

# September2013

⚠ FINAL ⚠

This report has been submitted to the Board. Please do not edit further.

Incubator PMC report for Sep 2013

The Apache Incubator is the entry path into the ASF for projects and codebases wishing to become part of the Foundation's efforts.

It's been another relatively quiet and smooth month for the Incubator.

## \* Community

New IPMC members:

Ben Hindman  
Jake Farrell  
Andrei Savu

People who left the IPMC:

(None)

## \* New Podlings

(None)

## \* Graduations

The board has motions for the following:

Curator

## \* Releases

The following releases were made since the last Incubator report:

Aug 15 Apache Falcon (incubating) v0.3  
Aug 27 Apache jclouds 1.6.2-incubating

It took 5-8 days for the third IPMC vote to arrive.

Release	RC VOTE start	Third IMPC +1	Days
Apache Falcon (incubating) v0.3	Aug 07	Aug 12	5
Apache jclouds 1.6.2-incubating	Aug 18	Aug 26	8

However, the first incubating release for Allura has now been waiting since at least August 28 (13 days as of this writing) with only 1 IPMC vote cast so far. Two rounds of polite pings have yet to scare up additional votes.

## \* Miscellaneous

- \* Several proposals are in various states of preparation. Two seem close to launching an entry VOTE: Storm, "a distributed, fault-tolerant, and high-performance realtime computation system that provides strong guarantees on the processing of data", and Aurora, "a service scheduler used to schedule jobs onto Apache Mesos."
- \* Tashi's developers are currently running a community retirement vote. It seems likely that the podling will be retired within the next few weeks.
- \* VXQuery is being pressured to show progress by some IPMC members and defended by others.

\* The tradition of sending a "welcome" email to new podlings once their mailing lists are set up has been quietly introduced. Such mails are cross-posted to the podling's dev list and general@incubator. This is another offshoot of the "Incubator Ombud" discussions which began a few months ago, like the WhatToExpect wiki page.

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

\* Still getting started at the Incubator

Samza  
Sentry  
Spark

\* Not yet ready to graduate

No release:

Allura  
Drill  
Kalumet  
MetaModel  
MRQL  
Olingo  
Tajo  
Wave

Community growth:

Falcon  
Provisionr  
S4  
Streams

Other/unspecified:

Open Climate Workbench

\* Ready to graduate

Curator

\* Did not report

(None)

-----  
Table of Contents

Allura  
Curator  
Drill  
Falcon  
Kalumet  
MetaModel  
MRQL  
Olingo  
Open Climate Workbench  
Provisionr  
S4  
Samza  
Sentry  
Spark  
Streams  
Tajo  
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. Streamline release process (including voting)
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?

A merge request from long ago was improved and merged. Daniel Hinojosa, SourceForge Community manager, introduced himself and a flyer he is working on to promote Allura. A few emails on the users list. All other commit and email activity has been from existing PPMC members.

Developers from SourceForge continue to move their internal discussions to public forums when possible. Ticket updates now go to the mailing list, and more discussions and sharing of plans are occurring on the mailing list and IRC.

How has the project developed since the last report?

First release has been created and approved by PPMC, after a few weeks of tweaks and voting. It is now pending IPMC approval. Infra has helped us get <https://forge-allura.apache.org/> up, so we will be able to migrate our wiki, tickets, etc. to it.

Allura has gained an import infrastructure, bulk export, wiki macros for youtube and gittip, more extension points, SCM log view and API improvements and a variety of other assorted fixes & improvements.

Date of last release:

As of Sep 3, our first release is pending IPMC votes.

When were the last committers or PMC members elected?

2013-04-11

Signed-off-by:

[ ](allura) Ross Gardler  
[ ](allura) Greg Stein  
[ ](allura) Jim Jagielski

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:

none - graduation vote succeeded

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

of?

Vote to graduate as a TLP was successful on both dev@ and general@ [1][2]

How has the community developed since the last report?

- \* Iannois Canellos added as a new committer
- \* Vote to graduate as a TLP was successful on both dev@ and general@ [1][2]

How has the project developed since the last report?

2 successful releases

Date of last release:

2013-08-08

When were the last committers or PMC members elected?

2013-07-13

Signed-off-by:

[ ](Curator) Enis Söztutar  
[x](Curator) Luciano Resende  
[ ](Curator) Mahadev Konar  
[x](Curator) Patrick Hunt

[1] <http://s.apache.org/6S3>

[2] <http://s.apache.org/Tnf>

Shepherd notes:

-----  
Apache: Project Drill

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 and and early users with a variety of skills and viewpoints
2. Continue to develop deeper community skills and knowledge by building additional releases
3. Demonstrate community robustness by rotating project tasks among multiple project members

The community has made significant progress on items 1 and 2.

Issues to Call to Attention of PMC or ASF Board:

none

How community has developed since last report:

The most important activity is the run up to the Milestone 1 release.

Additional events include:

- \* Apache Drill project website redesigned to have a new look:  
<http://incubator.apache.org/drill/>
- \* Interactive "How to Run Drill" demo added to the Apache Drill wiki:  
<https://cwiki.apache.org/confluence/display/DRILL/Demo+HowTo>

## Mailing list discussions:

Subscriptions to the Drill mailing lists have risen to 383. There has been active and increasing participation in discussions on the developer mailing list, including new participants and developers. Participation on the user list is growing although still small; mainly activity takes place on developer mailing list.

## Activity summary for the dev mailing list:

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

- \* September to date 05/010/2013: 397(mainly jira; some discussion)
- \* August 2013, 394 (jira, focused discussions)
- \* July 2013,370 (jira; focused discussions)
- \* June 2013,297 (jira, focused discussions)

## Recent topics on the dev mailing list have included:

- \* Usability and introductory tutorials
- \* SQL semantics and extensions to type inference cases
- \* Implementation of various storage engines, including Parquet and ORC.
- \* Optimizer rewrites and operator implementations.

## Code

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

There has been a very significant ramp up of code commits during this quarter, as shown in this chart from GitHub:

Ten contributors have participated in this GitHub code activity; there have been 77 forks of the Apache Drill project on GitHub which is a good indicator of strong interest outside the group of core contributors.

## Recent code commits include but not limited to:

- \* full end-to-end execution of queries
- \* reorganization of the source tree to simplify initial user experience
- \* a number of new operators for the execution engine
- \* a pro tempore query optimizer that allows a physical plans to be generated
- \* the entire code generation framework
- \* Value Vector implementation

## Community Interactions

The weekly Drill hangout continues, conducted remotely through Google hangouts Tuesday mornings 9am Pacific Time to keep core developers in contact in realtime despite geographical separation.

The community stays in touch through @ApacheDrill Twitter ID, and by postings on various blogs including Apache Drill User <http://drill-user.org/>

Viability of community is also apparent through interest in next meet-up event for the Bay Area Apache Drill User group in late September, which is already attracting a robust audience. Volunteers are coming forward from audience members of presentations, such as the Drill workshop in July (see following).

## Presentations

There have been presentations and a Drill workshop from community members at conferences and meet-ups.

Several Drill contributors have other talks scheduled with different meetups in the upcoming months.

## Sample presentations (out of many):

- \* Drill talks by @mhausenblas at Hive London and in Paris in June
- \* Talk on Apache Drill by @mhausenblas and @ted\_dunning at Berlin Buzzwords
- \* Apache Drill hands-on workshop by @ted\_dunning and @intjesus at OSCON in Portland, Oregon USA in July for ~40 participants.
- \* Apache Drill project featured by panelist @tshiran in Aug for the "Hadoop + SQL" Hive Data Think Tank event in California Bay Area.
- \* Next meeting for the Bay Area Apache Drill User group is planned for September with talk and demo by Steve Phillips

#### Slides

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

Example:

OSCON Apache Drill workshop posted 1 Aug 2013 by Ted Dunning and Jacques Nadeau, 436 views.

#### Articles

Examples of articles on Apache Drill since last report include:

- \* Article by @mhausenblas and @intjesus "Introduction to Apache Drill: Interactive Ad-Hoc Query for Large-scale Datasets" Michael Hausenblas and Jacques Nadeau. Big Data. June 2013, 1(2): 100-104. doi:10.1089/big.2013.0011. <http://bit.ly/15101Y7>
- \* A blog post by @Ellen\_Friedman reports on that Drill-via-Amazon-Cloud event and includes links to slides: <http://bit.ly/18aS3Lk>
- \* Drill blog article by S. J. Vaughan-Nichols "Drilling into Big Data with Apache Drill" in Aug: <http://bit.ly/1309MXA>
- \* A blog posting on Drill by T. Shiran as a prelude to the Hadoop + SQL event by Hive Data Think Tank can be found here: <http://bit.ly/1cvxn5D>

#### Social Networking

@ApacheDrill Twitter entity is active and has grown by ~20%, to 437 followers.

How project has developed since last report:

1. Website homepage has a new design
2. Wiki has been updated
3. Significant code drops have been checked in from a number of developers
4. Started to create release candidates for the milestone one [first] release of Drill
5. New developers are contributing.
6. Additional non-code contributors have become active and are being encouraged

Signed-off-by:

[ ](drill) Ted Dunning  
 [X](drill) Grant Ingersoll  
 [X](drill) Isabel Drost-Fromm

Shepherd notes:

(Dave Fisher/wave)

I am concerned the podling is missing the DISCLAIMER in its source code and on most web pages. There is a lot of activity but it does not seem to include any drive to make any release. There is tremendous community building activity. Heavy on community and light on Release/IP.

Rebuttal to shepherd notes: (Ted Dunning/tdunning)

I think that the Dave is substantially incorrect in his assessment that there "does not seem to include any drive to make any release". This is

factually and patently incorrect. I have asked that the shepherd reconsider this comment and retract it as it does not reflect the substantial progress that the community is making towards a release.

(Isabel Drost-Fromm/ Apache Mahout, mentor Drill)

Dave, as for the missing disclaimer - there are still a few branding issues with Drill. The project is currently on it's way to the first release - for a discussion on an initial RC and issues related to it see also <http://markmail.org/thread/gl2z6dbcsjkbdelx> In the process of preparing this release a couple issues surfaced that seemed more fundamental and are now in the process of being fixed. The Drill community reacted very positively to the input they got and are diligently fixing problems. Mentors provided several links both to Apache-wide as well as Incubator-specific documentation on what needs to be checked for an Apache release. The next RC will show how helpful that turns out to be with fixing the existing problems.

(Grant Ingersoll, mentor)

Unlike most Incubator projects, Drill is starting from scratch code-wise and community-wise and thus it will take longer to release, IMO. AFAICT, the community has sustainable development underway. That being said, I was just thinking it needed to do a release and then lo and behold there was an RC and a vote. As Isabel points out, the community has responded well to the feedback on the release. Once this release is out, I would expect the community to release more often, but other than that, I think the community is developing well, and is, in fact, doing some novel community building things like Google Hangouts that are open to anyone and which I think are very positive.

-----  
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?

- \* In an effort to bring more contributors and users to the fold, we have had a user meetup in Aug.
- \* More users have joined the falcon users group and mailing lists.

How has the project developed since the last report?

- \* First ever release of Falcon was made on Aug 15, 2013 (0.3-incubating)
- \* We have seen fairly heavy activity during the last month. 32 new JIRAs were created since the last report and 18 of them have been resolved.

Date of last release:

2013-08-15 (0.3-incubating)

When were the last committers or PMC members elected?

At time of incubation. No new committers or PMC members elected since.

Signed-off-by:

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

Shepherd notes:

-----  
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 two versions (0.6-incubating and 0.7-incubating)
2. Refactoring of the model
3. Refactoring of 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 did a couple of informal presentations during events (JUG, JAX, etc). We got good feedbacks, but the message is that users are waiting for a first release to move forward.

How has the project developed since the last report?

The 0.6-incubating release is in the vote (3rd try), including the legal files fixes identified during the previous vote.

Date of last release:

0.6-incubating in progress

When were the last committers or PMC members elected?

No new committers or PMC members since the last report.

Signed-off-by:

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

Shepherd notes:

(rvs) Look like project is on track. Nothing jumped out at me as requiring any special attention. Mentors seem to be pretty engaged and steering the project in the right direction.

-----  
MetaModel

MetaModel is a data access framework, providing a common interface for



exploration and querying of different types of datastores.

MetaModel has been incubating since 2013-06-12.

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

1. More contributors to increase diversity of the community
2. Make first release under ASF umbrella
3. Looking forward to setup CI builds under ASF infra

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 new contributions for JIRAs and lively discussions on dev list about technical implementations.

@dev list has 26 members

How has the project developed since the last report?

1. We have merge mostly all issues from MetaModel old JIRA to ASF JIRAs that we want to solve for MetaModel.
2. We have removed the component of the software that require incompatible license with ASF 2.0 license (Microsoft Access module)
3. We have started discussion about first release under ASF.
4. We have updated the website (<http://metamodel.incubator.apache.org>) to adhere to ASF Incubator guidelines.

Date of last release:

None

When were the last committers or PMC members elected?

June 2013

Signed-off-by:

[x](metamodel) Henry Saputra  
[x](metamodel) Arvind Prabhakar  
[ ](metamodel) Matt Franklin  
[ ](metamodel) Noah Slater

-----  
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

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 already addressed most issues that prevented us from releasing our

first release, which we now expect to take place in the next few weeks. We believe that, after we make the first release available, we will get many new users and contributors.

How has the project developed since the last report?

We have switched to maven as our project management tool. We refactored the source tree to make it compliant to maven; we created maven modules for compilation; we used the maven assembly plugin to create binary and source release artifacts; we updated the MRQL wiki page accordingly (getting started, how to contribute/build/commit). We have sent a request to the ASF Infrastructure to setup our project in Nexus so that we can deploy our snapshots there. We need also to verify the copyright information. As soon as we get these two tasks done, we will upload our first release on Nexus and stage it for vote.

Date of last release:

none yet

When were the last committers or PMC members elected?

2013-03-13

Signed-off-by:

[X](mrql) Alan Cabrera  
[ ](mrql) Anthony Elder  
[ ](mrql) Alex Karasulu  
[ ](mrql) Mohammad Nour El-Din

Shepherd notes:

(rvs) Project looks reasonably well off -- no major surprises. I was a little bit surprised by a very low JIRA activity but on the other hand they've got an exemplary Wiki and ML traffic is not too shabby either.

-----  
Olingo

Apache Olingo provides libraries which enable developers to implement OData producers and OData consumers. While starting with an initial code base implementing OData version 2.0 it is also a clear goal to start implementing a library for OData 4.0 once the OData standard is published at OASIS. The focus within the community is currently on the Java technology but it is up to the community to discuss if other environments find interest.

Olingo has been incubating since 2013-07-08.

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

1. Grow the community
2. Make a first release
- 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?

- Initial committers actively using the mailing list and Jira to coordinate project activities
- Jira Issues and resolutions contributed (not only by committers)
- Committers actively working on the code base
- Committers contributed new features

How has the project developed since the last report?

- Infrastructure setup
  - Build Infrastructure established
- Code base preparation to be able to produce the first release continued
  - Package name changes done
  - License Headers added
  - Code Cleanup + Bugfixes in progress
  - Readme and Notice files added to distribution
  - Discussion happened on what content should be provided within the first release (binaries, sources, javadoc, samples)
- Web Site enhanced / extended
- First proposals about contents for documentation available

Date of last release:

- No release so far

When were the last committers or PMC members elected?

Signed-off-by:

[X](olingo) Alan Cabrera  
 [X](olingo) Dave Fisher  
 [X](olingo) Florian Müller

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.

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 committers or PPMC members added since the last report.

How has the project developed since the last report?

- \* Michael Joyce made the 0.2-incubating release after 1 release candidate on August 18, 2013.
- \* Maziyar Boustani finished work on the new CLI for running evaluations.
- \* The initial work for the refactoring of the Core API has been completed.
- \* Cameron Goodale has completed some end-to-end tests that use the new Core API.
- \* Shakeh Khudikyan and Andrew Hart have made multiple improvements to the UI
- \* History/Results page for displaying results from previous runs.
- \* The community had a good discussion regarding JIRA component names. A new system was agreed upon and put in place.

Date of last release: 18-AUG-2013

Signed-off-by:

[ ](opencli~~m~~ateworkbench) Chris Mattmann  
[x](opencli~~m~~ateworkbench) Suresh Marru  
[X](opencli~~m~~ateworkbench) Chris Douglas  
[ ](opencli~~m~~ateworkbench) Nick Kew

Shepherd notes:

-----  
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. Clearly define use-cases and scope
2. Build a community
3. Make a few more releases focused on usability

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?

No changes. We had more interactions on JIRA.

How has the project developed since the last report?

No major functional changes - refactoring core APIs.

Date of last release:

2013-07-02 (July 2, 2013)

When were the last committers or PMC members elected?

No new committers elected.

Signed-off-by:

[X](provisionr) Roman Shaposhnik  
[X](provisionr) Tom White  
[ ](provisionr) Mohammad Nour El-Din

Shepherd notes:

-----  
S4

S4 is a general-purpose, distributed, scalable, 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. The frequency of releases has been low in the past. Working to put out more releases.
2. The community rise, to be more active and flourish.

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

There also was discussion related to "retirement". Currently people of the community stepped up and decided continue the project. As result, came up with building community, providing more patches, implementing new features. Moreover, Kurtt Lin wrote:

"Cheer up, guys. I'm seeing opportunities in this stream processing field as well as in S4. What I haven't mentioned is that, S4 had found successful applications in our research group even back in the 0.3.0 days and now in our start-up. We have built scalable real time search over microblogs with some S4 features enhanced (the paper appears in 2013's EDBT proceedings, link is <http://www.edbt.org/Proceedings/2013-Genova/papers/edbt/a31-lin.pdf>); we also built a platform on top of S4 for detecting and analyzing real-time traffic conditions, processing over hundreds of millions of records every day. Now we plan to contribute to the community. I personally have created and responded to some JIRAs, but I would like to do more. Actually, I am ongoing proposing a design documentation of integrating S4 with Helix along the roadmap; this would help enhancing the elasticity feature for S4. So, I suppose we could put the "retirement" discussion aside? Give us the new comers some time perhaps a couple of months to see what we can do to make the community healthier. Of course, some of us are new; we shall ask for your help and guidance occasionally, please be responsive (as what you are doing now), :) You all have done brilliant jobs! :) "

One particular problem we have been having is with growth of the community. Having said that in a recent e-mail thread, we have received very positive feedback from some developers, so we would like to remain in incubation for a little longer so that we can get some traction.

When were the last committers or PMC members elected?

Was reported by March 2013 (Daniel Gómez Ferro).

When did the project last make any releases?

June 2013: S4 0.6.0 released.

Two patches are available now, due to 0.7 release 3 of 18 issues was solved, the rest in progress.

How has the community developed since the last report?

In March, the dev mailing list was reasonably busy (132 messages) and then that's dwindled to less than 35 per month since March. There was no activity in git on August at all. A little in July, more in June. There has been no change in the community. We have been seeing some traffic on the mailing lists, mostly questions about how to use S4, but the code contributions have been fairly minimal. However, some developers have expressed interested in contributing so we expect to start receiving more contributions in the near future. To grow the community, we plan to have a closer interaction with other Apache projects like Gora and Helix.

Signed-off-by:

[x](s4) Patrick Hunt  
[ ](s4) Arun Murthy

Shepherd notes:

(Dave Fisher/Wave) S4 knows how make releases. Is concerned about community growth and think that may impede graduation. I think that they are actually ready to graduate.

-----  
Samza

Samza is a stream processing system for running continuous computation on infinite streams of data.

Samza has been incubating since 2013-07-30.

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

1. Bringing in new contributors
2. Building an Apache community around the code
3. Doing a release

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?

Discussion lists have become active with numerous questions. First patch from the community has been received and committed. Initial committers are very good about using public lists and seeking community input.

How has the project developed since the last report?

All initial bootstrapping has been completed and all requested infrastructure is available and being used. Forty-nine JIRA issues have been opened. First Incubation release could be considered in the next month or two.

Date of last release:

None.

When were the last committers or PMC members elected?

None yet. Project has just completed one full month of incubation, having entered July 30.

Signed-off-by:

[X](samza) Chris Douglas  
[ ](samza) Arun Murthy  
[X](samza) Roman Shaposhnik

Shepherd notes:

-----  
Sentry

Sentry is a highly modular system for providing fine grained role based authorization to both data and metadata stored on an Apache Hadoop cluster.

Sentry has been incubating since 2013-08-08.

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

1. Get rest of the project infrastructure in place, namely build and webpage
2. Release at regular intervals
3. Build community

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?

This is the first board report for Sentry. There are 15 sign-ups on the dev list.

How has the project developed since the last report?

This is the first board report for Sentry. At this time, the basic project infrastructure is setup (git repo, review board, JIRA and wiki). Most committers (also PPMC members) have their accounts and are subscribed to the dev, commits and private lists. Few JIRAs have been filed and they are being actively worked on.

Date of last release:

We have not done a release of Sentry since it started incubation.

When were the last committers or PMC members elected?

This is the first board report for Sentry. Last committers and PMC members were elected when the project entered incubation.

Signed-off-by:

[X](sentry) Arvind Prabhakar  
[X](sentry) Joe Brockmeier  
[X](sentry) David Nalley  
[X](sentry) Olivier Lamy  
[X](sentry) Patrick Hunt  
[X](sentry) Thomas White

Shepherd notes:

-----  
Spark

Spark is an open source system for fast and flexible large-scale data analysis. Spark provides a general purpose runtime that supports low-latency execution in several forms.

Spark has been incubating since 2013-06-19.

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

1. Make a first Apache release (we're in the final stages of this)
2. Move JIRA over to Apache (<https://issues.apache.org/jira/browse/INFRA-6419>)
3. Move development to Apache repo (in progress)

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

We still need some help importing our JIRA, though Michael Joyce and INFRA have looked into it (see <http://s.apache.org/fi>).

How has the community developed since the last report?

We're continuing to get a lot of great contributions to Spark. UC Berkeley also recently hosted a two-day training on Spark and related technologies (<http://ampcamp.berkeley.edu/3/>) that was highly attended -- we sold out at over 200 on-site attendees, and had 1000+ people watch online. User meetups included a well-attended meetup on Shark (Hive on Spark) contributions at Yahoo!.

How has the project developed since the last report?

We've made a lot of progress towards a first Apache release of Spark, including changing the package name to org.apache.spark, documenting the third-party licenses as required in LICENSE / NOTICE, and updating the documentation to reflect the transition. This month we've also moved our website to an apache.org domain (<http://spark.incubator.apache.org>) and updated the branding there. Finally, on the code side, we have continued to make bug fixes and improvements for the 0.8 release. Some recently merged improvements include simplified packaging and Python API support for Windows.

Date of last release:

No Apache releases yet

When were the last committers or PMC members elected?

June 2013

Signed-off-by:

[x](spark) Chris Mattmann  
[ ](spark) Paul Ramirez  
[x](spark) Andrew Hart  
[ ](spark) Thomas Dudziak  
[x](spark) Suresh Maru  
[x](spark) Henry Saputra  
[x](spark) Roman Shaposhnik

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. Growing the active developer community
2. Making our second incubating release
- 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?

The community has welcomed one new committer, Danny Sullivan, to the team.

How has the project developed since the last report?

Architectural discussions (e.g.: handling issues of condensing unusually high volume/noisy streams, publisher identification, message splitting, etc.) continue to happen on-list and provoke thoughtful responses. We have also heard from a number of outside individuals about issues and questions related to deployment of our first incubating release. Several of these have turned into deeper relationships including the addition of Danny Sullivan as a committer/PPMC member last month. Streams is now working to create our second incubating release.

Date of last release:

2013-01-09

When were the last committers or PMC members elected?

2013-07-18

Signed-off-by:

[ ](streams) Matt Franklin  
[ ](streams) Ate Douma  
[ ](streams) Craig McClanahan

Shepherd notes:



(rvs) Look like project is on track. Nothing jumped out at me as requiring any special attention, except it would be nice to have to have a second release out (since the first release was done immediately after entering an incubation).

-----  
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. The first release
2. IP clearance check
3. Various committers

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?

- \* The mailing list traffic grows.

How has the project developed since the last report?

- \* Tajo is rapidly evolving and is being stable by core committers and contributors.
- \* Since the last report, 45 issues have been resolved, and 58 issues have been created.
- \* The logical planning and SQL parser are improved.
- \* Tajo cluster can be on standby mode for more interactive query processing or on on-demand mode for sharing cluster resource by using Yarn.
- \* The overall query response times are significantly reduced.
- \* We've started to design the JIT and vectorized query engine.
- \* We've started to discuss the first Tajo release.

Date of last release:

Not Yet

When were the last committers or PMC members elected?

Not Yet

Signed-off-by:

[x](tajo) Chris Mattmann  
[ ](tajo) Owen O'Malley  
[ ](tajo) Alex Karasulu

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 Websockets, Javascript, 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. Make the initial release
2. Extend the community
3. Decide on the future of the existing codebase

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?

Following the busyness noted in the last report, the community has since returned to a relatively quiet state. Last month, a second public discussion (on Hangouts) was held, to extend Wave's public image. Some experiments using volunteers from the community, to test out possible changes to the OT algorithm, have been planned.

How has the project developed since the last report?

Following on from 0.4-RC3 being rejected at the Incubator list. Work on 0.4-RC4 has been completed (albeit fairly slowly), with a request for review for RC4 posted to the Wave-dev list only a few days ago. This addressed the remaining licensing problems, and merged in some of the trunk development that had occurred in the meantime (new translations, federation fixed).

Some questions have been asked about the future of the existing codebase beyond the 0.4 release, with a rewrite looking preferred if enough development time can be secured.

Date of last release:

N/A.

When were the last committers or PMC members elected?

The last 2 committers were elected around 30th June 2013.

Signed-off-by:

[X](wave) Christian Grobmeier  
[ ](wave) Upayavira

Shepherd notes:

Mentor notes:

Ben Laurie and Vincent Siveton have stepped down as Mentors, leaving the project with two active mentors. Currently this is not an issue.  
(christian)