



Nov/Dec 2017 | Newsletter

What's been done



[Apache Beam 2.2.0 Release](#) (by: Reuven Lax)

- Apache Beam 2.2.0 Release

[Bundle Retry for the Python SDK DirectRunner](#) (by: María García Herrero)

- Beam Python is now more reliable, as it includes a mechanism for the DirectRunner to process bundles up to 4 times and makes sure there are no partial write-outs if a retry occurs.
- The feature is currently opt-in by using the `--direct_runner_bundle_retry` flag.
- Fixes <https://issues.apache.org/jira/browse/BEAM-2718>.

[Extension for Data Sketches](#) (by: Arnaud Fournier)

- A new extension library of Combiners using probabilistic data structures (data sketches).
- The first sketch ApproximateDistinct (aka HyperLogLog+) is already in master, CountMinSketch and t-digest are expected to arrive before the end of the year.
- Design doc:
https://docs.google.com/document/d/1Xy6g5RPBYX_Hadplr_2WrUeusiwl0Jo2ACI5PEOP1kc/edit
- Code: <https://github.com/apache/beam/tree/master/sdks/java/extensions/sketching>

What we're working on...

Loading...

[Portability](#) (by: Lukasz Cwik, Kenn Knowles, Henning Rohde, et al.)

- Defining and migrating existing source to a cross language representation to enable execution of Java/Python and future SDKs in a language agnostic manner.
- See s.apache.org/beam-runner-api & s.apache.org/beam-fn-api for further details.

What's planned



[2018 Planning Summit](#) (by: X,Y)

- We'll kick off 2018 planning via the mailing list on 12/15. After this, we'll host a planning call in early January to discuss and decide on the direction we want Apache Beam to take. The decisions of those two conversations will be shared via the mailing list.
- January 2018

New Contributor Experience (by: Gris Cuevas & Pablo Estrada)

- We'll be partnering with IBM to onboard a team of their developers to contribute to Apache Beam. This experience will be built and documented by Pablo Estrada and Gris Cuevas as a way to prepare our project for further growth.
- January 2018
- If you're interested in helping welcome the new folks reach out to Pablo or Gris

New Members

New Committers

- Etienne Chauchot, works at Talend and is based in Nantes France, became a committer on november 2nd, he helped migrate Nexmark benchmark suite to Beam and added tests validating runners, he created ElasticsearchIO and improved some others IOs, and created some transforms in the SDK.

New PMCs

- Ismaël Mejía (iemejia), Talend, France, PMC since November 6, Proudest job so far is to have mentored new contributors, as a consequence Beam has now the Hive and Solr connectors and the extension for data sketches.

Talks & Meetups

Apache Beam @ SF Cloud Mafia

- [Slides here](#)

#events-and-meetups Slack Channel

- [Join](#) if you're interested in hosting or speaking at events.

Big Data Meetup Singapore @ PayPal

- [Slides](#)

Apache Beam: Portable and efficient data-intensive processing @ Paris OSS

- [Description here](#)

Apache Beam presentation at Strata Singapore

Resources

Meetups and Events Knowledge Base

- If you're interested in giving a talk about Beam, check out our [knowledge base](#).

Contribution Guide

- Thinking of contributing to the project? Read our [Contribution Guide](#).

Featured Talks

- Check-out [Tyler's talk on Foundations of Streaming SQL @ Big Data Spain '17](#)

Happy Holidays, and Until Next Time!