Tab 1

Hamilton ASF Incubator Proposal

Abstract

Hamilton is a lightweight python directed acyclic graph (DAG) orchestration framework used for defining and executing transformation pipelines that comes with a self-hostable observability UI.

Proposal

Hamilton is a lightweight in-process framework to define, execute, and observe directed acyclic graphs (DAGs) that express data transformations. In Hamilton one can express complex DAGs of transformations, e.g. from dataframe transformations (using pandas, polars, PySpark), machine learning pipelines, through to regular software engineering API request and LLM API based workflows. Observability hooks are built into the framework. The Hamilton UI is a self-hostable service to capture observability output from workflow runs. Apache Software Foundation incubation will establish Hamilton as a community-driven standard.

Background

Modern AI and data applications require structured, observable, and maintainable execution of computational graphs. However, existing solutions often lack modularity, are vendor-specific, or focus primarily on large-scale (macro) workflow orchestration rather than composable, fine-grained execution models (i.e. micro-orchestration).

Hamilton was designed to be a modular and composable framework for structuring and observing computational graphs expressed as a directed acyclic graph (DAG). Hamilton provides a function-based approach to defining DAGs for data, ML, and Generative AI pipelines, allowing teams to build structured and maintainable transformations across various execution environments, from Pandas to PySpark, and from notebooks to web-services. The framework runs anywhere a python process can run. The optional observability system can be run as a separate process or hosted via docker.

Hamilton was created by Stefan Krawczyk and Elijah ben Izzy, who have extensive experience in platform tooling to operationalize data, machine learning and AI systems. Hamilton was originally conceived at Stitch Fix in 2019, it was open sourced in 2021, forked in 2023 (with the original repository archived and all committers moving to the new repository). The project was designed to enable teams to move faster in collaboration with their platform counterparts.

Fast forward to March 2025, and data, machine learning, and generative AI needs have exploded, but there are no truly vendor neutral set of frameworks that address their challenges. With a growing contributor and user list, we're looking to take this project to the next level by making sure that efforts remain community driven and vendor neutral.

Rationale

The need for vendor-neutral, open-source frameworks for managing computational graphs has never been greater. Machine learning, generative AI, and data-driven applications increasingly rely on structured workflows to ensure reproducibility, observability, and maintainability. However, existing solutions are either too heavyweight (e.g., Apache Airflow for workflow orchestration), tied to proprietary ecosystems that limit flexibility and adoption, or require custom in-house code / frameworks that end up accumulating technical debt and slowing down projects over time.

Hamilton fills a gap by providing a lightweight, Python-native framework that standardizes the expression of DAGs, it promotes software engineering best practices to enable faster iteration cycles, and comes with self-hostable observability:

- **Testing:** both frameworks enable test driven approaches because each framework's paradigm guides developers to structure code in a unit and integration testing friendly manner.
- **Lineage & provenance**: by modeling workflows as a graph, the emission and introspection of what happened is a simple property of the graph that the frameworks provide access to.
- **Modularity & reuse:** both frameworks aim to decouple execution and platform concerns from the actual logical flow of a graph. This decoupling helps avoid technical debt and ensures code can be reused or more easily modified to meet the needs of evolving organizations and software projects.

By bringing Hamilton into the Apache Software Foundation (ASF), we aim to:

1. **Establish Open Standards** – Current solutions lack a widely accepted, open-source standard for structuring computational graphs, particularly for AI-driven workflows. ASF incubation will help position Hamilton as foundational tools in this space.
2. **Foster Community-Driven Development** – ASF's governance model will enable a diverse contributor base, ensuring sustainable growth and innovation. Hamilton has already gained traction, and ASF can provide the structure needed for broader enterprise adoption.

3. **Promote Vendor-Neutrality** – Many existing graph execution and orchestration tools are tied to commercial platforms. ASF incubation will ensure that Hamilton remains open, interoperable, and accessible to all developers and organizations.
4. **Encourage Adoption Across AI and Data Ecosystems** – With the growing complexity of AI and data engineering, lightweight, transparent execution frameworks are critical. Hamilton simplifies structured DAG computation without the overhead of full-fledged workflow orchestrators.

Initial Goals

The initial goals of bringing the Hamilton project into the ASF are several-fold:

- Move the existing codebases to Apache and integrate it with the Apache development process.
    - https://github.com/dagworks-inc/hamilton
    - https://github.com/dagworks-inc/tryhamilton (source for www.tryhamilton.dev)
- Set up the governance structure of the project.
- Plan for incremental development, and releases along with the Apache guidelines.
- Ensure all donated code is appropriately licensed under the Apache License.
- Attract contributions from the greater Apache and Python communities.
- Building out the observability and annotation capabilities to help drive faster SDLCs for data, machine learning, and GenAI workflows.
- Build out better execution support on top of a wide variety of platforms and cloud infrastructures, e.g. Ray, Apache Airflow, AWS, GCP Vertex, etc.

Current Status

Since open sourcing Hamilton on GitHub, we have been fostering a growing community.

Hamilton Metrics:

- 2K+ stars, 500K+ downloads, 500+ community in slack - users from banks to start-ups, 130+ forks
- 60+ contributors from around the globe
- Over 800 Pull Requests and 300 issues
- Four other open source libraries are built on top of / for Hamilton.
- 100+ members in the Hamilton meet-up group.

What needs to be addressed:

- Complete license change to Apache 2.0 from BSD-3 for Hamilton (PR will be linked here)
- Addressing one javascript UI dependency that is Eclipse 2.0 licensed.

- Moving project decision discourse to public mailing lists versus split across GitHub and Slack.

Meritocracy

Hamilton follows a meritocratic model where project direction & contributions—whether in code, documentation, or community support—are recognized and rewarded. Contributors who demonstrate sustained, high-quality contributions along with interest to do more, will be invited to become committers, and active committers may join the PMC through consensus.

We foster an inclusive environment by providing clear contribution pathways, encouraging mentorship, and making decisions transparently via public github issues and discussions; we will transition to using mailing lists as part of the transition to ASF. By aligning with ASF's governance model, we will ensure long-term sustainability and an open, community-driven project.

Community

Since inception, we've been building an active slack community of 500+ members. Hamilton also has a virtual meet-up group of 100+ members. We will use these as the basis for creating the new community at Apache.

Planned Community Initiatives

To nurture and expand the Hamilton community at Apache, we are planning:

- Regular Meetups & Office Hours: In-person and virtual gatherings to connect contributors and users, share knowledge, and foster collaboration.
- Virtual Conferences: Dedicated events to showcase use cases, provide tutorials, and encourage community participation.
- Workshops and Webinars: Educational sessions to onboard new contributors and empower existing users.

Core Developers

The core development team consists of individuals with a breadth of experience prototyping, and productionizing data, machine learning, and generative AI production applications. These individuals range from PhD students through to industry veterans.

Alignment

There is no vendor neutral alternative to popular generative AI frameworks. At the same time Generative AI frameworks aren't general purpose enough to cover other use cases. Hamilton's design philosophy is to enable users to logically model their workflows across data, machine learning, and generative AI workflows that decouple platform concerns from application logic. This approach aligns well with the spirit of community driven and vendor neutrality projects Apache sponsors. We also look forward to tighter integrations with other Apache projects like Apache Airflow.

Known Risks

Project Name

The name **Hamilton** has been in use within the open-source community and we do not have known conflicts with existing trademarks or software projects (e.g. we have good SEO results for "hamilton python"). A preliminary search has not identified legal or branding issues, but we will work with the Apache Software Foundation to conduct a thorough trademark review during incubation.

While the name is a reference to historical figures, its usage in a technical context is distinct and unlikely to cause confusion. If any naming conflicts arise, we are open to adjustments to ensure compliance with ASF guidelines.

Orphaned Products

First, the core developers will continue to work on Hamilton. Additionally, Hamilton is already used by Best Egg, Rippling, Coinbase, Wealth.com, Capitec Bank, and many other companies across the world. With an active community on Slack, continued stream of interested enterprises (e.g., Wells Fargo, Nvidia), and steady weekly download numbers, the risk of Hamilton being deprecated is minimal. The project is widely relevant supporting use cases across data science, machine learning, analytics, and generative AI.

Inexperience with Open Source

Hamilton has been an open source project for 3.5 years. All contributors are familiar with the open source development process.

There will be an adjusting period to learn and practice ASF's approach to release management, etc., as the core committers have never been an official committer or PMC on an ASF project. With the right mentorship we don't foresee any issues here.

Length of Incubation

Given the maturity of Hamilton's codebase, size of community and the experience of its committers with the ASF, we expect to graduate to TLP within 1 year.

Homogenous Developers

Over the years the main committers to the project have been Elijah ben Izzy and Stefan Krawczyk. In the past year we have been focusing on bringing more outside contributors to the core internals of each project, with success - total core internal code contributors now totals six persons from various companies and backgrounds. We plan to continue to expand this so that more committers are familiar with the internals of both projects to enable more informed community decision making.

Reliance on Salaried Developers

It is expected that Hamilton development will occur on both salaried time and on volunteer time, after hours. The majority of initial committers used the project in their day to day and were allowed by their employer to contribute to this project. However, in the case that we lose salaried development, we are confident that the project will continue even if no salaried developers contribute to the project. We are committed to recruiting additional committers including non-salaried developers.

Relationships with Other Apache Projects

Currently Hamilton has a plugin for Apache Spark.
Otherwise depending on the project direction, there are projects where tighter integration could be possible:

- Apache Airflow (enabling one to more easily run Hamilton with Airflow)
- Apache Druid (as a backend for metrics / observability for the projects)
- Apache Pinot (as a backend for metrics / observability for the projects)
- Apache Spark (more work for debugging and optimization of Spark Jobs using Hamilton)
- Apache NiFi (as a means to execute Hamilton graphs)
- Apache Parquet & Apache Iceberg (as a means to store observability into)

An Excessive Fascination with the Apache Brand:

We acknowledge that the Apache brand is one of the most recognized and respected in the open-source domain. However, the primary goal of this proposal is not to seek publicity but to expand the developer base and foster a broad, diverse community around Hamilton. We believe that successful open-source projects thrive on inclusivity and collaboration, and becoming an Apache project provides a clear and proven framework to attract and nurture new contributors.

The core development team has been deeply committed to Hamilton over the past few years, addressing the fast-growing needs of iteration and observability in today's AI landscape. While we remain dedicated to continuing its development regardless of its status, we see this as an opportunity to give back to the Apache community. Having benefited extensively from various Apache libraries and projects throughout our careers, we view this proposal as a way to contribute to the vibrant ecosystem that has supported so many innovative technologies.

Documentation

Hamilton:

- Documentation - https://hamilton.dagworks.io/en/latest/
- Interactive tutorial - https://www.tryhamilton.dev/
- Reusable modules - https://hub.dagworks.io/
- Blog - https://blog.dagworks.io/

Initial Source

Hamilton:

- https://github.com/stitchfix/hamilton (archived & forked March 2023) → https://github.com/dagworks-inc/hamilton
- https://github.com/dagworks-inc/tryhamilton (source for www.tryhamilton.dev)

Source and Intellectual Property Submission Plan

The Hamilton codebase is currently hosted on Github: https://github.com/dagworks-inc/hamilton. This is the exact codebase that we would migrate to the Apache foundation. Only 15% of the original BSD-3 licensed pre-forked code base remains as determined by git blame; we don't foresee IP issues by switching to the Apache 2.0 license.

DAGWorks Inc. would also donate https://github.com/dagworks-inc/tryhamilton (www.tryhamilton.dev) which is not open source right now. Along with the www.tryhamilton.dev domain (if it makes sense).

While not formally using CLAs, we have had no friction when asking contributors about relicensing from BSD-3 to Apache 2.0. If this proposal passes the conditions for being an Incubator project under the ASF, DAGWorks Inc. would transition source code ownership

to the ASF via the Software Grant Agreement. All core contributors would then sign Apache CLAs.

External Dependencies

With python software, by default dependencies are not packaged or bundled into a single binary. Hamilton doesn't bundle any other code. To operate the python portion these dependencies span MIT, Apache 2.0, BSD, and PSF licenses, which are all compatible with Apache 2.0 usage.
sf-hamilton - no bundled dependencies
sf-hamilton-contrib - this contains code that is also MIT licensed and is appropriately marked in source files.
sf-hamilton-sdk (tracker to use sf-hamilton-ui) - no bundled dependencies
sf-hamilton-ui does include bundled JavaScript dependencies:

package	license	repo	Notes
@headlessui/react	MIT	https://github.com/tailwindlabs/headlessui.git	
@heroicons/react	MIT	https://github.com/tailwindlabs/heroicons.git	
@propelauth/react	MIT	https://github.com/PropelAuth/react.git	
@reduxjs/toolkit	MIT	https://github.com/reduxjs/redux-toolkit.git	
@tailwindcss/forms	MIT	https://github.com/tailwindlabs/tailwindcss-forms.git	
@tailwindcss/typography	MIT	https://github.com/tailwindlabs/tailwindcss-typography.git	
@testing-library/jest-dom	MIT	https://github.com/testing-library/jest-dom.git	
@testing-library/react	MIT	https://github.com/testing-library/react-testing-library.git	
@types/moment-timezone	MIT	https://github.com/moment/moment-timezone	
@types/react	MIT	https://github.com/DefinitelyTyped/DefinitelyTyped.git	
@types/react-dom	MIT	https://github.com/DefinitelyTyped/DefinitelyTyped.git	
@types/react-router-hash-link	MIT	https://github.com/DefinitelyTyped/DefinitelyTyped.git	
@types/react-select	MIT	https://github.com/JedWatson/react-select/tree/master	
@types/redux-persist	MIT	https://github.com/rt2zz/redux-persist	
@uiw/react-json-view	MIT	https://github.com/uiwjs/react-json-view.git	
chart.js	MIT	https://github.com/chartjs/Chart.js.git	
chartjs-adapter-date-fns	MIT	https://github.com/chartjs/chartjs-adapter-date-fns.git	

chartjs-plugin-zoom	MIT	https://github.com/chartjs/chartjs-plugin-zoom.git	
date-fns	MIT	https://github.com/date-fns/date-fns.git	
dayjs	MIT	https://github.com/iamkun/dayjs.git	
elkjs	EPL-2.0	https://github.com/kieler/elkjs.git	We need to figure out how to drop this dependency or make it optional.
fuse.js	Apache-2.0	https://github.com/krisk/Fuse.git	
headlessui	MIT	https://github.com/tailwindlabs/headlessui	
http-proxy-middleware	MIT	https://github.com/chimurai/http-proxy-middleware.git	
moment-timezone	MIT	https://github.com/moment/moment-timezone.git	
posthog-js	MIT	https://github.com/PostHog/posthog-js.git	
prism-react-renderer	MIT	https://github.com/FormidableLabs/prism-react-renderer.git	
react	MIT	https://github.com/facebook/react.git	
react-chartjs-2	MIT	https://github.com/reactchartjs/react-chartjs-2.git	
react-diff-viewer-continued	MIT	https://github.com/aeolun/react-diff-viewer-continued	
react-dom	MIT	https://github.com/facebook/react.git	
react-draggable	MIT	https://github.com/react-grid-layout/react-draggable.git	
react-icons	MIT	https://github.com/react-icons/react-icons	
react-markdown	MIT	https://github.com/remarkjs/react-markdown.git	
react-micro-bar-chart	MIT	https://github.com/KyleAMathews/react-micro-bar-chart.git	
react-redux	MIT	https://github.com/reduxjs/react-redux.git	
react-router-dom	MIT	https://github.com/remix-run/react-router.git	
react-router-hash-link	MIT	https://github.com/rafgraph/react-router-hash-link.git	
react-scripts	MIT	https://github.com/facebook/create-react-app.git	
react-select	MIT	https://github.com/JedWatson/react-select.git#master	
react-syntax-highlighter	MIT	https://github.com/react-syntax-highlighter/react-syntax-highlighter.git	
react-tailwindcss-datepicker	MIT	https://github.com/onesine/react-tailwindcss-datepicker	
react-time-ago	MIT	https://gitlab.com/catamphetamine/react-time-ago.git	

react-visibility-sensor	MIT	https://github.com/joshwnj/react-visibility-sensor.git	
reactflow	MIT	https://github.com/xyflow/xyflow.git	
redux-persist	MIT	https://github.com/rt2zz/redux-persist.git	
tailwind-scrollbar-hide	MIT	https://github.com/reslear/tailwind-scrollbar-hide.git	
tailwindcss-question-mark	MIT	https://github.com/GavinJoyce/tailwindcss-question-mark.git	
typescript	Apache-2.0	https://github.com/microsoft/TypeScript.git	
web-vitals	Apache-2.0	https://github.com/GoogleChrome/web-vitals.git	

Cryptography:

Not applicable

Required Resources

Mailing lists

Note since this is a two project proposal we think a single set of mailing lists for both would be best to start.

We plan to use the following mailing lists (TBD based on name):

- private@hamilton.apache.org (with moderated subscriptions)
- dev@hamilton.apache.org
- commits@hamilton.apache.org
- users@hamilton.apache.org

Git Repositories

We would like to use Git for source code management and enable GitHub mirroring functionality. We need the following Git repos:

- https://github.com/apache/hamilton
- https://github.com/apache/hamilton-101 (for tryhamilton.dev)

Issue Tracking

The community would like to continue using GitHub Issues.

Other Resources

At this moment, no special resources or infrastructure are required.

Initial Committers

TODO: send Apache CLAs.

Stefan Krawczyk, stefank at cs dot stanford dot edu

Elijah ben Izzy, elijah_ben_izzy at alumni dot brown dot edu

Thierry Jean, thierry dot jean dot ac at gmail dot com

Jernej Frank, jernej dot frank at physics dot ox dot ac dot uk

Charles Swartz, charles dot w dot swartz do vi at gmail dot com

Ryan Whitten, ryan dot whitten at bestegg dot com

Sponsors

Affiliations of the above committers are as follows:
Stefan Krawczyk, DAGWorks Inc.
Elijah ben Izzy, DAGWorks Inc.
Thierry Jean, dltHub
Charles Swartz, QUANTICS Inc.
Jernej Frank, Oxford University
Ryan Whitten, Best Egg Inc.

Champion

- TBD

Nominated Mentors

- PJ Fanning, fanningpj at apache dot org
- Kevin Ratnasekera djkevincr at apache dot org
- Ayush Saxena ayushtkn at gmail dot com

Sponsoring Entity

The Apache Incubator