

YetusProposal

Abstract

Yetus provides libraries and tools that enable contribution and release processes for software projects.

Proposal

Yetus helps community driven software projects improve their contribution and release processes by providing:

- a robust system for automatically checking new contributions against a variety of community accepted requirements
- the means to document a well defined supported interface for downstream projects
- tooling to help release managers generate release documentation based on the information provided by community issue trackers and source repositories

Background

Over the last several months a few folks have been working to generalize several tools used within the Hadoop project so that they can be reused across both ASF and projects outside of the foundation. The majority of this effort has been around the pre-commit patch testing facility used to evaluate contributions prior to review by a committer. The effort has reached a point where it's ready to start evangelizing for downstream users, which necessitates moving our fledgling effort out from under the Hadoop banner so that we can establish a release cadence and drive our own community growth.

Rationale

All software development projects that are community based (that is, accepting of outside contributions) face a common QA problem for vetting incoming contributions. Hadoop is fortunate enough to be sufficiently popular that the weight of the problem drove tool development (i.e. test-patch). That tool is generalizable enough that a bunch of other TLPs have adopted their own forks. Unfortunately, in most projects this kind of QA work is an enabler rather than a primary concern, so often the tooling is worked on ad-hoc and improvements are rarely shared across projects. Since the tooling itself is never a primary concern, any artifacts made are rarely reused outside of ASF projects. Having a solid set of build tools that are customizable to fit the norms of different software communities is a bunch of work. Making it work well in both the context of automated test systems like Jenkins and for individual developers is even more work. By focusing these efforts in the Yetus project, we can gain some additional reuse and network effects across both ASF projects and development projects that might adopt us outside of the foundation.

Initial Goals

- Establish release cadence sufficient to allow use by downstream projects for their critical contribution pipeline
- Demonstrate sufficient utility to convince existing ASF projects to move from one-off pre commit evaluation
- Ensure interoperability with non-ASF issue tracking / repositories likely to be used by potential non-ASF projects (e.g. [GitHub](#))

Current Status

Several months of work by several Hadoop contributors have established a generalized pre commit process and proof of concept integration for a half dozen or so projects. The current work is already in a releasable state as far as the code is concerned.

The other components slated for inclusion (e.g. shelldoc and releasedocmaker) are earlier in their development but already provide good incremental benefit over current tools in the handful of projects we looked at.

Meritocracy

The initial PMC list covers folks from several established ASF communities and several ASF members; they are all well acquainted with the importance of building incremental project responsibility for new contributors. Meritocracy will not be an issue.

Community

The nature of the project should allow a good chance for users of the project to transition to developers. During our recent adjustment period within the Hadoop project we have already attracted an active contributor we hope to quickly foster into a committer. Longer term, the PMC will work to draw in a group of folks traditionally underrepresented in ASF projects, namely QA, technical writer, and operations folks.

Core Developers

The initial set of developers come from a variety of ASF projects, all with related needs for contribution QA and release tooling. The list also includes several ASF members.

- Andrew Bayer (ASF member, incubator pmc, bigtop pmc, flume pmc, jclouds pmc, sqoop pmc, all around Jenkins expert)
- Sean Busbey (ASF member, incubator pmc, accumulo pmc, hbase pmc)
- Nick Dimiduk (hbase pmc, phoenix pmc)
- Chris Nauroth (ASF member, incubator pmc, hadoop pmc)
- Andrew Purtell (ASF member, incubator pmc, bigtop pmc, hbase pmc, phoenix pmc)
- Allen Wittenauer (hadoop committer)

Alignment

The community and the code making up the start of the Yetus project already exist at the ASF, the formation of the project merely formalizes our ability to make them something available to the public at large rather than an internal project detail.

Known Risks

Orphaned Products

The initial PMC are all involved actively in existing ASF projects that have expressed a desire for a common library that can be relied on for this kind of testing. Presuming those projects move forward with plans to integrate with Yetus, there should be sufficient demand to keep a stable amount of development interest.

Inexperience with Open Source

All initial PMC members have an established record of working within ASF projects and the current code base has already been developed within an established ASF project.

Homogenous Developers

The initial set of developers are employed by a variety of companies, located across the US and abroad, and used to working on a variety of distributed projects.

We have an apparent lack of diversity within several protected classes, though not out of line with demographics within the ASF itself. The PMC will seek to improve this when practical.

Reliance on Salaried Developers

The initial set of developers are paid by their employers to work on ASF projects as a part of their job duties, but their participation on this project is self motivated. We do not expect their interest to be directly tied to current employment, but will actively seek to grow our volunteer base regardless.

Relationships with Other Apache Products

A large number of ASF projects, mostly related to the Hadoop project in one way or another, currently rely on either a direct fork of the code Yetus grew out of or something inspired by that work. We have already been actively working to make sure we can provide the tooling needed for the Hadoop, HBase, and NiFi projects in the short term. As previously mentioned, we also have proof of concept integration for several other projects and will start working toward broader adoption as soon as we have a release.

A Excessive Fascination with the Apache Brand

The success of Yetus will be tied to its ability to improve the community process for downstream software projects. Adoption by a large number of ASF projects would provide a great indicator of its utility due to the reputation of the Apache Brand in regard to community focus. However, such a positive association with the Apache brand would happen as a result of adoption and not the fact that Yetus happens to also live under the governance of the foundation nor carries that brand itself.

Documentation

[1] Discussion of Yetus moving out from Hadoop

<http://s.apache.org/yetus-discuss-hadoop>

[2] Current test-patch documentation (linked on github for ease of rendering)

<https://github.com/apache/hadoop/tree/HADOOP-12111/dev-support/docs/>

[3] Umbrella ticket for Yetus within Hadoop Common

<https://issues.apache.org/jira/browse/HADOOP-12111>

Initial Source

The Yetus code currently is being developed within a feature branch of the Hadoop repository.

<https://git1-us-west.apache.org/repos/asf?p=hadoop.git;a=shortlog;h=refs/heads/HADOOP-12111>

Once Yetus has a repository of its own, a filtered history will be used to pare down to just the code intended for the project.

Source and Intellectual Property Submission Plan

All the code in question is already hosted on ASF infrastructure and authored by people with ICLAs on file.

External Dependencies

The current Yetus code base does not bundle any third party dependencies.

Cryptography

Yetus contains no special cryptographic components, though it does rely on common tooling for SSL encrypted communication.

Required Resources

Mailing Lists

- private@yetus.apache.org (moderated subscriptions)
- commits@yetus.apache.org
- notifications@yetus.apache.org
- dev@yetus.apache.org

Repositories

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

Issue Tracking

JIRA tracker with project YETUS

Other Resources

Yetus expects to make extensive use of Jenkins and related build infrastructure. Existing offerings on [builds.apache](https://builds.apache.org/) should suffice and where they do not the project will actively work with ASF Infra.

domain name:

- <https://yetus.apache.org>

Initial PMC

- Andrew Bayer <[abayer at apache dot org](mailto:abayer@apache.org)>
- Sean Busbey <[busbey at apache dot org](mailto:busbey@apache.org)>
- Nick Dimiduk <[ndimiduk at apache dot org](mailto:ndimiduk@apache.org)>
- Chris Nauroth <[cnauroth at apache dot org](mailto:cnauroth@apache.org)>
- Andrew Purtell <[apurtell at apache dot org](mailto:apurtell@apache.org)>
- Allen Wittenauer <[aw at apache dot org](mailto:aw@apache.org)>

Affiliations

PMC members are employees of (alphabetically) Altiscale, Cloudera, Hortonworks, and Salesforce.

Additional Interested Contributors

Those interested in getting involved with the project as it starts are encourage to list themselves here.

- Roman Shaposhnik < rvs at apache dot org>
- Jun Aoki <jaoki at apache dot org>
- Kengo Seki <sekikn at nttdata dot co dot jp>
- Joey Echeverria <joeey42 at gmail dot com>
- Bill Havanki <bhavanki at apache dot org>
- Jarek Jarcec Cecho < jarcec at apache dot org >
- Mark Grover < mark at apache dot org>
- Joe Witt < joewitt at apache dot org>
- < add here >

Sponsors

Champion

Sean Busbey <busbey at apache dot org>

Nominated Mentors

- Sean Busbey <busbey at apache dot org>

Sponsoring Entity

ASF Board. Note: this project is expected to go direct to TLP.