

March2013

Shepherds

Podling	Shepherd
Allura	mfranklin
Bloodhound	N/A
cTAKES	N/A
Drill	mfranklin
Hadoop Development Tools	mfranklin
Kalumet	mfranklin
Knox	New
Open Climate Workbench	mfranklin
S4	mfranklin
Streams	
Tez	New
Wave	grobmeier

Incubator PMC report for Mar 2013

This month saw a significant discussion around a proposal to modify the Incubator's policy to eliminate the option for podlings to graduate as sub-projects of existing PMCs. While this discussion still has unresolved issues, including moving short-form IP Clearance to legal, there is significant support for eliminating the option for a proposed podling to explicitly target a sub-project graduation path. Many participants agreed that retaining it as a viable exit path, by exception, was a logical compromise to eliminating it all together; specifically, when it is apparent that achieving TLP is not a viable option and there is a PMC willing to accept the code/community. The effect of this discussion can already be seen in the modification of the Curator proposal to keep TLP as a potential graduation target.

Additionally, the Incubator saw a substantial influx of new, accpeted projects with more proposals currently under discussion. In prior reports, the issue of mentor time & attention was raised and will need to be watched closely with the addition of so many new projects in a short timeframe. The only point of immediate concern in this space is the failed experiment for Shepherd's to self-organize. As a PMC, the Incubator needs to determine how we will handle the assignment and execution of the up-till-now successful shepherd role for any given report without the direct intervention of a single individual, such as the PMC chair.

Lastly, the MRQL project proposal was championed by an individual who is a PMC chair, but is not a foundation member or IPMC member. Discovery of this fact, combined with the perception that the vote was rushed, led some in the IPMC to immediately invalidate the project proposal and take actions to cancel creation of its resources. Further discussions have prompted other IPMC members to step forward as proposed champions and mentors in an effort to get the proposal and project creation back on track.

o Community

New IPMC members:

None

People who left the IPMC:

None

o New Podlings

The Incubator PMC voted to accept 6 new podlings since the last report

Curator - a set of Java libraries for working with Apache Zookeeper

Knox - a single point of authentication for Apache Hadoop services

MRQL* - a query processing and optimization system for large-scale data analysis

Open Climate Workbench - a framework focused on the rapid comparison of climate model output to remote sensing data.

Provisionr - a service to manage pools of virtual machines on multiple clouds

Tajo - a relational and distributed data warehouse system for Hadoop

Tez - a framework for processing arbitrarily complex data-processing tasks

* The MRQL acceptance has been mired in controversy due to an issue with the Champion's status. (See commentary)

o Graduations

The board has motions for the following:

Bloodhound
cTAKES

The following projects are graduating to an existing PMC:

EasyAnt to the Ant PMC

o Releases

Apache Onami Parent 0.3-incubating
Apache EasyAnt 0.9-incubating
Apache Crunch 0.5.0-incubating
Apache Mesos 0.10-incubating

o Legal / Trademarks
No issues at this time

o Infrastructure
No issues at this time

Still getting started at the Incubator
These projects are still getting started, so no immediate progress towards graduation is yet expected.

Knox
Tez

Not yet ready to graduate

Allura
Drill
HDT
Open Climate Workbench
S4
Streams

Not ready to graduate require attention

Kalumet
Wave

Ready to graduate

Bloodhound
cTAKES

Table of Contents

Allura
Bloodhound
cTAKES
Drill
Hadoop Development Tools
Kalumet
Knox
Open Climate Workbench
S4
Streams
Tez
Wave

Allura

Forge software for the development of software projects, including source control systems, issue tracking, discussion, wiki, and other software project management tools.

Allura has been incubating since 2012-06-25.

Three most important issues to address in the move towards graduation:

1. Continue to grow community
2. Move project development to ASF hardware
3. Continue to remove non-AL code from the repo

Any issues that the Incubator PMC (IPMC) or ASF Board wish/need to be aware of?

No issues at this time

How has the community developed since the last report?

1. Lots of contributions from Stefano Invernizzi and Simone Gatti.
2. Small patch received from SourceForge user 'honyczek'. No other new contributors.
3. Some new faces on IRC and the mailing list.
4. Professor mentoring Stefano and Simone has expressed willingness to recruit more students to work on Allura.

How has the project developed since the last report?

1. Allura instance set up at <http://allura-vm.apache.org/> but nothing moved to it yet. Need to finish setup and address any problems before moving development to the new vm.
2. GPL code made optional. Licensing review by Peter shows that only one Creative Commons javascript library needs to be addressed. LICENSE and NOTICE files created. Still need to apply Apache License header to all files.

Please check this [x] when you have filled in the report for Allura.

Signed-off-by:

Ross Gardler: [](allura)

Greg Stein: [](allura)

Jim Jagielski: [](allura)

Rich Bowen: [x](allura)

Shepherd notes:

Commit activity & mailing list traffic look good. I see that there are license issues that the podling has been dealing with, specifically removing non AL code. I am looking forward to seeing discussions about making a release in the near future.

Bloodhound

Bloodhound is a software development collaboration tool, including issue tracking, wiki and repository browsing.

Bloodhound has been incubating since 2011-12-23.

The most important issue the Bloodhound community needs to address is diversifying the community, but we believe this is no longer any impediment to graduation.

We're hoping to graduate soon, before the next report. While growing our community further continues to be important, we have seen more outside interest recently. This addresses the only barrier to graduation raised after the December2012 report. We have further prepared by completing the PODLINGNAMESEARCH, and just recently completed the community vote on graduation readiness.

The project now has two online demo instances running of Apache Bloodhound, one of which shows the current state of the trunk branch [<http://bh-demo1.apache.org>] and providing

a first implementation of the new search functionality, responsive layout and multi-product architecture that has been developed. Together these represent the main strands of what Apache Bloodhound set out to achieve. Unfortunately, we do not yet have the ability to make use of the repository browser, which was something that was brought up in the previous report.

We added two new committers to the project in January, and a fourth incubation release. We have also seen more interest in the project from a number of channels including irc and the dev mailing list which we are, as ever, hoping to convert into growth of the developer community and continue to improve upon diversity.

Please check this [x] when you have filled in the report for Bloodhound.

Signed-off-by:

Hyrum Wright: [](bloodhound)

Greg Stein: [](bloodhound)

Branko ibej: [x](bloodhound)

P.S.: The Incubator vote to graduate Bloodhound passed on 2013-03-13.

A resolution to establish the Apache Bloodhound TLP has been submitted to the Board.

Shepherd notes:

cTAKES

cTAKES (clinical Text Analysis and Knowledge Extraction System) is an natural language processing tool for information extraction from electronic medical record clinical free-text.

cTAKES has been incubating since 2012-06-11.

Three most important steps moving towards graduation

Attract new contributors

Make at least one cTAKES release

Get everyone's ICLA on file and start developing code at Apache and using the infrastructure

Anything required IPMC attention?

None

Community:

We added over 10 new committers after the podling was formed.

We initiated a community VOTE for graduation (http://mail-archives.apache.org/mod_mbox/incubator-ctakes-dev/201303.mbox/browser)

Development:

We successfully produced one released (ctakes-3.0.0-incubating).

Roadmap has been created on Jira and currently working on future releases.

Signed-off-by:

Jörn Kottmann: [](ctakes)

Grant Ingersoll: [](ctakes)

Chris Mattmann: [X](ctakes)

Shepherd notes:

Apache: Project Drill

Description:

Apache Drill is a distributed system for interactive analysis of large-scale datasets that is based on Google's Dremel. Its goal is to efficiently process nested data, scale to 10,000 servers or more and to be able to

process petabytes of data and trillions of records in seconds.

Drill has been incubating since 2012-08-11.

Three Issues to Address in Move to Graduation:

1. Continue to attract new developers with a variety of skills and viewpoints
2. Develop community skills and knowledge by building some releases
3. Demonstrate community robustness by rotating project tasks among multiple project members

Issues to Call to Attention of PMC or ASF Board:

none

How community has developed since last report:

Mailing list discussions:

There has been active participation in discussions on the developer mailing list, including new participants and developers. A few have participated in the users list; mainly activity takes place on developer mailing list.

Activity summary:

http://mail-archives.apache.org/mod_mbox/incubator-drill-dev/
March 2012, 21 by 6th of March (mainly jira; some discussion)
February 2013, 227 (jira, focused discussions)
January 2013, 169 (jira; focused discussions)
Dec 2012, 51 (jira, focused discussions)

Topics in discussion on the dev mailing list included but not limited to:

- * JSON scanner API
- * implementation of reference interpreter
- * building SQL parser
- * implementation of a variety of reference operators including flatten and WindowsPane
- * Mocking Library
- * Drill plus behavioral data

Presentations

There have been more than a dozen presentations from community members at international Hadoop conferences, Strata Conference, HUGs, JUG and an Apache Drill Users Group in at least four countries.

Slides

Slides from Drill presentations posted online such as at slideshare get a large number of views. Examples:

Japan Hadoop Conf. 2013 Winter, 2114 views
Boulder/Denver HUG, 848 views
PJUG Portland Oregon, 404 views
HUG Munich, 475 views

Articles

An invited article on Apache Drill, "Apache Drill: Newcomer in the Hadoop Ecosystem" appeared in the 30 January 2013 Software Developers Journal, authored by Ted Dunning and Jacques Nadeau.

In addition there have been a variety of blog postings about Drill.

Social Networking

@ApacheDrill Twitter entity is active and has grown to 147 followers.

How project has developed since last report:

1. Wiki has been built
2. Significant code drops have been checked in from a number of new developers
3. Added our first additional committer and PMC member, additional candidates are developing
4. Additional non-code contributors have become active and are being encouraged

Please check this [] when you have filled in the report for Drill.

Signed-off-by:

Ted Dunning: [x](drill)

Grant Ingersoll: [](drill)

Isabel Drost: [](drill)

Shepherd notes:

Drill appears to be healthy. Mailing lists are seeing a ton of traffic and work in the sandbox seems to be progressing at a reasonable pace. Question to the community: When do you estimate that you would want to start putting a preliminary release of some kind together? I assume this would require identifying at least some components that should be moved from "sandbox".

Hadoop Development Tools

Eclipse based tools for developing applications on the Hadoop platform.

Hadoop Development Tools entered the Incubator on 11/09/2012.

Three most important issues to address in the move towards graduation:

- Support multiple versions of Hadoop in a single IDE instance. During this time building understanding of the Apache processes around working and releasing.
- Release
- Grow the podling community in terms of users and contributors.

Any issues that the Incubator PMC (IPMC) or ASF Board wish/need to be aware of?

- None

How has the community developed since the last report?

- Bob Kerns was added to the PPMC
- Rahul sharma has begun contributing to the project (HDT-18)

How has the project developed since the last report?

- Wiki group permissions sorted out with mentors (thanks Adam Berry)
- Confluence permissions sorted out (thanks Adam Berry)
- requested FISHEYE instance for HDT code base <https://support.atlassian.com/browse/FSH-11186> (thanks Bob Kerns)
- discussed moving to Confluence for wiki
- David Crossley mentioned that HDT needed to fix its project metadata
- Bob Kerns put up HDT Product Experience Roadmap on the wiki <http://wiki.apache.org/hdt/HDTProductExperience>
- Discussing 0.1-incubating release

Please check this [X] when you have filled in the report for Hadoop Development Tools.

Signed-off-by:

Suresh Maru: [X](hadoopdevelopmenttools)

Chris A Mattmann: [X](hadoopdevelopmenttools)

Roman Shaposhnik: [](hadoopdevelopmenttools)

Shepherd notes:

The report adequately represents the state of the project. My only minor comment is that there appears to be a single individual responsible for most of the commits; but, it is way too early in the podling's lifecycle to determine whether or not there is a lack of community involvement.

Kalumet

Kalumet a complete environment manager and deployer including J2EE environments (application servers, applications, etc), softwares, and resources.

Kalumet has been incubating since 2011-09-20.

Three most important issues to address in the move towards graduation:

1. cut off a couple of incubating releases
2. we identified "high" priority features/changes for 0.7.0-incubating release. Kscripts and different

"ecosystems" (OSGi, JavaEE, etc) providers could be included in this release.

Any issues that the Incubator PMC (IPMC) or ASF Board wish/need to be aware of?
None so far

How has the community developed since the last report?
We submitted a couple of Kalumet 0.6.0-incubating releases to vote.

Unfortunately, due to "discussion" around legal files (especially NOTICE), the votes didn't pass.

We are fixing the legal issue and will submit a new 0.6.0-incubating release to vote.

How has the project developed since the last report?
We created the Jira corresponding to the changes that we want to include in Kalumet 0.7.0-incubating release.

We decided to increase the release cycle in order to give more visibility to the users.

We completed a first documentation. The documentation is part of the 0.6.0-incubating release and is also available directly on the website.

Please check this [] when you have filled in the report for Kalumet.

Signed-off-by:
Jim Jagielski: [](kalumet)
Henri Gomez: [](kalumet)
Jean-Baptiste Onofre: [X](kalumet)
Olivier Lamy: [](kalumet)

Shepherd notes:

I am a concerned about the recent drop in activity for the podling. There was a flurry of mail around the 0.6.0-incubating release and very little of substance since then, with absolutely nothing in the month of February. I wouldn't be so concerned if the vote was successful; but, since it was a cancelled vote, I would have assumed the community would try to fix the issues and get it out as soon as possible. These observations are, of course, made from a quick review, so I could be missing an important detail.

Knox

Knox Gateway is a system that provides a single point of secure access for Apache Hadoop clusters.

Knox has been incubating since 2013-02-22.

Three most important issues to address in the move towards graduation:

1. Complete the infrastructure setup for the project.
2. Complete the initial code grant.
3. Clear the project name with legal and pick a new name if required.

Any issues that the Incubator PMC (IPMC) or ASF Board wish/need to be aware of?
1. None.

How has the community developed since the last report?
1. Since we don't have all of our infrastructure setup yet we haven't made much progress on this front.
2. Have received a number of inquiries about participation in spite of this.

How has the project developed since the last report?
1. Have some of our infrastructure, such as mailing lists, setup.
2. Created PODLINGNAMESEARCH-25 to determine if Knox is a viable project name.
3. Auditing and cleaning up the existing code base in preparation for SVN import.
4. Working on website but that requires completing code grant.

Please check this [x] when you have filled in the report for Knox.

Signed-off-by:
Owen O'Malley: [X](knox)
Chris Douglas: [X](knox)
Mahadev Konar: [](knox)
Alan Gates: [X](knox)
Devaraj Das: [X](knox)
Chris Mattmann: [X](knox)

Tom White: [X](knox)

Shepherd notes:

Open Climate Workbench

Apache Open Climate Workbench (Incubating) is an effort to develop software that performs climate model evaluation using model outputs from a variety of different sources (the Earth System Grid Federation, the Coordinated Regional Downscaling Experiment, the U.S. National Climate Assessment and the North American Regional Climate Change Assessment Program) and temporal/spatial scales with remote sensing data from NASA, NOAA and other agencies. The toolkit includes capabilities for regridding, metrics computation and visualization.

Open Climate Workbench has been incubating since 2013-02-15.

Three most important issues to address in the move towards graduation:

1. Develop an Apache community for Open Climate Workbench and connect to other relevant Apache efforts (Tika, Hadoop, SIS, OODT)
2. Make an initial release.
3. Add new contributors to the project.

Any issues that the Incubator PMC (IPMC) or ASF Board wish/need to be aware of?

None at this time.

How has the community developed since the last report?

The project is still adding members of the initial project proposal. So far 71% of the proposal initial committers on the projects have ICLAs on file and accounts either already or under processing. 29% of the project committers (8 people) are still working on filing their ICLAs. Chris Mattmann is working to get these accounts opened.

How has the project developed since the last report?

Chris filed INFRA-5874 to bootstrap the podling. All mailing lists are setup, and now archiving, SVN repo has been created and is underway. The site has been created (using the CMS) by Suresh Marru.

Chris is also working with Craig Russell on filing a Software Grant for the RCMES initial codebase drop. As of Thursday March 14, 2013, the grant has been filed in r40293 of the foundation repository.

Please check this [X] when you have filled in the report for Open Climate Workbench.

Signed-off-by:

Chris Mattmann: [X](openclimateworkbench)
Suresh Marru: [X](openclimateworkbench)
Chris Douglas: [X](openclimateworkbench)
Nick Kew: [](openclimateworkbench)

Shepherd Comments:

None

S4

S4 (Simple Scalable Streaming System) is a general-purpose, distributed, scalable, partially fault-tolerant, pluggable platform that allows programmers to easily develop applications for processing continuous, unbounded streams of data.

S4 has been incubating since 2011-09-26.

Three most important issues to address in the move towards graduation:

1. growing the community
2. verifying (changing?) the name of the project. See PODLINGNAMESEARCH-10
- 3.

Any issues that the Incubator PMC (IPMC) or ASF Board wish/need to be aware of?

How has the community developed since the last report?

We have a new committer (dferro).

How has the project developed since the last report?

We kept iterating on the integration with Apache Helix. We also added major updates to the codebase for improving configurability (and usability).

We added the sources of the website to our Apache git repository, and prepared updates to the website and documentation for the forthcoming release (0.6). This includes migrating the documentation from the confluence wiki to the S4 website.

We are aiming at cutting a release candidate during the second week of march for a new release.

Please check this ☒ when you have filled in the report for S4.

Signed-off-by:

Patrick Hunt: ☒ (s4)

Arun Murthy: ☐ (s4)

Shepherd notes:

Things seem to be going well despite the smaller size. Do you think you will be proposing graduation before the next reporting period?

Streams

Apache Streams is a lightweight server for ActivityStreams.

Streams has been incubating since 2012-11-20.

Three most important issues to address in the move towards graduation:

1. Diverse participation in development. More of the community needs to be actively engaged.
2. Increase the codebase
3. Develop a larger community.

Any issues that the Incubator PMC (IPMC) or ASF Board wish/need to be aware of?

Not at this time

How has the community developed since the last report?

Craig McClanahan gave a presentation on Apache Streams at ApacheCon North America.

How has the project developed since the last report?

Discussions initiated at ApacheCon have recently been continued on the list and the major issues are being addressed.

Please check this ☐ when you have filled in the report for Streams.

Signed-off-by:

Matt Franklin: ☒ (streams)

Ate Douma: ☒ (streams)

Craig McClanahan: ☐ (streams)

Andrew Hart: ☐ (streams)

Shepherd notes:

Tez

Tez is a framework for processing arbitrarily complex data-processing tasks.

Tez has been incubating since 2013-02-24.

Three most important issues to address in the move towards graduation:

1. Develop collaborations with other Apache projects, including Hadoop, YARN
2. Make an initial Tez release.
3. Grow the Apache Tez community.

Any issues that the Incubator PMC (IPMC) or ASF Board wish/need to be aware of?

None at this time.

How has the community developed since the last report?

The community is still bootstrapping and we will need to probably finish getting ICLAs on board. Chris started a thread on dev@tez to request clarification on who needs ICLAs filed, and who needs accounts set up. Chris can help with this as can the other mentors.

How has the project developed since the last report?

Arun checked in the initial code drop on March 15, 2013 at <http://svn.apache.org/viewvc/incubator/tez/trunk/>. Mailing lists for dev, private, and commits are up and running.

Please check this [X] when you have filled in the report for Tez.

Signed-off-by:

Alan Gates: [](tez)

Arun C Murthy: [](tez)

Chris Douglas: [](tez)

Chris Mattmann: [X] (tez)

Jakob Homan: [] (tez)

Owen O'Malley: [] (tez)

Shepherd notes:

Wave

A wave is a hosted, live, concurrent data structure for rich communication. It can be used like email, chat, or a document.

Wave has been incubating since 2010-12-04.

Three most important issues to address in the move towards graduation:

1. Make a release (stalled).

The licensing issues have been resolved, and build scripts have been put into the repository. Documentation (see 2) is missing, after which point we can finalize a commit suitable for release.

2. Improve documentation.

This is currently scattered between the old WaveProtocol site, the mailing list, and elsewhere. It needs to be pruned (of no-longer relevant documentation), structured and put on the wiki.

3. Increase community size.

We are lacking any contributors who are able to make/support large code changes at the moment. A release should help increase wider knowledge of the project, to hopefully bring in more development support.

Any issues that the Incubator PMC (IPMC) or ASF Board wish/need to be aware of?

None at this time. (Though see Shepard's notes).

How has the community developed since the last report?

The mailing list traffic continues to consist of new people asking questions about setting-up/running a WIAB server. (Which could be resolved by #2). Other than that, mostly stagnation.

How has the project developed since the last report?

Some progress towards a release (licensing) has been made. Internationalization support has been added.

Please check this ☐ when you have filled in the report for Wave.

Signed-off-by:

Santiago Gala: ☐ (wave)

Upayavira: ☐ (wave)

Andrus Adamchik: ☐ (wave)

Vincent Sivetton: ☐ (wave)

Ben Laurie: ☐ (wave)

Christian Grobmeier: ☒ (Note: I am NOT a mentor, but i follow the project and due to timing issues I would like to confirm this report)

Shepherd notes:

grobmeier: The project is very silent with less commit activity (last commit before 5 weeks). Even when the committers are usually responsive on mailing-lists it is very clear that nobody there can spend enough time to actually drive Wave forward. The next weeks the question needs to be asked if the Incubator goals can ever be reached or if GitHub would be a better home for Wave. A new committer has been elected before a couple of months but the situation didn't change. The missing project report does unfortunately reflect the situation. From Mentor side of view, only Upayavira is actively and constructively contributing to the project. Actually the other mentors should be asked if they are still interested in mentoring Wave.