

# June2013

REPORT SUBMITTED. DO NOT EDIT FURTHER.

Incubator PMC report for Jun 2013

There is nothing very novel or interesting to report this month. The current chair exits as of this report. The reports below show the usual mixture of informative and not-so-informative.

Several IP clearances passed via 'lazy consensus.' They couldn't be lazier; no one on the IPMC (except the Foundation Secretary) indicated that they reviewed this transactions.

----- Summary of podling reports -----

Still getting started at the Incubator

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

Not yet ready to graduate

Ready to graduate

-----  
Table of Contents

Allura  
Curator  
Drill  
Falcon  
HCatalog  
jclouds  
Kalumet  
Knox  
MRQL  
Open Climate Workbench  
Provisionr  
S4  
Streams  
Tajo  
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. Make a release
2. Continue to grow community
3. Move project development to ASF hardware

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

No

How has the community developed since the last report?

Some new folks on the mailing lists, including developers from the EU-funded project,

PROSE, who are deploying Allura at <http://opensourceprojects.eu/>. Patches contributed from Jon Schewe and Chris Tsai.

How has the project developed since the last report?

1. We believe licensing work is complete, so we should be able to make our first release soon.

2. Many bug fixes & feature developments. Several updates to our vagrant image for easier development.

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

Signed-off-by:

Ross Gardler: [X](allura)

Greg Stein: [ ](allura)

Jim Jagielski: [ ](allura)

Rich Bowen: [x](allura)

Shepherd notes:

-----

Curator

Curator - ZooKeeper client wrapper and rich ZooKeeper framework

Curator has been incubating since 2013-03-11.

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

1. Build community
- 2.
- 3.

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

No

How has the community developed since the last report?

A new committer is in the process of being on-boarded: Eric Tschetter

4 subscribers were added to the dev list.

8 subscribers were added to the user list.

How has the project developed since the last report?

Curator 2.0.0-incubating and 2.0.1-incubating have been released.

Date of last release:

Curator 2.0.1 released May 31st.

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

Signed-off-by:

Enis Söztutar: [ ](curator)

Luciano Resende: [x](curator)

Mahadev Konar: [ ](curator)

Patrick Hunt: [x](curator)

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/](http://mail-archives.apache.org/mod_mbox/incubator-drill-dev/)

June to date 5 June, 29 (mainly jira; some discussion)

May 2013, 135 (jira, focused discussions)

April 2013, 188 (jira; focused discussions)

March 2013 260 (jira, focused discussions)

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

- \* Evolution of logical plan syntax with addition of operators including the Value and Union Distinct operators
- \* Advantages and disadvantages of Parquet versus ORC
- \* ValueVector construct and requirements
- \* The relative performance of Janino based compilation versus javax.tools.Javacompiler
- \* Initial development of execution engine environment
- \* Discussion of various types of large array and off heap data structure libraries
- \* RPC protocol and framework

Code

For details of code commits, see <http://bit.ly/14YPXN9> and <http://bit.ly/19IyID1>

There has been great progress around both evolution of the reference interpreter and

In the last three months, there have been many commits including:

- \* Initial implementation of RPC framework
- \* Base client and Zookeeper based client abstraction
- \* SQL parser with JDBC driver
- \* Distributed query scheduling framework
- \* ValueVector implementations
- \* Large number of reference interpreter tests and fixes

Community Interactions

There is now a weekly Drill hangout conducted remotely through Google hangouts Tuesday mornings 9am Pacific

Time to keep core developers in contact in realtime despite geographical separation. Results from these discussions are shared with the discussion list through meeting minutes and all are welcome to attend. This has been helpful in speeding development and averages attendance of 8-10 developers each week.

#### Presentations

There have been presentations from community members at conferences, meet-ups and through the weekly Google hangout.

\* As you can see from <http://drill-user.org/> there were few more HUGs/BUGs where Drill was presented/discussed (in Europe) - the blog itself might also be considered to manifest a contribution (?)

\* We have published an article on Drill in the Big Data journal <http://www.liebertpub.com/big>

#### Sample presentations:

\* Introduction to Apache Drill, Bay Area Analytics Group 2 April 2013 by Tomer Shiran

\* Interactive Ad hoc query at scale: talk at Hadoop User Group UK by @mhausenblas

\* Apache Drill Technical Overview: talk at Google Hangout, May 22 by Jacques Nadeau available at <http://slidesha.re/123mSDh>

\* Drill Technical update @April 16 Hangout by Jacques Nadeau available at <http://slidesha.re/ZDBvWP>

\* Drill Dissection at NoSQL matters (April) @mhausenblas video available at <http://bit.ly/13Ffk7b>

\* All You Need to Know About Drill, talk during Big Data Week #bdw13 by Michael Hausenblas on 26 April <http://bit.ly/17LlrD>

\* Deep Dive into Drill Implementation 3 June at Berlin Buzzwords by Ted Dunning and Michael Hausenblas

#### Slides

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

#### Articles

An invited interview with Ted Dunning in an O'Reilly white paper by Mike Barlow titled "Real Time Big Data Analytics: Emerging Architecture" discussed Apache Drill; there have been a number of blog posts.

#### Social Networking

@ApacheDrill Twitter entity is active and has grown to 362 followers.

How project has developed since last report:

1. Wiki has been updated regularly
2. Significant code drops have been checked in from a number of developers
3. Significant design documents have been created and discussed
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: [x](drill)

Isabel Drost-Fromm: [x](drill)

#### Falcon

Falcon is a data processing and management solution for Hadoop designed for data motion, coordination of data pipelines, lifecycle management, and data discovery. Falcon enables end consumers to quickly onboard their data and its associated processing and management tasks on Hadoop clusters.

Falcon has been incubating since 2013-03-27.

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

1. Add new and diverse committers

2. Build and grow community
3. Releases at frequent and regular intervals

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

- No

How has the community developed since the last report? We now have users subscribed to the dev mailing list and have added one new contributor since the last report.

How has the project developed since the last report?

24 new JIRAs have been created since the last report and 7 have been resolved so far with 9 more in patch available status. We intend to release version 0.3 in the next month.

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

Signed-off-by:

Arun Murthy: [ ](falcon)  
Chris Douglas: [X](falcon)  
Owen O'Malley: [ ](falcon)  
Devaraj Das: [ ](falcon)  
Alan Gates: [X](falcon)

Shepherd notes:

-----  
jclouds

A cloud agnostic library that enables developers to access a variety of supported cloud providers using one API

jclouds has been incubating since 2013-04-29.

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

1. We are working on our first release - iterating over issues currently, making sure we have a proper ASF release ready to go.
2. Since this is the first ASF project for some of the committers, it's taking some time for everyone to adjust to the Apache way, but we're getting there.
3. We need to figure out details on our wiki and site implementation.

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

Not particularly - we're definitely edging against some of the boundaries of what's allowed in an Apache project re: license headers, specifically in test expectation files, but we're getting there.

How has the community developed since the last report?

The community is roughly the same - no changes in committer list. We have continued to receive patch submissions and JIRAs from new people, which is, of course, great.

How has the project developed since the last report?

A lot. => More specifically:

- \* We've moved our source to the Apache git repos.
- \* We've moved to ASF JIRA and mailing lists.
- \* We've renamed our Maven groupIds to org.apache.jclouds, and are publishing SNAPSHOTS to repository.apache.org.
- \* The jclouds.{org,com,net} domain names have been transferred to Apache, and the jclouds trademark is en route.
- \* We've been working on RCs for Apache jclouds 1.6.1-incubating, ironing out the many little gotchas, license issues, etc that are to be expected when folding such a large, complex existing project into the ASF. We hope to have a viable final RC presented to the IPMC for voting within a week or so, assuming our mentors sign off on the next RC.

Date of last release:

No ASF release yet.

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

Signed-off-by:

Brian McCallister: [ ](jclouds)  
Tom White: [X](jclouds)  
Henning Schmiedehausen: [ ](jclouds)  
David Nalley: [X](jclouds)  
Jean-Baptiste Onofré: [ ](jclouds)  
Mohammad Nour El-Din: [X](jclouds)  
Olivier Lamy: [ ](jclouds)  
Tomaz Muraus: [ ](jclouds)  
Suresh Marru: [X](jclouds)  
Carlos Sanchez: [ ](jclouds)

Shepherd notes:

Really? 10 mentors and no report? Is everyone relying on everyone else? (rgardler)

-----

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. Release 0.6.0-incubator (fixes on the legal files)
2. Complete the documentation
3. Refactore the console

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 Kalumet 0.6.0-incubating release to vote, not approved by an IPMC as it requires some fixes on legal files (NOTICE especially).

We submitted a talk to ApacheCon NA, but the talk has not been choosed.

This talk presents Kalumet, with the current features and the roadmap. It has been given to different "local" events in Europe.

How has the project developed since the last report?

We created the Jira corresponding to the changes that we want to include in Kalumet releases.

We decided to increase the release cycle in order to give more visibility to the users, as soon as 0.6.0-incubating will be out.

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

Date of last release:

- 0.6.0-incubating submitted 2 times (September, 2012 and November, 2012), but not cancelled.
- This release is in preparation with the legal fixes

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:

-----

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. Expand community to include more diverse committers.
2. Align technically with security work going in in Hadoop.
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. No new PPMC members or committers added since last report, though commits now coming from Larry and Kevin instead of mostly Kevin which is great.

How has the project developed since the last report?

1. 53 open issues out of 71 issues currently in JIRA.
2. ReviewBoard is set up per INFRA-6068.
3. Discussion underway to set up Git/JIRA integration.

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

Signed-off-by:

Owen O'Malley: [ ](knox)

Chris Douglas: [X](knox)

Mahadev Konar: [ ](knox)

Alan Gates: [X](knox)

Devaraj Das: [ ](knox)

Chris Mattmann: [X](knox)

Tom White: [X](knox)

Shepherd notes:

-----

MRQL

MRQL is a query processing and optimization system for large-scale, distributed data analysis, built on top of Apache Hadoop and Hama.

MRQL has been incubating since 2013-03-13.

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

1. Complete the first release
  - a) Ensure proper transfer of code
  - b) Verify distribution rights
2. Establish whether "Apache MRQL" is a suitable name

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?

We have posted a project description on the HadoopSphere blog, which initiated some discussion and interest. Since the last report, there was little activity in recruiting new users and developers, but we are very close to completing the first release. We believe that, as soon as we make the first release available, many users will start using the system, will request bug fixes, improvements, and additional functionality, and will become new contributors.

How has the project developed since the last report?

We are in the process of completing the MRQL codebase to make it ready for the first release, which we expect to take place in the next few weeks. The only major component missing that delays this release is switching to maven as our project management tool. We believe that using maven is important because it will facilitate contributions by (and recruitment of) committers and will ease version management. The

reasons for the delay are: 1) no one of the current developers has prior experience with maven 2) the codebase is non-standard because it uses non-standard tools to generate Java code.

Date of last release:

none yet

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

Signed-off-by:

Alan Cabrera: [X](mrql)

Anthony Elder: [ ](mrql)

Alex Karasulu: [ ](mrql)

Mohammad Nour El-Din: [X](mrql)

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?

We're up to 86% (24) of the initial project members having ICLAs on file and accounts. We've got 14% more (4) people still who have yet to submit their ICLA. Chris Douglas and Chris Mattmann have reached out to Bruce Hewitson to submit ICLA. 3 NASA participants from the original proposal are working on their getting their ICLAs approved. Chris Jack from University of Cape Town just had his account created.

Community is now discussing many issues on list: 5 threads (deprecation policy; ideas to improve the metrics module; proposed tool refactoring; tools for Apache Open Climate Workbench; and coding style are all being actively discussed.

How has the project developed since the last report?

- \* Wiki now being used: <https://cwiki.apache.org/confluence/display/CLIMATE/Index>
- \* Review Board reviews are being leveraged; as is JIRA.
- \* Commit activity is high; and not singled out to Mike Joyce and Cameron Goodale any longer (Whitehall; Huikyo Lee; now contributing)
- \* The project discussed and decided to add new JIRA issue labels for users pointing out which issues are e.g., easy, more difficult, etc., to tackle. Worked with Gav in infra to set this up.
- \* The team is discussing cutting a 0.1-incubating RC.

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)



-----

Provisionr

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

Provisionr has been incubating since 2013-03-07.

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

1. Make a release
2. Build a community
3. Document all the important bits

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

No

How has the community developed since the last report?

Some new folks on JIRA and on the mailing list (18 subscribers)

How has the project developed since the last report?

1. We believe licensing work is almost complete and we should be able to prepare the first release candidate soon

2. General code cleanups

Date of last release: none

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

Signed-off-by:

Roman Shaposhnik: [ ](provisionr)

Tom White: [X](provisionr)

Mohammad Nour El-Din: [X](provisionr)

Shepherd notes:

-----

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?

Ant Elder brought up the possibility of graduating. This is being discussed in the community.

How has the community developed since the last report?

New release created/approved.

How has the project developed since the last report?

Released a new version (0.6.0)

Date of last release: June 3rd 2013

Please check this [X] when you have filled in the report for S4.

Signed-off-by:  
Patrick Hunt: [x](s4)  
Arun Murthy: [ ](s4)

Shepherd notes:

-----  
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?  
Streams was discussed at the Apache BarCamp in Boston during the month of May.

How has the project developed since the last report?  
Some discussions have occurred on list for improved documentation for new users as well as some architectural discussion. After a period of inactivity, the discussions have picked up as of late.

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

Signed-off-by:  
Matt Franklin: [x](streams)  
Ate Douma: [x](streams)  
Craig McClanahan: [ ](streams)  
Andrew Hart: [ ](streams)

Shepherd notes:

-----  
Tajo

Tajo is a distributed data warehouse system for Hadoop.

Tajo has been incubating since 2013-03-07.

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

1. Make an initial Tajo release
2. Grow the Apache Tajo community
3. Foster more committers

Any issues that the Incubator PMC (IPMC) or ASF Board wish/need to be aware of?  
No

How has the community developed since the last report?  
\* We voted for the official Tajo logo.  
\* Two projects were accepted to GSoC 2013.

How has the project developed since the last report?  
\* We changed data types to SQL data types.  
\* We've started to use jenkins as a continuous integration (CI) tool.

Date of last release:

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

Signed-off-by:

Chris Mattmann: [X](tajo)

Owen O'Malley: [ ](tajo)

Alex Karasulu: [ ](tajo)

Shepherd notes:

-----

Tez

Tez is an effort to develop a generic application framework which can be used to process arbitrarily complex data-processing tasks and also a re-usable set of data-processing primitives which can be used by other projects.

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?

No new PPMC members or committers added since the last report. We need to work to do a better job of identifying new contributors, but there is great activity so I don't think this will be a big issue.

Another here that we need to do a better job of here is discussing issues on list. I see lots of JIRAs and commits, but few [DISCUSS] and other threads. There was a discussion of merging TEZ-1 which is great though.

How has the project developed since the last report?

1. 109 issues resolved and 81 currently open in JIRA (lots of work going on).
2. Discussion to merge TEZ-1 (branch) into master.
3. Giri and Hitesh created CI for Tez: <https://builds.apache.org/job/Tez-Build/8/console>

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

Signed-off-by:

Alan Gates: [X](tez)

Arun Murthy: [ ](tez)

Chris Douglas: [X](tez)

Chris Mattmann: [X](tez)

Jakob Homan: [ ](tez)

Owen O'Malley: [ ](tez)

Shepherd notes:

-----

Wave

Wave is rich, web-based, distributed, collaboration platform that allows users to interact in near real time. The wave platform includes a web-based user interface containing an rich-text. The system is extendable though widgets, robots, and editor doodads. The Wave In a Box implementation is developed in java using a variety of web technologies such as Web Sockets, Java Script, GWT, and supported by an operational transform based conflict resolution algorithm.

Wave has been incubating since 2010-12-04.

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

1. Finish the initial release
2. Continue increasing the community size
3. Improve documentation for new users

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 last month has seen significantly increased activity on the mailing list regarding long-term visions on where the Wave project should be heading, with an influx of new people (mostly as potential users) discussing how they can help the project. At one of their suggestions, a public discussion was held to help with Wave's publicity, with the aim to bring more people to the project.

How has the project developed since the last report?

Much progress has been made towards an initial release (named 0.4), with RC2 and RC3 having votes on the wave-dev list. RC3's vote ends on the 8th June and is looking likely to be submitted for a vote on the general incubator list within the next few days.

Meanwhile development has continued on the trunk, receiving 5 review requests with the last week from 3 different people (2 new to the project).

The documentation has also seen work, with 3 people have been granted wiki access to help.

Date of last release: N/A

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

Signed-off-by:

Santiago Gala: [ ](wave)

Upayavira: [X](wave)

Andrus Adamchik: [ ](wave)

Vincent Siveton: [ ](wave)

Ben Laurie: [ ](wave)

Shepherd notes:

The report is very accurate and Wave is actually ways more active than before. All ideas of moving it to the attic/github should be delayed, as the project might have a chance to succeed, if the activity persists. From the mentors, only Upayavira is active. Not heard of any of the others the past period. (grobmeier)