

# TwillProposal

## Abstract

Twill is an abstraction over Apache Hadoop® YARN that reduces the complexity of developing distributed applications, allowing developers to focus more on their business logic.

## Proposal

Twill is a set of libraries that reduces the complexity of developing distributed applications. It exposes the distributed capabilities of Apache Hadoop® YARN via a simple and intuitive programming model similar to Java threads. Twill also has built-in capabilities required by many distributed applications, such as real-time application logs and metrics collection, application lifecycle management, and network service discovery.

## Background

Hadoop YARN is a generic cluster resource manager that supports any type of distributed application. However, YARN's interfaces are too low level for rapid application development. It requires a great deal of boilerplate code even for a simple application, creating a high ramp up cost that can turn developers away.

Twill is designed to improve this situation with a programming model that makes running distributed applications as easy as running Java threads. With the abstraction provided by Twill, applications can be executed in process threads during development and unit testing and then be deployed to a YARN cluster without any modifications.

Twill also has built-in support for real-time application logs and metrics collection, delegation token renewal, application lifecycle management, and network service discovery. This greatly reduces the pain that developers face when developing, debugging, deploying and monitoring distributed applications.

Twill is not a replacement for YARN, it's a framework that operates on top of YARN.

## Rationale

Developers who write YARN applications typically find themselves implementing the same (or similar) boilerplate code over and over again for every application. It makes sense to distill this common code into a reusable set of libraries that is perpetually maintained and improved by a diverse community of developers.

Twill's simple thread-like programming model will enable many Java programmers to develop distributed applications. We believe that this simplicity will attract developers who would otherwise be discouraged by complexity, and many new use cases will emerge for the usage of YARN.

Incubating Twill as an Apache project makes sense because Twill is a framework built on top of YARN, and Twill uses Apache Zookeeper, HDFS, Kafka, and other Apache software (see the External Dependencies section).

## Current Status

Twill was initially developed at Continuity under the name of Weave. The Weave codebase is currently hosted in a public repository at github.com, which will seed the Apache git repository after renaming to Twill.

## Meritocracy

Our intent with this incubator proposal is to start building a diverse developer community around Twill following the Apache meritocracy model. Since Twill was initially developed in early 2013, we have had fast adoption and contributions within Continuity. We are looking forward to new contributors. We wish to build a community based on Apache's meritocracy principles, working with those who contribute significantly to the project and welcoming them to be committers both during the incubation process and beyond.

## Community

Twill is currently being used internally at Continuity and is at the core of our products. We hope to extend our contributor base significantly and we will invite all who are interested in simplifying the development of distributed applications to participate.

## Core Developers

Twill is currently being developed by five engineers at Continuity:

Terence Yim, Andreas Neumann, Gary Helmling, Poorna Chandra and Albert Shau.

Terence Yim is an Apache committer for Helix, Andreas is an Apache committer and PMC member for Oozie, and Gary Helmling is an Apache committer and PMC member for HBase. Poorna Chandra and Albert Shau have made many contributions to Twill.

## Alignment

The ASF is the natural choice to host the Twill project as its goal of encouraging community-driven open source projects fits with our vision for Twill.

Additionally, many other projects with which we are familiar and expect Twill to integrate with, such as [ZooKeeper](#), YARN, HDFS, log4j, and others mentioned in the External Dependencies section are Apache projects, and Twill will benefit by close proximity to them.

## Known Risks

### Orphaned Products

There is very little risk of Twill being orphaned, as it is a key part of Continuity's products. The core Twill developers plan to continue to work on Twill, and Continuity has funding in place to support their efforts going forward. Many other Big Data companies can benefit from Twill, and we have already received interest from various entities that would like to use and contribute to Twill.

### Inexperience with Open Source

Several of the core developers have experience with open source development. Terence Yim, Andreas Neumann and Gary Helmling are currently Apache committers for Helix, Oozie and HBase respectively Homogeneous Developers

The current core developers are all Continuity employees. However, we intend to establish a developer community that includes independent and corporate contributors. We are encouraging new contributors via our mailing lists, public presentations, and personal contacts, and we will continue to do so. Various entities have already expressed interest in becoming involved with Twill.

### Reliance on Salaried Developers

Currently, these developers are paid to work on Twill. Once the project has built a community, we expect to attract committers, developers and community other than the current core developers. However, because Continuity products use Twill internally, the reliance on salaried developers is unlikely to change, at least in the near term.

### Relationships with Other Apache Products

Twill is deeply integrated with Apache projects. Twill uses Apache YARN as its underlying resource management and task scheduling system and Apache Zookeeper for coordination. In addition, Twill uses Apache HDFS and Apache Kafka. A number of other Apache projects are Twill dependencies and are listed in the External Dependencies section.

### An Excessive Fascination with the Apache Brand

While we respect the reputation of the Apache brand and have no doubt that it will attract contributors and users, our interest is primarily to give Twill a solid home as an open source project following an established development model. We have also given additional reasons in the Rationale and Alignment sections.

## Documentation

The current documentation for Weave is at <https://github.com/continuity/weave>. This will be adapted to the new name Twill.

## Initial Source

The Weave codebase is currently hosted at <https://github.com/continuity/weave>. Weave will be renamed to Twill to seed the Apache git repository.

## External Dependencies

The dependencies all have Apache-compatible licenses:

- avro (Apache 2.0)
- hadoop (Apache 2.0)
- gson (Apache 2.0)
- guava-libraries (Apache 2.0)
- hbase (Apache 2.0)
- hdfs (Apache 2.0)
- kafka (Apache 2.0)
- netty (Apache 2.0)
- snappy-java (Apache 2.0)
- yarn (Apache 2.0)

- zookeeper (Apache 2.0)
- asm (BSD)
- junit (EPL v1.0)
- logback (EPL v1.0 )
- slf4j (MIT)

## Cryptography

Twill will depend on secure Hadoop, which can optionally use Kerberos.

## Required Resources

### Mailing Lists

- twill-private for private PMC discussions (with moderated subscriptions)
- twill-dev for technical discussions among contributors
- twill-commits for notification about commits

### Subversion Directory

Git is the preferred source control system: [git://git.apache.org/twill](https://git.apache.org/twill)

### Issue Tracking

JIRA Twill (TWILL)

### Other Resources

The existing code already has unit tests, so we would like a Hudson instance to run them whenever a new patch is submitted. This can be added after project creation.

## Initial Committers

- Terence Yim
- Andreas Neumann
- Gary Helmling
- Poorna Chandra
- Albert Shau

## Affiliations

- Terence Yim (Continuity)
- Andreas Neumann (Continuity)
- Gary Helmling (Continuity)
- Poorna Chandra (Continuity)
- Albert Shau (Continuity)

## Sponsors

### Champion

Vinod K <[vinodkv at apache dot org](mailto:vinodkv@apache.org)> (Apache Member)

### Nominated Mentors

- Arun C Murthy <[acmurthy at apache dot org](mailto:acmurthy@apache.org)>
- Tom White <[tomwhite at apache dot org](mailto:tomwhite@apache.org)>
- Patrick Hunt <[phunt at apache dot org](mailto:phunt@apache.org)>
- Andrei Savu <[asavu at apache dot org](mailto:asavu@apache.org)>

# Sponsoring Entity

We are requesting that the Incubator sponsor this project.