
APACHE FLINK'S® JOURNEY FROM ACADEMIA INTO THE ASF

FABIAN HUESKE, APACHE FLINK PMC MEMBER

dataArtisans

Apache®, the names of Apache projects and their logos, and the multicolor feather logo are registered trademarks or trademarks of The Apache Software Foundation in the United States and/or other countries.

WHAT IS APACHE FLINK?

- Open Source project of the Apache Software Foundation
- Software to process data streams with low latency at scale
 - Programming interfaces / APIs
 - Execution engine
- Running on clusters of 2 to 10000s of machines
- Backed by an active and helpful community
 - 100s of people, several big companies



COMMON USE CASES FOR APACHE FLINK

- Fraud detection (financial transactions, trading, ...)
- Security & Anomaly Detection (network intrusion, compliance, bots, ...)
- Real time analytics (dashboards, infrastructure and process health, ...)
- Machine learning (search, recommendations, ETA models, ...)
- Event-driven Applications (micro services, websites, inventory keeping...)
- Or just crunching streaming data in general...



POWERED BY APACHE FLINK



HOW IT ALL BEGAN



2010: THE STRATOSPHERE RESEARCH PROJECT

- Research project funded by the Deutsche Forschungsgemeinschaft
- Mission: "Develop the next generation Big Data Analytics platform"
 - Fuse state-of-the-art data processing technology (MapReduce & DBMS)
- Five research groups at TU Berlin, HU Berlin, and HPI Potsdam
- Additional funding from EIT ICT Labs, EU FP7 & Industrial Partners
- 40+ scientific publications (first in 2009)



2013: BYE BYE ACADEMIA

- End of 2013, the first generation of PhD students was about to leave
- The state of the prototype
 - 3+ years of development
 - System is the state-of-the-art and offers unique and distinguishing features
 - Released and available under Apache License 2
- How to continue?
 - Don't want to abandon the project.
Continue developing the project as Open Source?
 - How to find a job in Berlin to work on similar technology?



2013: HOW TO BOOTSTRAP AN OS PROJECT?

- Releasing software as open source is one thing
 - Getting people to use it another...
 - Growing a community and attracting people to join the project another...
 - Figuring out licensing and setting up the infrastructure another...
 - Ensuring legal framework and contributor protection another...
- Lots of challenges to tackle
 - Open source organizations help to address these issues
- The Apache Software Foundation was a natural choice for us



HOW WE GOT IN TOUCH WITH THE ASF

- Several ASF people worked with us or visited us.
- We spoke at events with a focus on Open Source projects.
- Many related data infrastructure systems are Apache projects.



Apache®, the names of Apache projects and their logos, and the multicolor feather logo are registered trademarks or trademarks of The Apache Software Foundation in the United States and/or other countries.



WHAT IS THE APACHE SOFTWARE FOUNDATION?



WHAT IS THE APACHE SOFTWARE FOUNDATION?

- *“The mission of the Apache Software Foundation (ASF) is to provide software for the public good.
We do this by providing services and support for many like-minded software project communities of individuals who choose to join the ASF.”*
- The ASF is a nonprofit charitable organization, US 501(c)(3)
 - Established in 1999.
- Provides a legal framework for Open Source projects
 - Intellectual property, financial contributions, software license
- The home of more than 350 individual Open Source projects



PRINCIPLES OF THE APACHE SOFTWARE FOUNDATION

- All ASF projects follow the *Apache Way*
- Community over Code
 - Users, contributors, committers, PMC members
 - All contributors act as individuals
- Project decisions by people who contributed to the project
 - Not only code contributions!
 - Communication happens on open and archived mailing lists
- Software is released under the liberal Apache License 2 (AL2).



CREATING FLINK AND DATA ARTISANS



2014: APACHE FLINK IS BORN

- The core Stratosphere developers decided to donate the code to the ASF
 - In April, the Stratosphere code base enters the ASF Incubator
- A new name: Apache Flink!
 - German for nimble, swift
 - Name had to be changed due to trademark issues
 - Cute squirrel logo
- Stratosphere developers (PhD & master students) and ASF mentors are the initial committers



2014: INCUBATION AND GRADUATION

- Setting up the project's infrastructure and solving licensing issues
- Growing the community
 - Five committers added during incubation
 - Students from affiliated EU partners (Stockholm, Budapest) or TU Berlin
- Receiving code contributions and publishing releases
 - Adding a stream processing API & runtime
 - The first official ASF release in August 2014
- Flink graduates from the Incubator as a top-level project in December 2014



2014: FOUNDING DATA ARTISANS

- Start-up founded by Stratosphere core team & Flink committers
- Based in Berlin Kreuzberg
- Venture capital funded



2014: DATA ARTISANS' (INITIAL) GOAL

- 100% of company resources devoted to foster Apache Flink
- Grow an Open Source community
 - Help users, mentor contributors & add committers
 - Gain mindshare and traction in the Open Source Big Data ecosystem
- Develop Apache Flink
 - Extending Flink's feature set
 - Improving Flink's usability
 - And of course fixing bugs



GROWING A COMMUNITY



2014: OUR "COMPETITIVE" LANDSCAPE

- There is a clear need for our technology
 - More advanced analytical big data tools are required
- Several other projects in the space: Spark, GraphLab, Impala, H2O, ...
- Other projects backed by big companies or massive VC funding (>10x our funding)
- Other projects come out of US/Silicon Valley tech scene.
- How can we establish a vital community & successful project with
 - 7 people in Berlin (who never did an OS project before) and
 - comparatively little funding?



2015: WORKING ON GAINING TRACTION

- Spreading the word about Flink
 - Speaking at conferences and meetups
 - Writing blog posts
 - Find early users/partners, grow them into evangelists
- data Artisans organizes a Flink-focused conference “Flink Forward”
 - Organizes hackathons and trainings
 - Good contents and crowd, seeded to a good extent from academic partners
- Flink gains a few early adopters
 - ResearchGate, King.com, Bouyges Telecom



2015: FINDING THE NICHE - STREAM PROCESSING

- Competing against bigger players on their turf is very hard
 - Flink is considered as technologically interesting or even superior,
 - But less mature/adoption
- Being just better at the same thing does not cut it (unless 10x better)
 - Community, mindshare, hype make up for quite a bit of technological lack
 - Technology is easier to advance than growing mindshare
- Stream processing (real-time data) emerging as a new trend
 - Existing solutions not adequate, Flink handles such use cases uniquely well



2016: THE TURNING POINT

- In late 2016, the adoption of Flink speeds up
- Flink pioneers many important stream processing features in OS
- More and more companies use Flink for business critical tasks
 - Flink gets more production exposure, feature requests, and bug reports
- Employees of many companies that use Flink start contributing
 - Present their use cases at conferences and meetups
 - Active contributors become committers
- Breakthrough users: Alibaba, Netflix, Uber



APACHE FLINK TODAY

- Community
 - 400+ contributors, 38 committers, 10k+ downloads per month
 - Among the top 10 ASF projects with the most mailing list traffic in 2017
- Adoption
 - Some of the world's largest stream processing applications run on Flink
 - Users funnel trillions of events / petabytes of data through Flink per day



LESSONS LEARNED



THE ASF AND SEED COMMUNITY

- Going with ASF was a really good choice
 - Gives legal protection and helps with exposure/mindshare
- Have an initial community from academia helped a lot in the beginning
 - Students at TU Berlin and EU partners
 - Grow a community and promote contributors to committers
 - Bootstrap our own conference with a respectable audience



GROWING AN OPEN SOURCE COMMUNITY

- Growing a community is a lot of work, but the most important thing
 - Community trumps technology to a large extent
- Finding the right niche turned out to be essential
 - Adding value that no other system provided at that point,
 - Being the first of a kind is important, unique technology also attracts community
- Enterprise (Open Source) software a competitive space
 - Projects without full-time developers have little chance
 - A successful project needs a vital community and a backing entity.



TODAY'S CHALLENGES

- Keeping up with the growth of the community
- Maintaining a community is a lot of work
 - Helping users
 - Discussing roadmaps, feature requests, designs
 - Supporting contributors, review contributions
 - On-boarding new committers
 - Balance user requests and our own long-term vision for Flink



WORKING ON OPEN SOURCE IS VERY REWARDING

- Fun way of developing, interacting directly with users
- Makes you proud to see your work being used in cool ways you never imagined before



FLINK FORWARD



organized by **dataArtisans**

The Apache Flink® Conference
Stream Processing | Event Driven | Real Time

3 SEPTEMBER 2018: TRAINING
4-5 SEPTEMBER 2018: CONFERENCE

BERLIN, GERMANY

Register at berlin.flink-forward.org

Early bird prices available until June 22

@dataArtisans #flinkforward



THANK YOU!

@fhueske

@ApacheFlink & @TheASF

@dataArtisans

dataArtisans