

Beam report to ASF Board, 2017-06

Due: Wednesday, June 14, 2017.

Board meeting: Wednesday, June 21, 2017.

Previous report: [March 2017](#).

Status: submitted.

Description:

Apache Beam is a unified programming model for both batch and streaming data processing, enabling efficient execution across diverse distributed execution engines and providing extensibility points for connecting to different technologies and user communities.

Issues:

There are no issues that require the Board's attention at this time.

Activity:

We have achieved a major milestone for the project -- the completion of the first stable release, version 2.0.0. It signifies a statement from the community that it intends to maintain API stability with all releases for the foreseeable future, and making Beam suitable for enterprise deployment. Additionally, version 2.0.0 improves user experience across the project, focusing on seamless portability across execution environments, including engines, operating systems, on-premise clusters, cloud providers, and data storage systems.

Beam continues to interconnect additional execution engines and data storage/messaging systems, and serves as a glue in the ecosystem. On the execution side, the work continues on the Apache Gearpump (incubating) runner, and a new effort on the JStorm runner has started. On the IO connector side, connectors for Apache Cassandra and Apache Hive's HCatalog have been contributed, and additional connectors for Redis, Apache DistributedLog (incubating), Apache Solr, Apache Parquet, RabbitMQ, and Advanced Message Queuing Protocol (AMQP) are in progress. Finally, we have started a major effort to create a SQL extension, based on Apache Calcite.

We have published a press release and a blog post regarding the first stable release:

- <https://blogs.apache.org/foundation/entry/the-apache-software-foundation-announces12>
- <https://beam.apache.org/blog/2017/05/17/beam-first-stable-release.html>

We have also refreshed the design of our website.

Beam was covered at seven major industry conferences over the past quarter, including the "Apache: Big Data" conference in Miami, FL, where we have had 4 talks, a birds-of-a-feather

session and a social event. Additionally, we organized the first meetup in the Bay Area, hosted by Hortonworks and Future of Data.

Going forward, the main focus continues to be on the user growth, with outreach continuing across conferences and meetups. On the technical side, the next major milestone is the completion of the portability framework across all components of the project, which would, among other benefits, extend Python SDK to all Beam runners.

Health report:

The community continues to grow steadily, as follows:

- The number of contributors continues to increase. We are now at 179 unique code contributors, with 76 individuals contributing to the latest release alone (which spanned less than 2 months).
- Releases continue at a regular pace of 1-2 months per release.
- The activity on the user@ mailing list more than doubled.

PMC changes:

Currently 14 PMC members. No new PMC members have been added since graduation six months ago.

We are watching for potential new PMC members.

Committer base changes:

Currently 24 committers. Four new committers have been added since the last report:

- Aviem Zur was added as a committer on Fri Mar 17 2017.
- Chamikara Jayalath was added as a committer on Fri Mar 17 2017.
- Ismaël Mejía was added as a committer on Fri Mar 17 2017.
- Eugene Kirpichov was added as a committer on Fri Mar 17 2017.

Releases:

Since the last report, Apache Beam has published two releases:

- 0.6.0 was released on Mon Mar 13 2017.
- 2.0.0 was released on Mon May 15 2017.

Version 2.0.0 is the first release that comes with API stability guarantees. Going forward, we expect to publish a release every 1-2 months.

Mailing list activity:

Mailing list activity continues to increase across all metrics, with the number of user@ emails more than doubling compared to the previous quarter.

- dev@beam.apache.org
 - 424 subscribers (up 63 in the last 3 months).
 - 1162 emails sent to list (1094 in previous quarter).

- user@beam.apache.org
 - 384 subscribers (up 73 in the last 3 months).
 - 547 emails sent to list (250 in previous quarter).

JIRA activity:

JIRA activity continues to increase across all metrics, with the number of resolved issues nearly doubling.

- 725 JIRA tickets created in the last 3 months (542 in the previous quarter).
- 650 JIRA tickets closed/resolved in the last 3 months (347 in the previous quarter).