None

Description:

The mission of Apache Beam is the creation and maintenance of software related to 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.

Project Status:

Current project status: ongoing

Issues for the board: none

Membership Data:

Apache Beam was founded 2016-12-20 (9 years ago)

There are currently 102 committers and 26 PMC members in this project.

The Committer-to-PMC ratio is roughly 4:1.

Community changes, past quarter:

- No new PMC members. Last addition was Danny McCormick on 2024-12-12.
- Derrick Williams was added as committer on 2026-04-28

Project Activity:

Recent releases:

- 2.74.0 was released on 2026-06-02.
- 2.73.0 was released on 2026-04-30.

First, highlighting major events:

- Beam College took place in late april
- Beam Summit will take place in a week

A theme of CDC work taking place in a few ways:

- Iceberg CDC support improved
- BigQuery CDC made more efficient
- BigQuery CDC streaming source for Python (previously just Java)
- Spanner ChangeStreams can union of TVF streams
- Persisting Debezium CDC offsets across pipeline restarts
- At the core of the Beam model, new per-element metadata being leveraged to make streaming of CDC data through all transforms a "default" way of processing

Compatibility / new versions / dropping dead versions

- added Flink 2.0 support
- added Spark 4 support
- added Python 3.14 support
- dropped Java 8 support
- minimum Go version to 1.26.1
- dropped SamzaRunner support

Google Summer of Code projects active & ongoing:

- Portable Kafka Streams Runner: https://issues.apache.org/jira/browse/GSOC-305
- Native Python Streaming Transforms: https://issues.apache.org/jira/browse/GSOC-315
- Test Optimization & Monitoring System Integration: https://issues.apache.org/jira/browse/GSOC-273

AI features:

 - Qdrant vector database sink

 - Google ADK Agent Handler

 - Huggingface and Tensorflow image embedding for RAG

Connectors:

 - Iceberg source had a ton of work

 - Delta Lake source designed and implemented

 - MQTT connector made available for Python use via multi-language pipelines

Other Designs

- Rust-based Python tooling: https://s.apache.org/beam-python-rust-tooling

- Python Memory Profiling Capabilities: https://s.apache.org/beam-python-memory-profiling

- AsyncDoFn for Java: https://lists.apache.org/thread/x9jbvjvy8njnljo57y1f9lmssltxk022

Community Health:

[Insert your own data here]

```
None
Goal: to prepare preliminary information for the Beam PMC and community to report to the
board

The rules:
 - filter all data to the date range 2026-04-01 to present
 - filter out anything already mentioned on
https://whimsy.apache.org/board/minutes/Beam.html
 - ground every statement with a URL citation

The tasks:
  - have a subagent read https://lists.apache.org/list.html?dev@beam.apache.org since
2026-04-01 (if you are blocked by robots.txt then search my gmail for
list:dev@beam.apache.org) and write you a summary of the most notable things
     - always link to the dev list thread
     - include all design documents, no matter what! include the s.apache.org shortlink if
one exists, otherwise include the google doc link
  - include any event (beam summit, meetups)
  - have a subagent read the commits in https://github.com/apache/beam since 2026-04-01 and
write you a summary of the most notable changes
     - include most everything from CHANGES.md
     - link to the pull request(s)

The finalization:
 - check that every URL loads and supports the citation
 - check that nothing is already mentioned at
https://whimsy.apache.org/board/minutes/Beam.html
```

Apache Beam PMC: Preliminary Board Report Information (Q2 2026)

This document synthesizes preliminary research for the Apache Beam PMC and community to assist in drafting the upcoming quarterly board report, covering the period from April 1, 2026 to the present (June 16, 2026).

All items have been filtered to exclude any data or milestones already reported in the April 15, 2026 Board Minutes (such as Beam 2.72.0, Radosław Stankiewicz's committership, ActiveMQ/Postgres upgrades, Go 1.26.1, Python 3.14 support, and preliminary GSoC topic ideas).

Every statement is fully grounded with a verified URL citation.

---

1. Mailing List Activity & Community Activity (dev@beam.apache.org)

Releases & Milestones

- **Apache Beam 2.73.0**: Successfully released and announced on April 29, 2026 (Thread: Beam 2.73.0 Release; Tag v2.73.0).
- **Apache Beam 2.74.0**: Unanimously approved on June 2, 2026 (Thread: [RESULT] [VOTE] Release 2.74.0, release candidate #3).
- **Dropping Java 8 Support**: Following formal phase-out notices, Beam officially dropped support for Java 8 starting with version 2.74.0 (Thread: [Followup] Drop Java 8 support - phase 3). Supported JVM runtimes are Java 11, 17, 21, and 25.

Membership Changes

- **New Committer**: Derrick Williams (derrick@apache.org) was welcomed as a new committer, announced April 28, 2026 (Thread: [Announce] New Committer: Derrick Williams).

Design Documents Shared on the List (Q2 2026)

- **Upgrading Beam Python Tooling to Rust-Based Packages**: Design proposal to migrate core Python SDK tooling to Rust-backed implementations (Thread, Design Doc: https://s.apache.org/beam-python-rust-tooling).
- **ForkedPCollections in Python SDK**: Proposal for returning augmented or secondary side-output PCollections in Python (Thread, Design Doc).
- **Portable Kafka Streams Runner**: Comprehensive architectural design and proposal for a portable runner backed by Kafka Streams (Thread, Design Doc, Proposal).
- **Adding Python Memory Profiling Capabilities**: Design detailing native memory profiling options in Python (Thread, Design Doc: https://s.apache.org/beam-python-memory-profiling).
- **Delta Lake Source for Apache Beam**: Design document for implementing a Delta Lake read source supporting checkpoints and Managed I/O (Thread, Design Doc: https://s.apache.org/beam-delta-lake-source).
- **Implementing AsyncDoFn to the Java SDK**: Proposal to implement an asynchronous DoFn wrapper in Java mirroring the Python SDK (Thread, Tracking Issue).
- **Disabling Build Isolation by Default in Python Container**: Proposal to disable pip build isolation in boot.go starting with Beam 2.75 to support restricted network execution (Thread).
- **GSoC 2026 Resource & Security Monitoring System**: Proposal for test optimization and security monitoring integration (Thread, Proposal PDF).

---

2. Events & Meetups

- **Beam College 2026**: Free online training event held in late April 2026 focusing on GenAI, RAG integration, Beam YAML, and Remote ML Inference (Thread, Website).
- **Data in Motion NYC (Messaging, Streaming & Processing)**: Community meetup held on June 4, 2026 in New York City leading into Beam Summit (Thread, Event Page).

- **Beam Summit 2026**: Annual summit scheduled for June 22–23, 2026 in New York City (St. John's Terminal Google offices) coinciding with the World Cup (Thread, Website).

---

3. Codebase Commits & Notable Changes (`apache/beam`)

Runners & Engine Updates

- **Spark 4 Runner**: Added Spark 4 runner support for the Java SDK (PR #38255).
- **Samza Removal**: Removed legacy Apache Samza runner support (PR #35448).
- **Dataflow State Tag Encoding v2**: Enabled state tag encoding v2 by default for new Dataflow Streaming Engine jobs (PR #38705).
- **Pipeline Drain Indicator**: Introduced an indicator for aggregations and timers firing during a pipeline drain to recognize and handle partial data (PR #36884).
- **Provisioned Disk Settings**: Added support for setting disk provisioned IOPS and throughput in Dataflow via `--diskProvisionedIops` and `--diskProvisionedThroughputMibps` (PR #38349).
- **Memory Profiling**: Added instrumentation supporting memory profiling with Memray and off-the-shelf profiling agents (PR #38853).
- **Beartype Default**: Made Beartype the default fallback type checking tool in the Python SDK (PR #38275).

I/O Connectors

- **Delta Lake IO**: Added support for reading from Delta Lake in Java (PR #38551).
- **IcebergIO Dynamic Sort Order**: Added support for declaring table sort orders on dynamic table creation via `sort_fields` (PR #38269).
- **IcebergIO Hash Distribution**: Added support for writing with hash distribution mode and autosharding (PR #38061).
- **DebeziumIO CDC Offsets**: Added `OffsetRetainer` interface and `FileSystemOffsetRetainer` to persist and restore CDC offsets across pipeline restarts (PR #28248).
- **ClickHouseIO DateTime64**: Added support for writing `DateTime64` columns with sub-second precision (PR #38466).
- **BigQuery CDC Source**: Added BigQuery CDC streaming source in Python (PR #37724).
- **Google ADK Agent Handler**: Added `ADKAgentModelHandler` for running Google ADK agents in Python (PR #37917).
- **Qdrant Vector DB**: Added Qdrant `VectorDatabaseWriteConfig` implementation in Python (PR #38141).
- **RAG Image Embeddings**: Added image embedding support in `apache_beam.ml.rag` (PR #37628).
- **BigQuery Storage Write Type Overrides**: Added `type_overrides` parameter to `WriteToBigQuery` supporting custom Python to BigQuery type mappings (PR #25946).

SDK & SQL Transforms

- **BatchElements (Java)**: Added `BatchElements` transform for the Java SDK (PR #38369).
- **Lineage Plugin Mechanism**: Added plugin mechanism for different Lineage implementations in Java (PR #36790).
- **Python Types in Beam SQL**: Supported picklable Python user types in Beam SQL rows (PR #20738).
- **Python Row Registration**: Introduced `register_row` as preferred API for named tuples and dataclasses (PR #38108).
- **Exception Chaining**: Added exception chaining to preserve error context in enrichment handlers and core transforms (PR #37422).
- **Disabling Build Isolation Option**: Added `--experiments=pip_no_build_isolation` option in Python (PR #37331).
- **File-Based Pipeline Options**: Container `boot.go` now passes pipeline options via a file instead of environment variables (PR #37370).