

July 2012

July 2012 Board reports (see [ReportingSchedule](#)).

This report is **CLOSED**.

This is the sixth Incubator report since we started putting more effort into overseeing our podlings and keeping better track of their progress towards graduation.

Over that time we had 51 podlings reporting twice on normal schedule. Of those podlings, one was stuck at IP clearance, four continued without a release, seven had a low level of activity for the whole quarter, and another seven were in various other stages without significant progress towards graduation. Many of these cases are already resolved, and the remaining are being looked at in more detail.

Over the same six-month time period we saw thirteen projects graduate, three retire, and five enter the Incubator as new podlings. There are currently 49 podlings in the Incubator.

o Community

Jörn Kottmann and Rich Bowen joined the Incubator PMC since our last report.

The Lucene.Net podling is currently preparing for graduation.

The following proposal for a new incubating project was accepted:

- Apache Allura

A proposal for a new project called Blur was brought up for discussion.

Various inactive podlings are being retired. We decided to retire the Kato podling as mentioned in last month's report. This month both the AWF and Kitty podlings are being considered for retirement due to inactivity. See the relevant cursory reports for details.

The proposed resolution to establish the Apache Steve project was briefly discussed, and the general consensus was that there is no specific need for the project to first go through incubation as the codebase and the development community around it already exists within the ASF.

The role of a Sponsor, i.e. an existing TLP that plans to take a podling up as a subproject upon graduation, came up in discussion related to the EasyAnt podling. Such a role reflects the concept of an umbrella project and is thus no longer that relevant. Whether to discontinue or redefine the Sponsor role remains to be decided.

o Releases

The following incubating releases were made since our last report:

- June 21st, 2012: Apache Airavata 0.3-incubating
- June 22nd, 2012: Apache Wookie 0.11.0-incubating
- June 25th, 2012: Apache Syncope 1.0.0-RC2-incubating
- June 27th, 2012: Apache Kafka 0.7.1-incubating
- July 9th, 2012: Apache DirectMemory 0.1-incubating
- July 10th, 2012: Apache Stanbol Entityhub 0.10.0-incubating

In addition the release of Apache Any23 0.7.0-incubating is imminent, as the release vote has just passed.

o Legal / Trademarks

The instructions on checking project names for suitability as trademarks are a bit inconsistent and partially outdated. We're working on fixing that.

o Infrastructure

Some of the recently graduated podlings have had trouble clearly communicating to the infrastructure team everything that's needed for migrating project infrastructure to a TLP. We'll need to find a way to better instruct podlings on this as they prepare to graduate.

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

Still getting started at the Incubator (4 podlings)

Allura, CloudStack, Crunch, cTAKES

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

Not yet ready to graduate (12 podlings)

No release: Any23, Celix, JSPWiki, VXQuery
Low activity: AWF, DeviceMap, EasyAnt, Kitty, ODF Toolkit
Low diversity: Chukwa, Mesos, Tashi

We expect the next quarterly report of projects in this category to include a summary of their actions and progress in solving these issues.

Ready to graduate (3 podlings)

DirectMemory, Kafka, Oozie

We expect these projects to graduate within the next quarter.

Allura

The Allura Project is forge software for the development of software projects, including source control systems, issue tracking, discussion, wiki, and other software project management tools. Allura entered incubation on 2012-06-25.

Allura only got its mailing lists a couple of hours ago. There is nothing more to report yet.

Signed off by mentor:

Any23

Anything To Triples (any23) is a library, a web service and a command line tool that extracts structured data in RDF format from a variety of Web documents.

Any23 was voted into the Incubator by the IPMC on October 1, 2011.

Three most important steps moving towards graduation

- Grow the Any23 PPMC and community
- Ensure at least one Any23 release
- Identify and liaise with other semantic web/linked data projects within the ASF to establish common goals and objectives.

The Any23 PPMC was delighted to extend an invitation to Peter Ansell during June. The invitation was subsequently accepted and we are happy to say that Peter is now on board.

During June/July the community has focused on constructing the 0.7.0-incubating release candidate. We have experienced several problems along the way which have delayed this significantly, however as of writing the community is now VOTE'ing on the 0.7.0-incubating (release candidate #2) release thanks to Simone Tripodi, we anticipate a release shortly. Since last reporting we have integrated a number of commits to the project from a range of people. The total number of commits sitting at >1174 since Any23 entered the Apache Incubator.

The website at <http://incubator.apache.org/any23/> is also up and functioning, and appears to be in line with the Apache branding requirements.

Lewis John McGibbney worked with ASF infrastructure to get a zone VM up and running to host the Any23 web service.

The community has again been in touch with members from Stanbol, Jena and Clerezza and we have already discussed a committer meetup and liason with the aforementioned sister projects at the forthcoming ApacheCon Europe.

Signed off by mentor: mattmann
Shepherd: Dave Fisher

AWF

AWF is a non-blocking, asynchronous, event driven high performance web framework running on the JVM. AWF, originally named Deft, entered incubation on 2011-07-08.

- AWF is about to retire from the Apache incubation
- No community development since last report
- No commits since last report.

A vote to retire the AWF podling is in progress.

Signed off by mentor:
Shepherd: Jukka Zitting

Celix

Celix is an implementation of the OSGi Specification in C.

Celix entered incubation on November 2, 2010.

Over the last months we have started working on "Native-OSGi" which is an effort to make a specification for a C and C++ bases OSGi implementation. As detailed in the graduation plan [1] in the previous report we hope to attract more community members with this move. Details for Native-OSGi can be found at [2].

Also the donation of the Device Access code has been finished and committed to the SVN repository. This has been a good experience from both a technical and a process point of view.

Furthermore there has been a slight increase on the mailing list, hopefully this is a trend that will continue in the coming months.

Most important issues to address before we can graduate:

- Make a first release and grow a larger community. It is also needed to attract some more committers, there are currently 2.

Any issues the Incubator PMC or ASF board not te be aware of:

None at this time

How has the community developed since the last report:

- There has been an increase on the mailing list, furthermore with the start of the Native-OSGi project we hope this will increase in the coming months.

How has the project developed since the last report:

- From a technical point of view, we are working towards a first release. The build files have been updated to support a more modular project structure. Since the implementation is moving forward (al be it slowly), the community growth is the biggest concern for Celix.

[1]: <http://incubator.apache.org/celix/community/boardreports/boardreports.html#2012-04>

[2]: <https://github.com/abroekhuis/NativeOSGi>

Signed off by mentor: marris
Shepherd: Matt Franklin

Chukwa

Chukwa is an open source data collection system for monitoring large distributed systems. Chukwa is built on top of the Hadoop Distributed File System (HDFS), HBase and Map/Reduce framework and inherits Hadoop's scalability and robustness. Chukwa also includes a flexible and powerful toolkit for displaying, monitoring and analyzing results to make the best use of the collected data.

- Mailing list is picking up some activities.
- 3 new contributors submitted patches since last report in April.

Most important issues to address:

Growing the community, especially attracting new developers

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:

Plan to release version 0.6 to fix some of the major bugs in version 0.5.

How has the project developed since the last report:

Prior to Hadoop summit, there was a thread discussion of retiring Chukwa because there does not seem to be much activity in the community. All PPMC members seem to agree to this idea. There were some free advertisement for Chukwa in Hadoop summit from word of mouth and speakers mentioned Chukwa in their talks. There seems to be some contribution activities happening after Hadoop summit. Discussion continued for retiring Chukwa because mentor does not think a community is developing. We plan to open enrollment for Chukwa committers for people who are interested to continue development for the next quarter. If activities generate a active community, then we will apply for becoming top level project. If no activities happen, then we will retire Chukwa.

Signed off by mentor:
Shepherd: Benson Margulies

CloudStack

CloudStack is an IaaS ("Infrastructure as a Service") cloud orchestration platform. CloudStack has been in incubation since 2012-04-16

The top 3 issues to address to move towards graduation:

- CloudStack's source still contains works that are prohibited for release under ASF guidelines
- Performing a release
- Migration of infrastructure (bug tracker, CI, websites, etc)

How has the community developed since the last report:

As of the time of this report at least one new committer has been added. Additionally, a number of previously unknown developers have begun making substantial quantities of bugfixes and even working on new functionality.

How has the project developed since the last report:

The project itself continues to deal with process questions as well as learning how to operate in the new environment. A good deal of work has been happening on resolving the problems in the code base around licensing,

though not currently close to finishing.

Signed off by mentor:

Crunch

Crunch is a Java library for writing, testing, and running pipelines of MapReduce jobs on Apache Hadoop.

Crunch entered incubation on May 27, 2012.

The most important steps towards graduation:

- Infrastructure setup (JIRA, Confluence, etc.)
- CCLA licensing of the existing Crunch code
- Adding new contributors
- Creating a release

Nothing that currently requires IPMC attention.

Community:

The developer mailing list has been very active with bug fixes, new features, and discussions of infrastructure setup and project policies, both from the existing committers and other developers with an interest in the project. The first patch from a non-committer is currently being prepared for submission: the code is written, but we were blocking on getting JIRA setup so that the copyright on the code could cleanly be assigned to the ASF. The JIRA issues were resolved earlier this week.

All ICLAs are in place. Cloudera has gathered all of the copyright assignments for the existing Crunch code from non-Cloudera developers and is preparing the CCLA to assign the copyrights on the existing Crunch code to the ASF.

Development:

The 15 commits on the project this month were primarily for documentation and bug fixes, although we are evaluating two larger patches that bring additional functionality to the library: 1) adding map-side joins and 2) supporting interactive pipeline creation and execution via the Scala REPL.

Signed off by mentor: phunt, tomwhite

cTAKES

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

cTAKES was voted into the Incubator by the IPMC on Monday, June 11, 2012.

Three most important steps moving towards graduation

- Attract new contributors
- Make at least one cTAKES release
- Get everyone's ICLA on file and start developing code at Apache and using the infrastructure

Anything required IPMC attention?

No

Community:

We are still wrapping up getting a few ICLAs for Guergana Savova and for Sean Finan. So far, little mailing list discussion, but the project is getting bootstrapped. The next step will be to start developing on list and getting the project going.

Jörn Kottmann who originally volunteered to Champion the project, was elected to the IPMC and can now officially be our Champion. Congrats Jörn!

The light discussion on the mailing list is because there will be another release on the old sourceforge site before the migration to the apache svn and name space updates.

Development:

So far, nothing much other than getting mailing lists set up, and other infrastructure tasks that can be tracked at:
<https://issues.apache.org/jira/browse/INFRA-4910>

Signed off by mentor: mattmann

----- DeviceMap

Apache DeviceMap is a data repository containing device information, images and other relevant information for all sorts of mobile devices, e.g. smartphones and tablets.

Entered incubation on January 3rd, 2012.

There are no issues that require the Incubator PMC's or the board's attention.

Little has happened since our last report in April, a prototype of client-side device probes has been created [1], a few messages have been exchanged about that but that's all.

At this point the future of the project is unclear, we should re-evaluate the situation for the next report in October to see if activity has picked up.

[1] <http://markmail.org/message/3bd63yqmuixn6co6>

Signed off by mentor: bdelacretaz, kevan
Shepherd: Matt Hogstrom

----- DirectMemory

(incubating since October 2011)

Apache DirectMemory is a multi layered cache implementation featuring off-heap memory management (a-la BigMemory) to enable efficient handling of a large number of java objects without affecting JVM garbage collection performance.

There is only one important issue to address in the move towards graduation

Understanding process/decision making guidelines (new committer process is undergoing testing, release process still yet to be worked out)

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

Tasks and proposals contributed by non-committers

How has the project developed since the last report.

- A first public release 0.1-incubating has been done.
- A talk of DirectMemory has been presented at DevoXX France (thanks to Olivier Lamy and Benoit Perroud). Slides (in French) are available at <http://www.slideshare.net/benoitperroud/direct-memory-3devoxfr201220418-12607286>
- A talk of DirectMemory has been presented (thanks to Benoit Perroud) at

Java User Group Lausanne. Slides are available at
<http://www.slideshare.net/benoitperroud/direct-memory-jugl20120308-12607297>

- A talk of Direct has been presented at BreizhCamp (thanks to OlivierLamy).
Slides are available at <http://www.slideshare.net/olamy/heapoff-wtf>

Signed off by mentor: Olivier Lamy
Shepherd: Benson Margulies

EasyAnt

EasyAnt is a build system based on Apache Ant and Apache Ivy.
Incubating since 2011-01-31.

Towards graduation, we need to:

- Create a release
- Build a community

Since the last report:

There was very low activity on the project. There was actually no commit on the project. Some work has been done though a bug was found on a dependency, Apache Ant, and a bug was reported, and a patch committed since one of the EasyAnt committer happens to have commit rights there.

After acknowledging this, a mail was sent on the dev mailing to remotivate people to focus on doing a release. A positive response and some commits were following.

Signed off by mentor: bodewig
Shepherd: Dave Fisher

JSPWiki

JSPWiki has been incubating since September 2007.

JSPWiki is a JSP-based wiki program.

Still following the objective of making the first Apache release:

4 JIRA issues were fixed since last period, the project website has been revamped, and few other niceties & fixes have got into trunk (for example lucene library was upgraded to 3.6.0).

2 Release candidates have been made, but they didn't succeed because of technical issues. Hopefully the third RC will become the first ASF release.

The main issues blocking graduation keep being the same as in the last report: resolution acceptance, community and IPMC graduation vote.

Signed off by mentor:
Shepherd: Jukka Zitting

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.

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

None.

Previously, there was some concern about the diversity of the project that we believe has been addressed:

- The project now has active committers from four companies
- and continues to receive patches from other contributors from various institutions.

We are considering whether to pursue graduation immediately or focus exclusively on the development of 0.8.

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:

- The mailing list continues to be very active with bug reports, patch submissions, feature requests and use cases.
[(Apr,May,Jun): kafka-users (151, 261, 251); kafka-dev (142, 195, 380)].
- Added two new committers.
- Conducted the first Kafka user group meeting with over 50 attendees (archived video link available here: <https://cwiki.apache.org/confluence/display/KAFKA/Kafka+papers+and+presentations>).
This was a very successful event that gave users and contributors an opportunity to meet and discuss Kafka usage at various companies, ongoing development efforts, feature requests, etc.
- Received and reviewed several major patches (including contributions from non-committers).

How has the project developed since the last report:

- Intra-cluster replication (KAFKA-50)
 - Considerable progress has been made on the design and implementation of this much-anticipated feature.
 - Regular status updates are being sent out on the kafka-dev mailing list.
- Another significant project that is under discussion and development is the consumer redesign (KAFKA-364) and consumer coordinator implementation (KAFKA-264).
- Released Kafka 0.7.1 in June. This is the second Kafka release since its introduction to Apache incubator

Signed off by mentor: Alan Cabrera
Shepherd: Matt Franklin

Kitty

Kitty is a lightweight, production focused, Java-based application server performance diagnostic and management utility. Kitty entered incubation on October 3rd, 2010.

A vote to retire the Kitty podling due to inactivity is in progress.

Signed off by mentor:
Shepherd: Mohammad Nour

Mesos

Mesos is a scalable cluster manager that can dynamically share resources between multiple computing frameworks, including Hadoop, Spark, and MPI.

Mesos entered incubation on December 23, 2010.

Progress since the last report:

- Released 0.9.0!
- Added support for Hadoop CDH3u3 and updated the tutorial.
- Revamped MPI framework support (thanks Harvey Feng!).
- New JS based web interface (thanks Jon Fuchs!).
- Initial cgroups abstractions/interfaces for better resource isolation on Linux (thanks Jie Yu!).
- Refactor of allocator for more sophisticated allocations (thanks Thomas Marshall!).
- Numerous bug fixes (FD_CLOEXEC, JSON escaping, killtree.sh issues, etc).

Top priorities prior to graduation:

- Same as before, finalize addition of new committers to the project!
- Do another release.

Issues for Incubator PMC or ASF Board:

- None at this time.

Signed off by mentor: tomwhite

Shepherd: Benson Margulies

ODF Toolkit

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.

Our last release was January 14th, 2012.

We voted in our most recent committer on November 16th, 2011.

Most important issues to address:

Growing the community, especially attracting new developers

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:

- We are mentoring two GSoC students on ODF- Toolkit related projects
- One of our leading developers has had a job change and is no longer involved with the project
- Considering content-related track at ApacheCon, perhaps with other related projects like OpenOffice and POI

How has the project developed since the last report:

Working on next release, to feature the new document encryption support

Signed off by mentor:

Shepherd: Ross Gardler

Oozie

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

Oozie entered the incubation on July 11, 2011.

The three most important issues to address in the move towards graduation:

- Improve the documentations: user and development for quicker adoption
- Automate the formal release process.

- Diversify the dev and committer base.

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:

- 88 JIRAs were resolved and 145 JIRAs were closed in last quarter.
- 103 JIRAs were created since last report.
- Oozie dev/users are active in the email lists (around 1000 emails in oozie-dev and 250+ in oozie-user).
- Around 6 new developers are contributing to the process and coding.

How has the project developed since the last report.

- The second Oozie release (3.2.0) from Apache incubator has been successfully accomplished.
- A lot of new features are added. Fixed a lot of bugs too. Nearly 230 JIRAs were closed/resolved in last three months.
- Oozie contribution process is now more stream-lined. Further improvement is going on.

Signed off by mentor: ddas

Shepherd: Matt Hogstrom

Tashi

Tashi has been incubating since September 2008.

The Tashi project aims to build a software infrastructure for cloud computing on massive internet-scale datasets (what we call Big Data). The idea is to build a cluster management system that enables the Big Data that are stored in a cluster/data center to be accessed, shared, manipulated, and computed on by remote users in a convenient, efficient, and safe manner.

Tashi originally encompassed just the tools to manage virtual machines using Xen and QEMU, but has been merged with Zoni, which manages the physical aspects of a cluster like power control, network settings and handing out physical machines.

In the period from April to July, the project did not ask to make another incubating release, but is ready to start the process for a new release incorporating the development efforts of this period.

Development efforts this period have included making user actions display assurance messages via the client in case of successful operations, extending the SQL database backend to support all Instance and Host fields that are already recorded via the alternative "pickled" backend.

The primitive scheduler gained additional resilience to refrain from scheduling load on hosts that are down transiently. The node manager service now tries to ensure that undelivered messages to the cluster manager are resubmitted regularly. Virtual machine migration was revised to ensure stale state wasn't being shadowed by new data, only to reappear when the migrated VM was shut down.

The code underwent a complete automatic analysis, fixing several issues. Furthermore, a few other minor additions, fixes and documentation updates were made.

The project has received code contributions from two non-committers in this period. MIMOS via Luke Jing Yuan have contributed "convertz" to the code base, a utility to convert a VM image to an image deployable to a physical machine provisioned by Zoni. Alexey Tumanov of CMU provided a communications timeout wrapper to handle the problem of threads hanging

forever, trying to communicate over a broken network connection.

Upcoming software goals are to separate the client into an administrative and a user interface, to investigate what is needed to support IPv6, replace RPyC, and to provide the ability to hand out server slices (operating system level virtualization). Besides CPU and memory, disk storage should also be a schedulable resource.

The project has a user community, but it is small. Growth mostly has happened by word of mouth. To show potential users at large the utility of this project, the author of this report is creating web pages to demonstrate how to accomplish distributed computing tasks. Base images of (free) OS installs will be provided to allow new users to get started quickly. Hopefully this will increase visibility of the project.

Items to be resolved before graduation:

- Generate more publicity for the project.
- Develop members of the user community to submit feature extensions.

Signed off by mentor:

Shepherd: Jukka Zitting

VXQuery

The VXQuery Project implements a standard compliant XML Query processor for parallel evaluation.

It has been in incubation since 2009-07-06.

VXQuery changed its focus from being a pure library implementation of XQuery in Java to being an engine for parallel XQuery evaluation. For the runtime VXQuery now leverages the Hyracks system which is an Apache-licensed parallel execution engine developed at UCI.

So far the top issues were

- 1) low activity
- 2) small community
- 3) create a release.

The activities on these issues are:

ad 1)

Since this change of focus - and the start of a GSoC project with this new focus - development activity has picked up significantly. This due to both activity of the GSoC student (Preston Carman) and of the existing committers.

ad 2)

Preston Carman a) has provided good patches for VXQuery and b) interacts very well with the existing community. Consequently he is a candidate to be the next committer for VXQuery. After the initial vote on vxquery-private failed (all committers voted +1, but there was no mentor/IPMC member vote), Jukka Zitting voted +1. The next step is to finalize the administrative part to make Preston a committer.

Further, there is a plan to publish a paper to show the projects utility and to increase the visibility of the incubation effort.

ad 3)

After the change of focus some things work, but not enough to build a release that is easily consumable. VXQuery should be released before a paper is published to enable readers or the paper to run the software.

Signed off by mentor:

Shepherd: Matt Franklin
