2024-01-08 Arrow ASF Board Report

Arrow PMC Chair Note: Please add any relevant comments/content to this document. Neal Richardson will submit to the ASF board on January 8, 2025 (about one week prior to the scheduled board meeting).

The rationale and process for this report: https://www.apache.org/foundation/board/reporting
Past report: 📄 2024-10-09 Arrow ASF Board Report

The metrics in this report are derived from https://reporter.apache.org/wizard/?arrow

Description:
The mission of Apache Arrow is the creation and maintenance of software related to columnar in-memory processing and data interchange. More information can be found at https://arrow.apache.org/overview/

Project Status:

Current project status: Ongoing (high activity)

Issues for the board:

- None

Membership Data:

Apache Arrow was founded 2016-01-19 (9 years ago)
There are currently 112 committers and 53 PMC members in this project.
The Committer-to-PMC ratio is roughly 7:4.

Community changes, past quarter:
- Curt Hagenlocher was added to the PMC on 2024-10-29
- Gang Wu was added to the PMC on 2024-12-02
- Adam Reeve was added as committer on 2024-11-18
- Laurent Goujon was added as committer on 2024-11-25
- Rossi Sun was added as committer on 2024-10-22

Project Activity:

Selected subproject updates:

Arrow C Data Interface

The proposed asynchronous version of the C Data Interface was merged: https://github.com/apache/arrow/pull/43632

Format

We've added a new experimental standard for statistics: https://arrow.apache.org/docs/dev/format/StatisticsSchema.html

Go

We've released a new version from the split repository (apache/arrow-go). We'll be able to maintain the Go implementation in this development style.

Java

Like Go, the Java implementation has been moved to its repository to make independent releases. This will not require users to make any changes, but (like Go) will need work to set up independent CI and release processes.

MATLAB

Initial support for reading and writing the Arrow IPC Stream Format was added. This work was done in part to enable [Arrow-over-HTTP](https://github.com/apache/arrow-experiments/tree/main/http/get_simple) workflows.

nanoarrow

Version 0.6.0 was added, which included an IPC writer, CMake and meson configuration improvements, and DLPack/non-CPU device support in the Python bindings.

R

A CVE was filed for arbitrary code execution when loading a malicious data file: https://www.cve.org/CVERecord?id=CVE-2024-52338

Rust

The Rust implementation continues monthly releases, with breaking API changes restricted to every 3 months. This process seems to be working well and we have had less community

pushback due to breaking changes. We continue to encourage additional community help maintaining / reviewing PRs, though PRs are still bottlenecked by reviewer bandwidth.

Overall the focus has been on filling out DataType support in kernels, and interoperability with other implementations (Dictionary handling, ListArray field naming, etc.)

Most work for the Rust object store has been related to incremental features, and small bug fixes. We are prioritizing stability.

Recent Releases:

- RS-OS-0.11.2 was released on 2024-12-24.
- RS-54.0.0 was released on 2024-12-23.
- JULIA-2.8.0 was released on 2024-12-06.
- 18.1.0 was released on 2024-11-21.
- RS-53.3.0 was released on 2024-11-20.
- ADBC-15 was released on 2024-11-13.
- JULIA-2.7.4 was released on 2024-10-30.
- 18.0.0 was released on 2024-10-28.
- RS-53.2.0 was released on 2024-10-24.
- GO-18.0.0 was released on 2024-10-22.
- RS-OS-0.11.1 was released on 2024-10-20.
- NANOARROW-0.6.0 was released on 2024-10-13.

Community Health:

GitHub statistics show a 20-30% drop in activity this quarter:

github@arrow.apache.org had a 30% decrease in traffic in the past quarter (10460 emails compared to 14773)
1878 commits in the past quarter (-21% change)
1616 PRs opened on GitHub, past quarter (-27% change)
1666 PRs closed on GitHub, past quarter (-22% change)
1034 issues opened on GitHub, past quarter (-31% change)
899 issues closed on GitHub, past quarter (-23% change)

Given that the number of commits is down less than the traffic on the `github@` mailing list, we can infer that this means that PR review activity was more significantly reduced: there were fewer issues, PRs, and commits, and those that did occur got less discussion.

Dev mailing list traffic is stable/slightly increased, as is the number of unique code contributors.

This is all consistent with the effects of the major layoffs of paid Arrow contributors mentioned in the previous board update. If that is in fact the cause, we should expect that decrease to stabilize in the next quarter. Moreover, some of those affected by the layoffs have found Arrow-related positions in other companies (with at least one open source company stepping up to provide maintenance for Arrow), so we may see these numbers recover in the future.