pegasus.incubator.apache.org
- commits@pegasus.incubator.apache.org

Git Repositories
Upon entering incubation, we want to move the existing repository from https://github.com/XiaoMi/pegasus to Apache infrastructure like https://github.com/apache/incubator-pegasus.

**Issue Tracking**

Pegasus currently uses Github to track issues. Would like to continue to do so while we discuss migration possibilities with the ASF Infra committee.

**Other Resources**

The existing code already has unit tests so we will make use of existing Apache continuous testing infrastructure. The resulting load should not be very large.

**Source and Intellectual Property Submission Plan**

Most of the current code is Apache 2.0 licensed and the copyright is assigned to XiaoMi. If the project enters incubator, XiaoMi will transfer the source code & trademark ownership to ASF via a Software Grant Agreement.

But due to historical issues, Pegasus was based on an MIT licensed code that was initially written by microsoft/rDSN, which has long been actively developed by Pegasus because the original project is unmaintained (modified code is licensed under Apache License 2.0). We aren’t sure if we should request Microsoft for any Contributor License Agreement (CLA) during the IP clearance process.

**Initial Committers**

- Zuoyan Qin (https://github.com/qinzuoyan, qinzuoyan@xiaomi dot com)
- Weijie Sun (https://github.com/shengofsun, luckyweijie@gmail dot com)
- Yuchen He (https://github.com/hycdong, heyuchen@xiaomi dot com)
- Tao Wu (https://github.com/neverchanje, wuta01@xiaomi dot com)
- Yingchun Lai (https://github.com/acelyc111, laiyinchun@xiaomi dot com)
- Wei Huang (https://github.com/vagetblechicken, huangwei5@xiaomi dot com)
- Shuo Jia (https://github.com/Shuo-Jia, jiashuo1@xiaomi dot com)
- Liwei Zhao (https://github.com/levy5307, zhaoiwei@xiaomi dot com)
- Liuyang Cai (https://github.com/LoveHeat)

**Affiliations**

Seven of the initial committers are employees of Xiaomi.

**Sponsors**

**Champion**

Von Gosling, vongosling@apache.org

**Nominated Mentors**

Kevin A. McGrail, kmcgrail@apache.org
Duo zhang, zhangduo@apache.org
Von Gosling, vongosling@apache.org
Liang Chen, chenliang613@apache.org

**Sponsoring Entity**

We are requesting the Incubator to sponsor this project.