August2013

```
Incubator PMC report for Aug 2013

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

By Incubator standards, it's been a quiet month.  At 193 emails for July,
traffic on general@incubator was as light as it's been since May 2011, the
month before the Open Office proposal arrived.

* Community

  There were no changes to the IPMC roster.

* New Podlings

  Two new podlings entered the Incubator:

    Samza
    Sentry

* Graduations

  (None)

* Releases

  The following releases were made since the last Incubator report:

    Jul 24 Apache Ambari 1.2.4-incubating
    Aug 08 Apache Curator 2.2.0-incubating

  It took 3-6 days for the third IPMC vote to arrive.

    Release                      RC VOTE start  Third IMPC +1   Days
    -----------------------------------------------------------------
    Apache Ambari 1.2.4-incubating   Jul 17         Jun 23          6
    Apache Curator 2.2.0-incubating  Jul 31         Aug 02          3

* Miscellaneous

  o Discussions about a potential Incubator Ombud continued.

  o Discussions about a "welcoming committee" or other ways to facilitate
    orientation of new podlings, spun off from the WhatToExpect wiki page and
    the Ombud discussion, progressed but have not been put into action.


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

* Still getting started at the Incubator

  o Olingo
  o Samza
  o Spark

* Not yet ready to graduate

  o Blur (no release)
  o Droids (activity)
  o Falcon (community growth)
  o Hadoop Development Tools (no release)
  o Knox (community growth)
```

```
     o MetaModel (plan around compatibility breaks from move to Apache)
     o Open Climate Workbench (community growth)
     o Tez (no release)

* Ready to graduate

     o Ambari

* Did not report

     o NPanday
     o Tashi (second missed report)
```

Ambari

Ambari is a monitoring, administration and lifecycle management project for
Apache Hadoop clusters.

Ambari has been incubating since 2011-08-30.

   * Release 1.2.2, 1.2.3, 1.2.4 was done.
   * Preparing for 1.2.5 release expected to be in the next week or so.
   * New committers have been added:
     Oleksandr Diachenko,Xi Wang,Oleg Nechiporenko,Dmitry Lysnichenko, Chad
     Roberts, Andrii Tkach
   * New PPMC members added:
     Sumit Mohantly, Srimanth Gunturi, Nate Cole, Tom Beerbower, Siddharth
     Wagle, Jaimin Jetly

   * Meetup was held on June 25th at Hadoop Summit with good attendance:
     http://www.meetup.com/Apache-Ambari-User-Group/events/119184782/

   * Increased participation from others in the community outside of Hortonworks.

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

   None.

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?

   Users have been active on the lists and contributions from folks outside
   of Hortonworks has accelerated.

   users@ - 156
   dev@ - 113

How has the project developed since the last report?

  A lot of new features have been added to newer Ambari Releases.

Date of last release:

  July 2nd, 2013: Ambari 1.2.4-incubating

Signed-off-by:

  [X](ambari) Owen O'Malley
  [X](ambari) Arun Murthy

Shepherd notes:

  mfranklin: The podling's activity looks great and they are constantly adding
  new committers. I would like to see more discussion around graduation on
  the dev list.

Blur

Blur is a search platform capable of searching massive amounts of data
in a cloud computing environment.

Blur has been incubating since 2012-07-24.

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

  1. Licensing/Notice files
  2. Release
  3. Release

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 continue to be small but active.
  - Subscriptions: user@ - 38[+3]; dev@ - 46[+8]

How has the project developed since the last report?

  The majority of effort has been around tightening things up for an upcoming
  release.

Date of last release:

  XXXX-XX-XX

Signed-off-by:

  [X](blur) Doug Cutting
  [x](blur) Patrick Hunt
  [X](blur) Tim Williams

Shepherd notes:

  mfranklin: Podling activity appears to be solid. It would be good to see
  the community rally around producing a release.

Droids

Droids aims to be an intelligent standalone robot framework that allows to
create and extend existing droids (robots).

Droids has been incubating since 2008-10-09.

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

  1. Activity

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?

There has been no change in the community. The summer months have been
very quiet. There are enough active people for a viable PMC.

How has the project developed since the last report?

Quiet quarter.

Date of last release:

  2012-10-15

Signed-off-by:

  [ ](droids) Thorsten Scherler
  [x](droids) Richard Frovarp

Shepherd notes:

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?

  * More users have joined the falcon users group and mailing lists
  * 1 new contributor has joined the project

How has the project developed since the last report?

  33 new JIRAs were created since the last report and 15 JIRAs have been
  resolved. The first release post incubation is now up for VOTE within the
  dev community of falcon. Post passin g the vote, vote would be called in
  incubator-general for the release.

Signed-off-by:

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

Shepherd notes:

(marvin: No shepherd assigned, since the podling reported out of cycle.)

Hadoop Development Tools

Eclipse based tools for developing applications on the Hadoop platform

Hadoop Development Tools has been incubating since 2012-11-09.

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

  1. Release
  2. Support multiple versions of Hadoop in a single IDE instance.
  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?

  - Srimanth Gunturi contributed Hadoop Eclipse project[HDT-32]
  - Mirko Kaempf & Rahul Sharma added as commiters
  - Few JIRAs have also been filed during this period
  - hdt-dev has seen 105 mails during this period

How has the project developed since the last report?

  - The code has received a big contribution from Srimanth Gunturi which would
    enable interactions with HDFS and Zookeeper.
  - Discussion on the mailing lists while bringing up HDT with hadoop-eclipse
    and work in is progress to add MR side of things on the same.
  - Wizards for Mapper/Reducers higrated to new MR API[HDT-21] Three most
    important issues to address in the move towards graduation:

Date of last release:

  No releases yet

Signed-off-by:

  [X](hadoopdevelopmenttools) Suresh Marru
  [X](hadoopdevelopmenttools) Chris Mattmann
  [X](hadoopdevelopmenttools) Roman Shaposhnik

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. Vote passed to invite a new committer to the project.
  2. Engaging several interested parties in contributing plugins.

How has the project developed since the last report?

  1. Continue to add and improve support for secure Hadoop clusters.
  2. Driving secure cluster fixes of other Hadoop components.
  3. Finalizing adding Knox to Apache Bigtop.
  4. Closing down remaining 8 issues in preparation for an 0.3.0 release.
  5. Resolved 30(+4) of 92(+10) total issues currently in JIRA.

Date of last release:

  0.2.0 04/22/2031

Signed-off-by:

  [ ](knox) Owen O'Malley
  [X](knox) Chris Douglas
  [X](knox) Alan Gates
  [ ](knox) Mahadev Konar
  [ ](knox) Devaraj Das
  [X](knox) Chris Mattmann
  [ ](knox) Tom White

Shepherd notes:

  acabrera: Nice active podling.  It's already done a release.  I'm not sure why
  they haven't started discussing graduation but I would support such a move.

  There seems to be a few security oriented software named Knox.  There may be
  a naming problem here, unless the product was consistently named Knox
  Gateway and not Knox; just my opinion.

--------------------
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. Roadmapping of our first Apache MetaModel release. Since the namespace
     change is going to break backwards compatibility of the project anyway, a
     couple of "old but breaking ideas" needs to be accepted or rejected.
  2. Bring more diverse committers and contributors to the project.
  3. We will probably be having several third party modules for MetaModel,
     because of (L)GPL dependencies. We want to figure out a good way to make
     this understandable for users (through website or similar means).

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?

  The pending INFRA work was finished in end-july, and we are starting to take
  it into use.

  We've decided to go for a Review-Then-Commit policy. We will allow lazy
  concensus after 3 days and a minimum of one +1 vote.

  There are currently no concrete plans for a release. We have some
  discussions and proposals of features on the mailing list which we want to
  settle on first.

  Since incubation we have not made any new appointments for either committer
```

or PMC member.

How has the project developed since the last report?

   The code has been moved from eobjects.org to Apache's Git server
   (https://git-wip-us.apache.org/repos/asf/incubator-metamodel.git). 5
   bugfixes from the base release 3.4.4 was applied to our codebase as well.

   The namespace of the project has been changed to org.apache.metamodel
   (previously org.eobjects.metamodel). To aid migration, we've also
   implemented a helping facility for deserializing objects of the old
   namespace into the new.

   A performance optimization for the CSV module has been proposed and applied.

   The project now has a website (http://metamodel.incubator.apache.org/). The
   website is meant mostly as an 'appetizer', and we're planning to put
   lengthier pieces of information onto our wiki
   (http://wiki.apache.org/metamodel/), and if needed link to those from the
   website. Some initial content has been put on the wiki as well.

   The project contains a Microsoft Access module, which depends on the LGPL
   licensed Jackcess library. We are looking for a replacement dependency or to
   remove the module from the project.

Date of last release:

   None

Signed-off-by:

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

Shepherd notes:

   The podling is off to a good start (Dave Fisher - wave@)

NPanday

NPanday allows projects using the .NET framework to be built with Apache
Maven.

NPanday has been incubating since 2010-08-13.

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

   1.
   2.
   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?

How has the project developed since the last report?

Date of last release:

   XXXX-XX-XX

```
Signed-off-by:

  [ ](npanday) Dennis Lundberg

Shepherd notes:

  acabrera: Not a lot of activity.  Most of the mail is comes from the Jenkins
  server.  :)  Still, the lone developer seems to be reasonably active.  I
  think it's telling that there is no report filed.

--------------------
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

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?

  - All initial committers have signed the ICLA
  - Initial committers are all on board and started using the mailing
    list to coordinate project activities.
  - Committers started actively working on the code base
  - The code base was contributed by SAP AG via Software Grant


How has the project developed since the last report?

  - Infrastructure setup
    -- Git Repository (for OData V2.0)
    -- Issue Tracker (Jira)
    -- Mailing Lists
  - Check in of OData Library V2.0 as initial code base
  - The code base is prepared and enhanced to be able to
    produce the first release
    -- Package name changes
    -- License Headers added
    -- Code Cleanup + Bugfixes
  - Initial Web Site created with Project Overview,
    Documentation and Support section


Date of last release:

  - No release so far

Signed-off-by:

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

Shepherd notes:

  rgardler: The project is still in the initial setup phase so not much to
```

report. Mentors are engaged where they need to be.

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. Identify a Champion/VP candidate.
   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?

   No new committers or PPMC members added since the last report.

How has the project developed since the last report?

   * Cameron Goodale made the 0.1-incubating release after 5 release candidates
     on July 29, 2013.
   * Mike Joyce has a VOTE up for the 0.2-incubating release.
   * The UI and backend are now able to fully replicate the climate analysis
     performed by Kim et al., J. Climate 2013.
   * Maziyar Boustani and Mike Joyce both made screencasts demonstrating the UI,
     and linked them on the wiki.
   * A discussion of binding VOTEs by the IPMC on releases occurred, and the PPMC
     worked through the issues and were better informed of Incubator processes.
   * Kyo Lee and Alex Goodman continued to improve the metrics and viz for the toolkit.
   * Shakeh Khudikyan is working on a History page for displaying previous runs.

Date of last release: 29-JUL-2013

Signed-off-by:

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

Shepherd notes:

   rvs: Open Climate Workbench looks like a pretty healthy
   community with a strong potential for graduation.

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. Getting codebase imported
   2. Generating Apache community

```
   3. Imparting ASF way to new PMC

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?

   First report.

How has the project developed since the last report?

   First report. Incubating for one week.  Bootstrapping project.  Have JIRA,
   website and git repo up and running.  Dev discussion moving to list and
   JIRA.

Date of last release:

   None yet.

Signed-off-by:

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

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. Finish bringing up Apache infrastructure (the only system missing
      is JIRA, but we also still need to move out website to Apache)
   2. Switch development to work directly against Apache repo
   3. Make a Spark 0.8 release through the Apache process

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

   Nothing major. We've gotten a lot of help setting up infrastructure and the
   last piece missing is importing issues from our old JIRA, which we're
   working with INFRA on (https://issues.apache.org/jira/browse/INFRA-6419).

How has the community developed since the last report?

   We've continued to get and accept a number of external contributions,
   including metrics infrastructure, improved web UI, several optimizations and
   bug fixes.  We held a meetup on machine learning on Spark in San Francisco
   that got around 200 attendees. Finally, we've set up Apache mailing lists
   and warned users of the migration, which will complete at the beginning of
   September.

How has the project developed since the last report?

   We are finishing some bug fixes and merges to do a first Apache release of
   Spark later this month. During this release we'll go through the process of
   checking that the right license headers are in place, NOTICE file is
   present, etc, and we'll complete a website on Apache.

Date of last release:
```

```
  None yet.

Signed-off-by:

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

Shepherd notes:




-------------------
Tashi

An infrastructure for cloud computing on big data.

Tashi has been incubating since 2008-09-04.

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

  1.
  2.
  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?



How has the project developed since the last report?



Date of last release:

  XXXX-XX-XX

Signed-off-by:

  [ ](tashi) Matthieu Riou
  [ ](tashi) Craig Russell

Shepherd notes:

  rvs: Tashi looks completely dormant at this point. Despite my repeated
  on-list and off-list emails it appears that I couldn't find anybody to
  compile a report. The only discussion that resulted from my attempts is
  captured over here: http://markmail.org/thread/mveeuubf2fcdmgcw

  Personally I think we need to figure out a path to *some* kind of a
  resolution here. I don't think Tashi benefits from being an incubator
  project and we need to figure out how to get it to a different trajectory.

-------------------
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.

How has the project developed since the last report?

1. 170 jiras filed and 120 odd jiras resolved since the first week of June 2013.
2. The first Tez meetup was held at the Hortonworks office on July 31st and had an attendance of around 30+ users/developers from across the Hadoop ecosystem community.
3. Seeing more adoption from the Hive community as well as some initial prototyping work being done in Pig.
4. Looking to make a release in the next couple of months after the release of hadoop-2.1.0-beta (which Tez depends on).

We are to looking to increase both the user base as well as get more contributors by having more meetups and also expect a release to drive more adoption of Tez.

Signed-off-by:

  [X](tez) Alan Gates
  [X](tez) Arun Murthy
  [X](tez) Chris Douglas
  [X](tez) Jakob Homan
  [X](tez) Chris Mattmann
  [ ](tez) Owen O'Malley

Shepherd notes: