

# September 2011

September 2011 Board reports (see [ReportingSchedule](#)).

These reports are due here by **Wednesday, 14 September 2011** so that the Incubator PMC can relay them to the board.

Your project might need to report even if it is not listed below, please check your own reporting schedule or exceptions.

Please remember to include:

- The "incubating since" info.
- The project's top 2 or 3 things to resolve prior to graduation.
- A short description of what your project's software does.
- The **Signed off by mentor**: is for Mentor(s) to show that the Report has been reviewed.

Remember that this is a "text" report, so no markup please.

-----  
Ambari (was HMS)

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

- \* Incubating since 30 August 2011.
- \* Changed name to Ambari over trademark concerns.
- \* In process of moving onto Apache infrastructure:
  - \* Jira and subversion created.
  - \* Mailing lists requested (6 Sep), but not created.
  - \* Confluence requested (6 Sep), but not created
  - \* Committer accounts created.
  - \* Working on initial code import and code grant.

Signed off by mentor:omalley

-----  
BeanValidation

Bean Validation was accepted into Incubator on 1 March 2010.

The Bean Validation project is an implementation of the Java EE Bean Validation JSR303 specification.

There are no other important issues open before a possible graduation. Actually the project is discussing its graduation as TLP or into Apache Commons, as natural successor of Commons Validator.

Any issues that the Incubator PMC or ASF Board might wish/need to be aware of

- \* none

How has the community developed since the last report

\* Users community activity is stable, users slightly decreased the activity of filling issues on JIRA and asking questions, we suppose codebase/provided documentation start being mature enough to satisfy users needs.

How has the project developed since the last report.

- \* Started a 'extras' module development where putting validators not included in the JSR303 spec.
- \* planning the development for implementing next JSR330 spec version.

Signed off by mentor:kevan

-----  
Bigtop

Bigtop is a project for the development of native packaging and stack tests of the Hadoop ecosystem.

Bigtop entered incubation on June 20, 2011.

Primary issues blocking graduation:

- \* Need for increased diversity and additional committers.
- \* Incubating release including testing framework.
- \* Incorporation of functional stack testing.

Issues which Incubator PMC and/or ASF Board might need/wish to be aware of:

- \* Due to limitations in available platforms on Apache Jenkins infrastructure and need for VM spin-up/spin-down for tests, we are working directly with OSUOSL on build/test setup.

Community development since last report:

- \* Community meetup held August 18th, with mentors and committers alike.

Project development since last report:

- \* 0.1.0-incubating released.
- \* Website created.
- \* Additional component project (Mahout) added.
- \* Supported platforms voted on.
- \* Initial implementation of package validation tests implemented and live.

Signed off by mentor: phunt

-----  
Deft

Deft is a non-blocking, asynchronous, event driven high performance web framework running on the JVM.

Issues before graduation

- \* Project rename (Deft seems to be trademarked)
- \* Put together a first incubation release
- \* Find new committers

The PPMC has discussed that we probably need to rename Deft. The reason for this is to avoid future complication because of trademarks associated with Deft.

No significant change has been noticed regarding the Apache Deft community.

The Apache Deft web page is up and running (still a lot of documentation to be done).

Signed off by mentor: ngn, Mohammad Nour (mnour)

-----  
Etch

Etch was accepted into Incubator on 2 September 2008.

Etch is a cross-platform, language- and transport-independent framework for building and consuming network services. The Etch toolset includes a network service description language, a compiler, and binding libraries for a variety of programming languages.

- Etch binding-cpp is currently in development and some parts of the basic framework (OS abstraction layer, collection types, basic Etch components) are already implemented and available in the trunk
- A new developer Martin Veith from the BMW Car IT provided a lot of patches and some documentation stuff
- Fixed some smaller bugs in the C, Java and C# bindings
- The new Apache Etch website is nearly complete <http://etch.staging.apache.org/etch/> and will be migrated to the public area while the next weeks. A detailed Etch documentation will be converted afterwards (Docbook PDF and HTML)
- Community ramp up a little bit and we hope to get a better grounding

future tasks:

- Prepare next release and publish it
- Migrate to new Apache Etch CMS
- Further development of the binding-cpp
- Community development

Signed off by mentor: dashorst

-----  
= Flume =

Apache Flume is a distributed, reliable, and available system for efficiently collecting, aggregating, and

moving large amounts of log data to scalable data storage systems such as Apache Hadoop's HDFS.

Flume entered incubation on June 12th, 2011.

== Issues before graduation ==

- \* Create Flume web site.
- \* Make an incubating release.
- \* Grow the community size and diversity.
- \* Licensing and trademark issues.

== Community ==

\* Development activities are going steadily with eighteen JIRA issues created in the past month, and eighteen resolved.

\* Active development is going on in flume-728 branch which is an effort to address critical problems observed in the trunk implementation.

- \* The core interfaces have been defined for the first cut.
- \* Active development is going on for implementing HDFS sink.
- \* Active development is going on for implementing a reliable channel.
- \* Core lifecycle and configuration aspects of the system are still being tweaked to ensure support for common use-cases.

== Project developments ==

- \* Initial inquiry into trademark status.

Signed off by mentor: rgoers, phunt

-----  
Giraph

Giraph is a large-scale, fault-tolerant, Bulk Synchronous Parallel (BSP)-based graph processing framework that runs on Hadoop. Giraph entered the incubator in August 2011.

Project developments:

- \* Project website created.
- \* Confluence wiki created.
- \* Accounts were created for two of the committers.
- \* Project is entirely on Apache infrastructure.

Next steps:

- \* Adding new committers.
- \* Making a release.
- \* One of the initial committers still hasn't filed an ICLA. We either need him to move forward or remove him.

Signed off by mentor: omalley, mattmann

-----  
Gora

Gora is an ORM framework for column stores such as Apache HBase and Apache Cassandra with a specific focus on Hadoop.

A list of the three most important issues to address in the move towards graduation

1. Port Gora code and license headers into ASF license headers.
2. Develop a strong community with organizational diversity and with infection into existing ASF projects like Nutch and Hadoop.
3. At least one Gora incubating release

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

No, not at this time.

How has the community developed since the last report?

In July 2011, we elected Ioannis Cannelos to the Gora PPMC and as a Gora Committer. We've also had a lot of interest from the Nutch community lately (specifically Lewis John McGibbney) as they are trying to help us get a stable version of Gora 0.2 trunk out the door and working with Maven Central so that Nutch as a downstream consumer can leverage Gora in it's 2.0 trunk version.

How has the project developed since the last report?

Gora was voted into the Incubator by the IPMC on September 26, 2010.

We are on RC #4 for Gora 0.1.1-incubating, a small patch to 0.1-incubating to get Maven dependencies working and a process in place. There has been some recent dev activity in trunk by Alexis Detreglode to get a new Cassandra back-end store in place, and improve upon the existing one. There are also efforts underway by Henry Saputra to get a CI build on Jenkins going for Gora.

Signed off by mentor: mattmann

-----  
Hama

Hama was accepted into Incubator on 20 May 2008. Hama is a distributed computing framework based on BSP (Bulk Synchronous Parallel) computing techniques for massive scientific computations.

== Top 2 or 3 things to resolve prior to graduation ==

- \* Invite new active committers

== Issues for the Incubator PMC or ASF Board ==

None.

== Community development ==

- \* Now we have 3 people ready to become committers (ChiaHung Lin, Thomas Jungblut, Miklós Erdélyi).
- \* 'Miklós Erdélyi' has contributed graph computing framework on top of BSP.

== Project development ==

- \* Now we support multi-task.
- \* Migrated from Forrest to Maven site.
- \* Some bug and performance issues are fixed.
- \* Plan to integrate with Hadoop nextGen.

Signed off by mentor:

-----  
HCatalog

HCatalog is a table and storage management service for data created using Apache Hadoop.

The most important issues in moving the project to graduation are expanding the community of developers and producing a release of the software.

Since the last report we have:

- \* Made several attempts at an initial release, each of which have had issues with NOTICE or DISCLAIMER files. We are preparing for another release candidate.
- \* Continued feature development, adding two major new features (ability to write multiple partitions at once and a notification interface for data consumers).
- \* Added significant testing

Currently there are 60 subscribers to the user list and 59 on the dev list. There were 32 and 30 respectively last report (June 2011).

Signed off by mentor: gates

-----  
Isis

Isis is an ALv2 licensed implementation of the Naked Objects pattern. It is based on contributions of the original Naked Objects Framework along with a number of sister projects that were developed for the book "Domain Driven

Design using Naked Objects " (pragprog 2009).

Isis was accepted into the Incubator in 2011, September 7th.

#### Project Development

- \* Isis-0.1.2-incubating released during July 2011
- \* Ongoing work on new json viewer, implementing the restfulobjects.org spec
- \* Enhancements to sql object store

#### Community Development

- \* Reasonably active mailing list; first "real" problem/change request raised (and fixed)
- \* Frequent commits
- \* Isis members attended BarCamp Oxford in Sept, presented on Isis

#### Top 3 Issues to address in move towards graduation

- \* More blogging/publicity from existing community...
- \* More users of the framework...
- \* More committers to the framework

None of these issues requires Board attention.

#### New Releases

- \* Next release expected in Nov 2011

Signed off by mentor: struberg, Mohammad Nour (mnour)

#### Kafka

(introduced to Apache incubator on Jul 4, 2011)

Kafka provides an extremely high throughput distributed publish/subscribe messaging system. Additionally, it supports relatively long term persistence of messages to support a wide variety of consumers, partitioning of the message stream across servers and consumers, and functionality for loading data into Apache Hadoop for offline, batch processing.

#### A list of the most important issues to address in the move towards graduation

1. Successful podling release.
2. Invite diverse new active committers

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?

Mailing list traffic for July-August-September[12th] (user: 40, 65, 19; dev: 116, 371, 63) both show healthy growth trends. Qualitatively the -dev discussion has trended away from topics like "how should we configure Jira?" to "What's the best way to deal with multiple language bindings?". Several patches were submitted by first time contributors.

How has the project developed since the last report?

The general theme over the past month has been polish: fixing bugs, better unit tests, getting log levels right, build system cleanup etc.

Signed off by mentor:

Alan Cabrera

Kato

Kato

Kato was accepted into the Incubator on 6 November 2008.

Kato is a project to develop the Specification, Reference Implementation, and TCK for JSR 326: the JVM Post-mortem Diagnostics API.

Recent Activity:

- \* Some JIRA items have been raised regarding the commandline tomcat commands.
- \* Discussions with Oracle have been continuing. An individual from Oracle has been identified, but no discussions have come from this yet.

The following is planned for next reporting period:

- \* Decide in what form the podling should continue, if at all.

Before this project can graduate we need to encourage more participation in the project and grow the community.

Signed off by mentor: rdonkin (Robert Burrell Donkin)

-----

ManifoldCF

--Description--

ManifoldCF is an incremental crawler framework and set of connectors designed to pull documents from various kinds of repositories into search engine indexes or other targets. The current bevy of repository connectors includes Documentum (EMC), FileNet (IBM), LiveLink (OpenText), Meridio (Autonomy), SharePoint (Microsoft), JDBC, CIFS file systems, CMIS repositories, RSS feeds, and web content. Output support includes Solr, MetaCarta GTS, and OpenSearchServer. ManifoldCF also provides components for individual document security within a target search engine, so that repository security access conventions can be enforced in the search results.

ManifoldCF has been in incubation since January, 2010. It was originally a planned subproject of Lucene but is now a likely top-level project.

--A list of the three most important issues to address in the move towards graduation--

1. We need at least one additional active committer, as well as additional users and repeat contributors
2. We want to finish the current release before graduating
3. We'd like to see long-term contributions for project testing, especially infrastructure access

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

All issues have been addressed to our satisfaction at this time.

--How has the community developed since the last report?--

A book has been completed, and is now available in early-release form, available from Manning Publishing, at <http://www.manning.com/wright>. We have signed up a new committer in this quarter and are discussing a second. One of our mentors (Grant Ingersoll) resigned.

We continue to have user community interest. We are participating this year in both Apache Eurocon and Apache North America. We've had a number of extremely helpful contributions from the field, including the CMIS connector and the OpenSearchServer output connector. We have started to discuss graduation from the incubator, which may come to pass by the end of the year.

--How has the project developed since the last report?--

An 0.1 release was made on January 31, 2011, and a 0.2 release occurred on May 17, 2011. Another release is scheduled for September

15, 2011, and will contain significant new features, including two new connectors and a client scripting language.

Signed off by mentor:

-----

MRUnit - a library to support unit testing of Hadoop MapReduce jobs.

MRUnit entered incubation on March 8th, 2011.

Community

- \* Still looking to develop a broader community.
- \* Release and contribution docs under development.
- \* Eric Sammer doing a great job pushing development forward.
- \* Discussions about producing first MRUNIT release candidate.

Issues before graduation

- \* Make an incubating release
- \* Grow the community size and diversity

Licensing and other issues

- \* none - MRUnit was originally a subproject of Hadoop

Signed off by mentor: mattmann, phunt

-----

ODFToolkit

The ODF Toolkit is a set of Java modules that allow programmatic creation, scanning and manipulation of OpenDocument Format (ISO/IEC 26300 == ODF) documents. Unlike other approaches which rely on runtime manipulation of heavy-weight editors via an automation interface, the ODF Toolkit is lightweight and ideal for server use.

\* ODF Toolkit entered incubation on Aug 1st, 2011.

\* Most important issues to address.

- 1) Growing the community, increasing diversity of committers
- 2) Technical migration from the ODF Toolkit Union to Apache infrastructure, including code repository, website, bugzilla and wiki.
- 5) Successful podling release.

\* Any issues that the Incubator PMC or ASF Board might wish/need to be aware of  
None at this time.

\* How has the community developed since the last report

The mailing lists are ready now for 23 days. We have 37 subscribers. We invited the existing users to this new community and are trying to attract more new people join us. This should be easier once we have code in the repository.

\* How has the project developed since the last report.

Requests are in queue with Apache Infra for loading the code repository and the issue tracker. 70% website and wiki migration work has been done.

Signed off by mentor:

-----

Oozie

Oozie is a workflow management and scheduler primarily for Hadoop based jobs.

Oozie entered the incubation on July 11, 2011.

\* A list of the three most important issues to address in the move towards graduation:

- \* Make the first Oozie release from Apache incubation.
- \* Improve the documentations: user, development for quicker adoption
- \* Establish the formal contribution process (such as CTR vs RTC)

\* Any issues that the Incubator PMC or ASF Board might wish/need to be aware of:

\* No issues.

\* How has the community developed since the last report:

- \* Oozie users from github started moving to Apache Incubator.
- \* Using new Apache Oozie JIRA for issue tracking.
- \* Using oozie-users and oozie-dev mailing list provided by Apache instead of the same from yahoo group.

\* How has the project developed since the last report.

- \* Oozie source code is migrated to Apache SVN.
- \* Oozie code originally had Yahoo License. Replace those text by Apache License.
- \* Oozie product web page is created. Further improvement is ongoing on.
- \* Old github Issues and JIRA have been migrated to Apache Oozie JIRA.

Signed off by mentor: Alan Gates (gates)

-----  
OpenOffice.org

\* OpenOffice.org entered incubation 2011-06-13.

OpenOffice.org is an open-source, office-document productivity suite providing six productivity applications based around the OpenDocument Format (ODF). OpenOffice.org is released on multiple platforms. Its localizations support 110 languages worldwide.

\* Most Important To Address

1) Migration of the legacy OpenOffice.org website's content and services to Apache infrastructure, including defect tracking, wiki, forums, mailing lists, and cross-service registration using customized software not already supported by Apache projects and infrastructure. Successful negotiation of governance migration of user-supported services brought under incubation. Resolution of copyright, license and notice for content migrated from legacy OpenOffice.org website.

2) Completion of the IP-review portions of the incubation checklist, which will require getting an amended SGA from Oracle to cover additional source files; scrubbing of incompatible notices from SGA-licensed code and resolving provenance of other existing materials being migrated.

3) A Successful Podling Release

\* Issues for IPMC or ASF Board Awareness

Notices concerning encryption methods employed in code now in the podling SVN have not been produced; legal-discuss is being consulted in regard to product class for OpenOffice.org.

\* Community Development Progress

As of 2011-09-12 there are 72 committers, with 55 on the PPMC, up from 71 and 52 at last report. Eleven initial committers have failed to submit iCLAs and are out of communication.

Discussion is underway with the operators of the existing OpenOffice.org user-support forums for migration of the forums into the project, with adjustment of governance to provide appropriate PPMC oversight.

We have created a ooo-users.i.a.o mailing list. A Japanese-language ooo-general-ja.i.a.o is also starting.

We have reviewed a request for permission to use the OpenOffice.org trademark by a German book publisher, and sent our approval recommendation to Apache Branding.

A "Building OpenOffice.org for Linux" buildfest was announced on the project blog and carried out over the Internet in the first full week of September.

\* Project Development Progress

The OpenOffice.org trademarks have been transferred to Apache. The OpenOffice.org domain-name registrations are being transferred to Apache.

The legacy OpenOffice.org Issue Tracking Bugzilla has been moth-balled as read-only and an Apache Bugzilla established for continuation of Issue Tracking under the podling.

The main source code base has been transferred to Apache SVN and is being actively tested and modified. Merging of additional work spaces from OpenOffice.org, and preservation of versioning history is being



pursued. The current effort is focused on successful build of a counterpart of the last complete build at OpenOffice.org.

Test configurations of the OpenOffice.org forum system and the OpenOffice.org Wiki have been brought up on Apache infrastructure fixtures. Cutover of the forum system is anticipated as part of the OpenOffice.org migration.

Detailed planning continues on public wiki: <https://cwiki.apache.org/confluence/display/OOOUSERS/>

Signed off by mentor: grobmeier

-----  
RAT

Rat audits releases.

A renewed push started to find a final status for Rat. A consensus emerged that the best destination for Rat would be a top level project, even if the scope is broad enough to allow a suite of related products to be developed by the community. Hopefully, Rat will be in a position to graduate soon.

Work has started on new code complementing the classic plugins:

- \* Apache Rat Eye assists bulk reviews (coded in Python)
- \* Apache Rat Whisker automates the verification and generation of legal documents (LICENSE, NOTICE, etc) for application composed from many components (coded in Java)

The 0.8 release of the classic plugin is expected soon.

Trademark, branding and marketing issues remain unresolved. The Incubator guides no longer accord well with developments in ASF policy in this area. It seems appropriate that before graduation these issues should be sorted out, though this may require work to develop incubator policy, which may potentially delay graduation.

Signed off by mentor:

-----  
Rave

Apache Rave is a new web and social mashup engine. It will provide an out-of-the-box as well as an extendible lightweight Java platform to host, serve and aggregate (Open)Social Gadgets and services through a highly customizable and Web 2.0 friendly front-end.

Rave entered incubation on 2011-03-01.

Current Status:

- \* The project has adopted a montly release cycle:
  - a first time 0.1-incubating release candidate was accepted by the IPMC on July 8
  - a 0.2-incubating release candidate was canceled by the PPMC in August because the required Incubating DISCLAIMER file was missing in some artifacts
  - a 0.3-incubating release candidate was accepted by the IPMC on September 16
- \* Reach out to and further cooperation and coordination with Shindig is growing
- \* An integration of Wookie (Incubating) is targeted on short notice (this or next release cycle)
- \* Preliminary steps are made showing how to extend and customize Rave for end users/developers
- \* Jasha Joachimsthal has been elected as new committer in June
- \* The commit rate has been steadily growing (more than doubled since the last report)
- \* Mailing list activity remains high
- \* Rave has been added to ReviewBoard (reviews.apache.org) to provide better support for community contributions and patches
- \* Website documentation is steadily improved and extended
- \* A presentation about Rave, focusing on Apache, community and collaboration, was given by Matt Franklin and Ate Douma at TransferSummit 2011/UK (Oxford) September 8th

Next steps:

- \* Continue to build up awareness of Rave and grow the community
- \* Further collaboration and coordination with Shindig and Wookie
- \* Further modularize Rave to support extending and customizing for end users/developers
- \* Keep up the pace for the monthly release schedule, working towards a 0.4-incubating release by end September 2011

Issues before graduation:

- \* Complete 1.0 release

- \* Expand the community/user base

Signed off by mentor: ate

-----

Sqoop

A tool for efficiently transferring bulk data between Apache Hadoop and structured datastores such as relational databases.

Sqoop was accepted into Apache Incubator on June 11, 2011. Status information is available at <http://incubator.apache.org/projects/sqoop.html>.

Progress since last report:

- \* Development activity became stronger over the last month with twelve issues resolved and nine new issues created in this period.
- \* Sqoop PPMC voted in a new committer on the project - Bilung Lee.
- \* Sqoop is seeing healthy input from the community with respect to filing JIRA issues and providing patches.

Issues before graduation

- \* Create Sqoop web site.
- \* Make an incubating release.
- \* Grow the community size and diversity.
- \* Review all license headers (all contains Cloudera).
- \* Change java package from com.cloudera.sqoop to org.apache.sqoop.

Signed off by mentor: Olivier Lamy (olamy), phunt, Mohammd Nour (mnour)

-----

Wave

Incubating since: Dec-2010

Description: Wave is a real-time communication and collaboration tool. Wave in a Box (WIAB) is a server that hosts and federates waves, supports extensive APIs, and provides a rich web client. This project also includes an implementation of the Wave Federation protocol, to enable federated collaboration systems (such as multiple interoperable Wave In a Box instances).

Most important issues are:

- \* Migrate source code from code.google.com to SVN.
- \* Building up community.

Community:

The community shows stable levels of activity.

Project development:

- The migration of the source code was delayed due to technical issues of migrating Mercurial repository from Google Code to Apache SVN without losing history. After some discussion we decided to move the code to Apache infra by 28-th September 2011 even if that would require a clean check in without history.
- About 26 commits with improvements and bug fixes.

Signed off by mentor: