

How Apache works

JB Onofré <jbonofre@apache.org>



HUAWEI

Who am I

JB Onofré <jbonofre@apache.org>
@jbonofre | <http://blog.nanthrax.net>

- **Fellow**/Software Architect at Talend
- **Member** of the Apache Software Foundation (started in 2004)
- **PMC member on ~ 20 Apache projects** from system integration (Karaf, Camel, ActiveMQ, Archiva, Aries, ServiceMix, ...) to big data (Beam, CarbonData, Lens, Falcon, ...)



Why open source ?

For Companies & Organisations

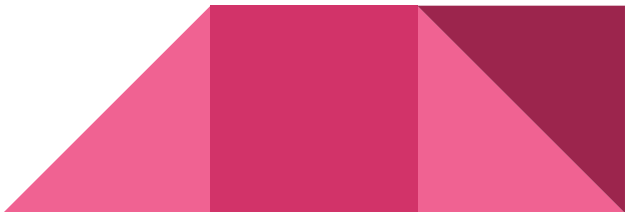
- **Impact on development and direction**
- Sense of membership in a community
- **Save on expensive resources and go to market**
- Ability to focus on what differentiates yourself
- Allows to nimbleness and agility
- **Increased vision, revenue and market share**



Why open source ?

For developers

- **Not alone, part of a community**
- Personal satisfaction
- Sense of accomplishment
- **Innovation: you do it !**
- Tinker !



Why open source ?

For users

- Access to code source
- **No vendor lock-in**
- **Better software, more reliable (security)**
- Much more nimble development (frequent releases)
- Direct user input
- **Open standards**



Open source promises

Let's pretend software is a cookie



=



Open source promises

Use !



Open source promises

Modify !



Open source promises

Share !



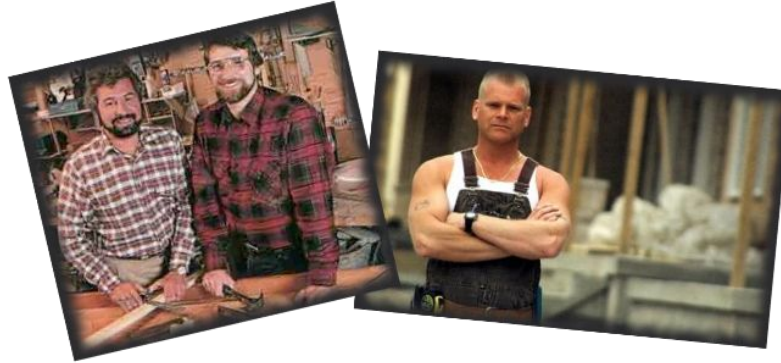
Open source licenses

Give me credit ! (Apache, BSD, MIT)



Open source licenses

Give me fixes (LGPL, EPL, MPL)



Open source licenses

Give me everything (GPL)



Open source community

Walled Garden (company community)

"All your base are belong to us."



Open source community

Meritocracy (open community)

"Out of chaos comes order."



Open source community

Building community

- Use mailing list
- Drive and accept consensus
- Avoid poisonous people
- Play nice, share, have fun



Apache Software Foundation



The Foundation

- Non-profit corporation
- 501(c)3 charity
- Volunteer organization
- Virtual world-wide organization
- Exists to provide the organizational, legal and financial support for various OSS projects
- Biggest open source foundation



From Apache group to the ASF

- Apache Group
- Informal corporate structure
- 8 members
- Resumed the work on NCSA httpd in Feb 1995 (“a patchy” server)
- Choose permissive licensing



ASF

Then

- Incorporated in 1999
- Started with 21 members
- 2 projects
- All servers and services donated



Now

- ~650 members
- ~300 projects (TLPs, podlings)
- ~5400 committers
- ~2200 PMCs
- ~8000 signed ICLAs
- Large and growing infrastructure

ASF mission

What ?

- Provide open source software to public free of charge
- Let the coders code - foundation deals with the rest



How ?

- Infrastructure for open source development
- Legal entity for donation purposes
- Shelter from law suits
- Protection of the Apache brand
- Project governance and guidance



ASF org chart

Development/Project

Project
People

PMC Members

Committers

Contributors

Users

Foundation/Administrative

Members

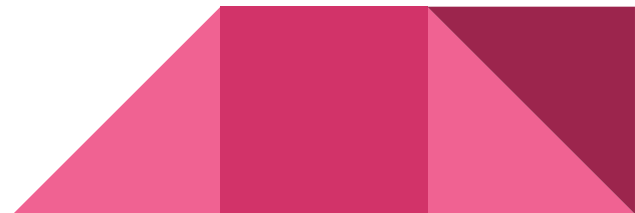
VP/Officers

delegates

Board

elect

report quarterly



Structure of the ASF - Project Management Community

- Volunteer driven organization
- Projects are managed by Project Management Committees (PMCs)
 - Vote in new PMC members and committers
 - Vote in new releases
 - Establish rules, report to board
 - Oversee legal aspect, and community health
- Each PMC has a Chair (appointed by the board) - eyes and ears of the board (oversight only)

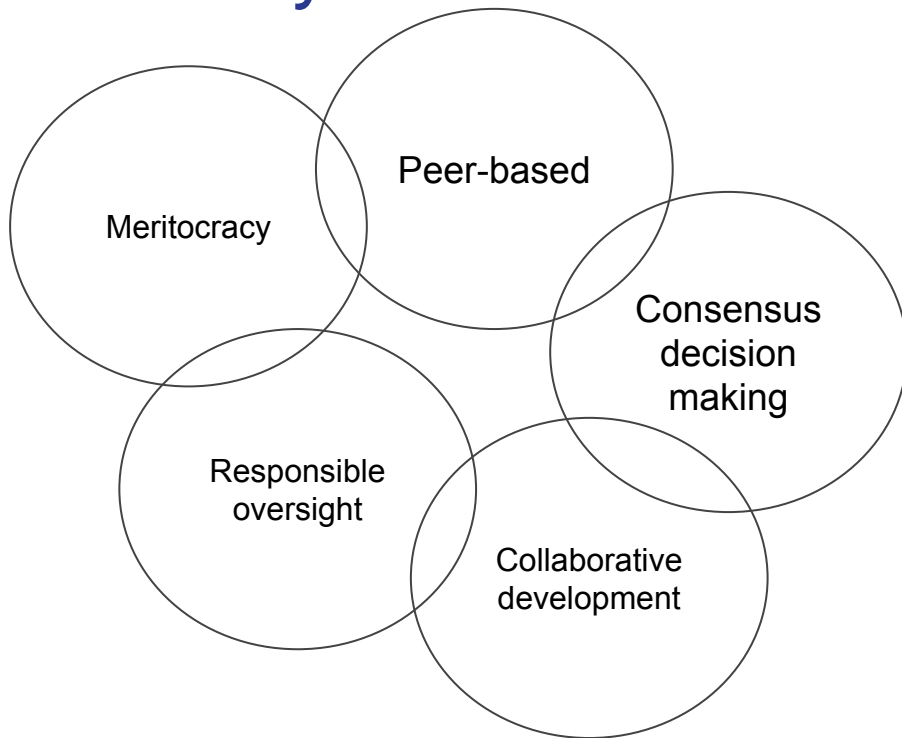


Structure of the ASF - Foundation Members and Board

- Again only volunteer effort
- Member-based corporation - individuals only
- Members nominate and elect new members
- Members elect a board - 9 seats
 - Manage and oversee the corporation, its assets and resources
- Annual meeting via IRC
- Board elects 10 officers
 - Oversee the day-by-day affairs



The Apache Way



Meritocracy

- Govern by **merit**
- Merit is based on what you do
- Merit never expires
- Those with merit, get more responsibility
- The best ideas win
- Any constructive contribution earns merit
- **Not just code** (bug reports, documentation, evangelism, ...)



Peer-based

- Developers represent themselves - **individuals**
- **Mutual trust and respect**
- All votes hold the same weight
- **Community over code**
 - Healthy community creates healthy code
 - Poisonous communities don't



Why community > code

- Since we are all volunteers, people's time and interests change
- A healthy community is “warm & inviting” and encourages a continued influx of developers
- Poisonous people/communities turn people off, and the project will die
- End result - **better code, long-term code**



Consensus and vote

- Key is the idea of voting: +1, 0, -1
- Main intent: gauge developer acceptance
- -1 must be justifiable



Commit processes

Commit Then Review (CTR)

- Commit on the code and review after
- Good for dev branches
- Depends on people doing the review
- Allows very fast development

Review Then Commit (RTC)

- Based on patches/pull requests
- Review on the pull request
- Merge the pull request
- Needs enough “reviewers”
- Lower pace, higher quality and governance
- *Probably the recommended and preferred process*

Lazy Consensus

- Variant of RTC
- Introduce deadline in the review
- Lower quality, higher flexibility



Collaborative development

- Code is developed by the community
- Voting ensures at least 3 active committers
- Decisions done on mailing list: ***“If it didn’t happen on the mailing list, it didn’t happen”***
- Shared vision
- Real-time updates
- Real-time help requests
- Self-service archives



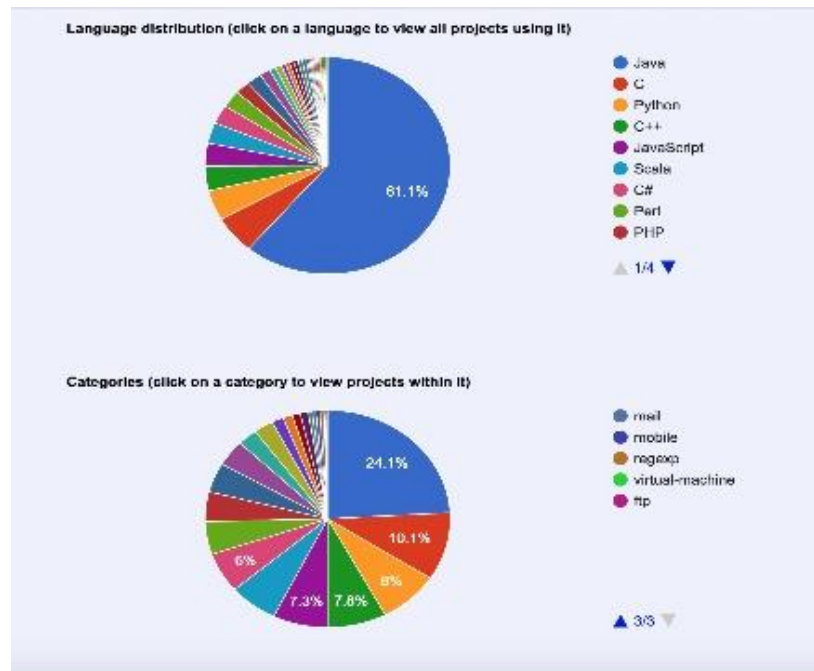
Responsible oversight

- Ensure clean IP in all Apache releases
- Ensure license compliance
- Quality code
- Quality community



Apache projects

- Responsible for their own code, community and direction, following the Apache rules
- Community over Code !
- Diversity: Java, C, C++, Perl, Scala, ...
- Leading technology (webservers, tools, big data, container, ...)
- Projects can be in the same area if they have their own community (ant/maven, axis/cxf, pig/hive, ...)



Apache projects



Apache projects



Apache incubator

- Entry point for all new projects
- Indoctrinates the Apache Way to the podling
- Ensures and tracks IP
- PPMC helped by mentors (foundation members)



Who pays ?

- Apache doesn't pay for development ! Voluntary contribution only !
- Many (not all) developers are paid by third-party to work on a project (form of sponsoring)
- Foundation bears indirect support costs
 - Infrastructure
 - Advertising and conference
 - ...



Who are we ?

“We are more than a group of projects sharing a server, we are a community of developers and users.”



How to start ?

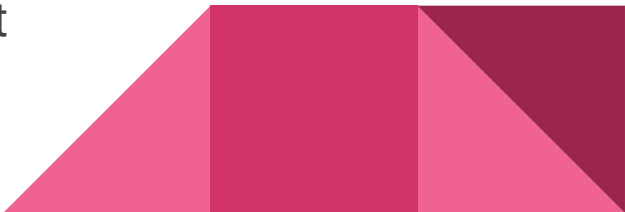


Why to contribute ?

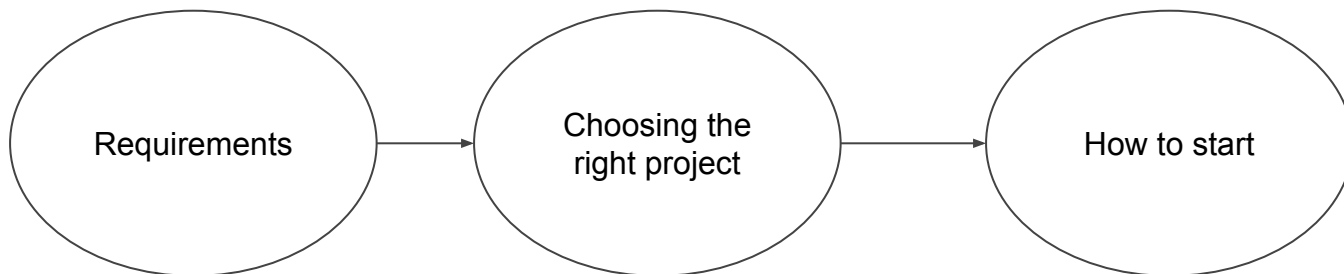
- **Better and easier recognition of work**
- Publicly verifiable resume
- **Work with great guys in great teams**
- No manager, no boss
- Work on what you like, when you like, where you want
- **Discuss technical designs and by part of the innovation**
- Build software used by millions around the world
- Networking opportunities



How to contribute ?

- **ASF wants voluntary contributions**
 - Documentations, tutorials, examples
 - Helping others with queries and questions
 - Issue / bug tracker triage
 - Testing new fixes, helping reproduce problems
 - Bug fixes and new features
 - Writing add-ons and extensions
 - Mentoring, volunteering for the foundation
 - Many different ways to get involved, all are important
- 

Start contributing



- English
- Programming language
- Debugging
- Passion
- Perseverance
- Time

- Filter by your interest area
- Filter by language
- Something you use
- Something you want to learn
- Using the project is very important

- Choose your project
- Subscribe to mailing lists
- Checkout code
- Play with binaries
- Find issues and features requests
- Ask
- Sign ICLA

Become a committer/PMC

Invitation of commit access

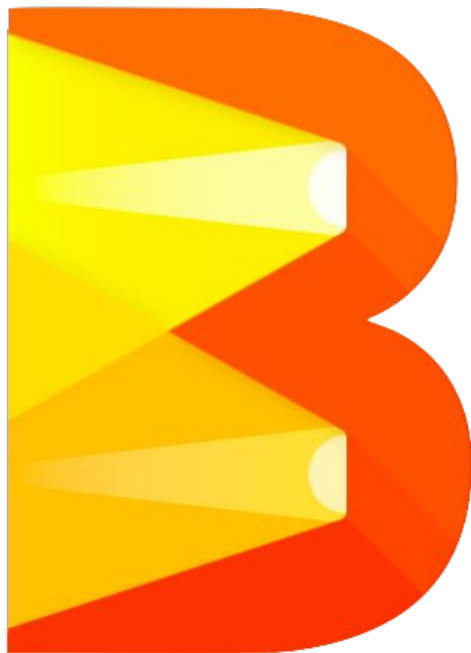
- Current PMC member nominates individual
- Discussion on the private@ mailing list
 - Key question ? Do I trust this individual ?
- PMCs are free to define the bar level

Beyond a committer

- Once you are committer, you can then become:
 - PMC member
 - Foundation member
 - Director/VP
 - Even President !
- You can
 - Nominate other foundation (or PMC) members
 - Can serve as mentor for incubating projects
 - Vote for board



Example: Apache Beam



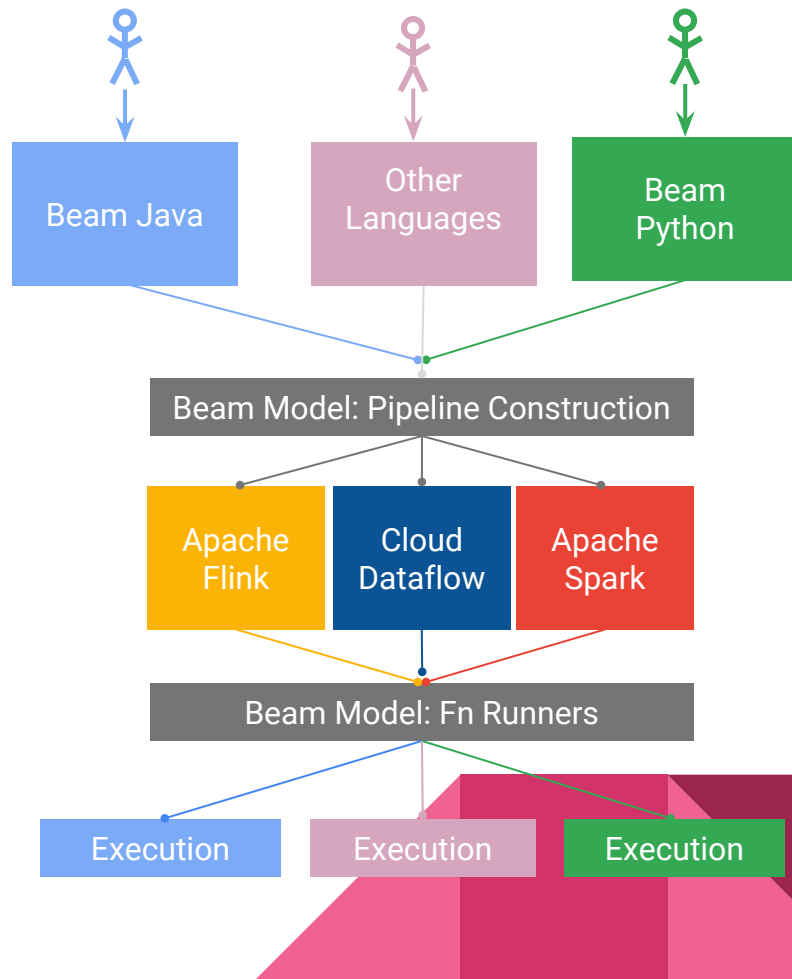
What's Apache Beam

- **Unified programming model** for big data (batch and streaming)
- SDKs (**Java, Python, ...**) & DSLs (**Scala, SQL, ...**)
- **Runners** for existing distributed processing engines (**Spark, Flink, Apex, Dataflow, ...**)
- **IOs** to connect with data sources (**Google GS, BigTable, BigQuery, Cassandra, Elasticsearch, ...**)



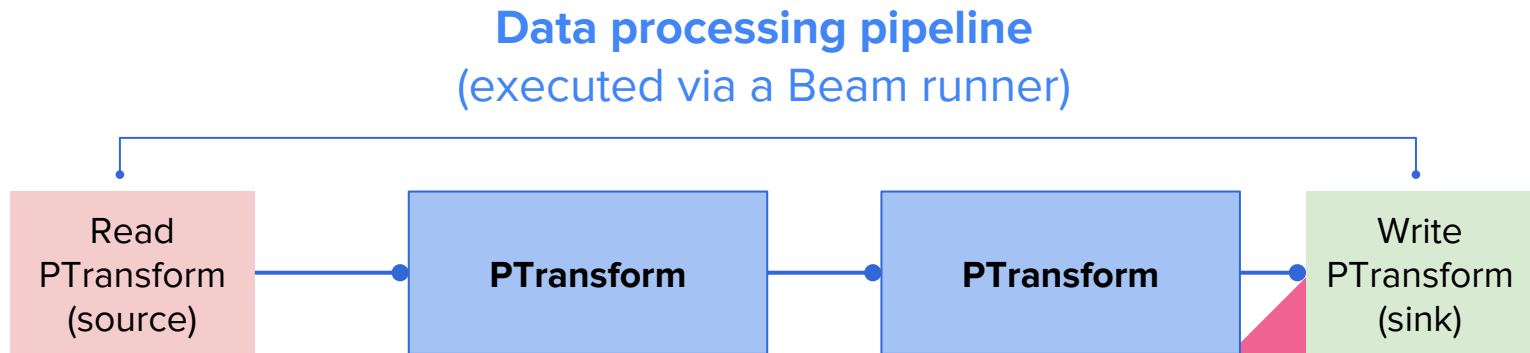
Apache Beam vision

1. **End users:** who want to write pipelines in a language that's familiar.
2. **SDK/DSL writers:** who want to make Beam concepts available in new languages.
3. **Runner writers:** who have a distributed processing environment and want to support Beam pipelines



Apache Beam model concepts

1. **Pipeline** - data processing job as a directed graph of transformation
2. **PCollection** - the data inside a pipeline
3. **PTransform** - a transformation step in the pipeline (**IO** or pure **transform**)



Apache Beam runners

- “**Translate**” pipeline to a **target runtime**
- **Same code** can run on **different engines without re-compilation**
- Runners can be **on-premise** (local Spark cluster, Yarn, Mesos, ...) or in a **cloud** (using Google Cloud Dataproc, Amazon EMR, ...)

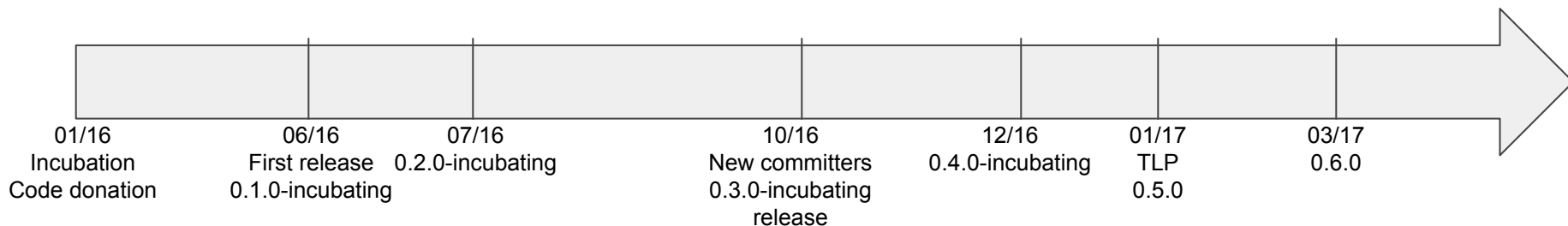


Why Apache Beam ?

- **Portable** - same code with different runners (agnostic) and backends on premise, in the cloud, or local
- **Unified** - same model for batch and streaming processing
- **Advanced features** - event windowing, triggering, watermark, checkpoint, lateness, ...
- **Extensible** - create your own IOs, SDKs, DSLs, runners



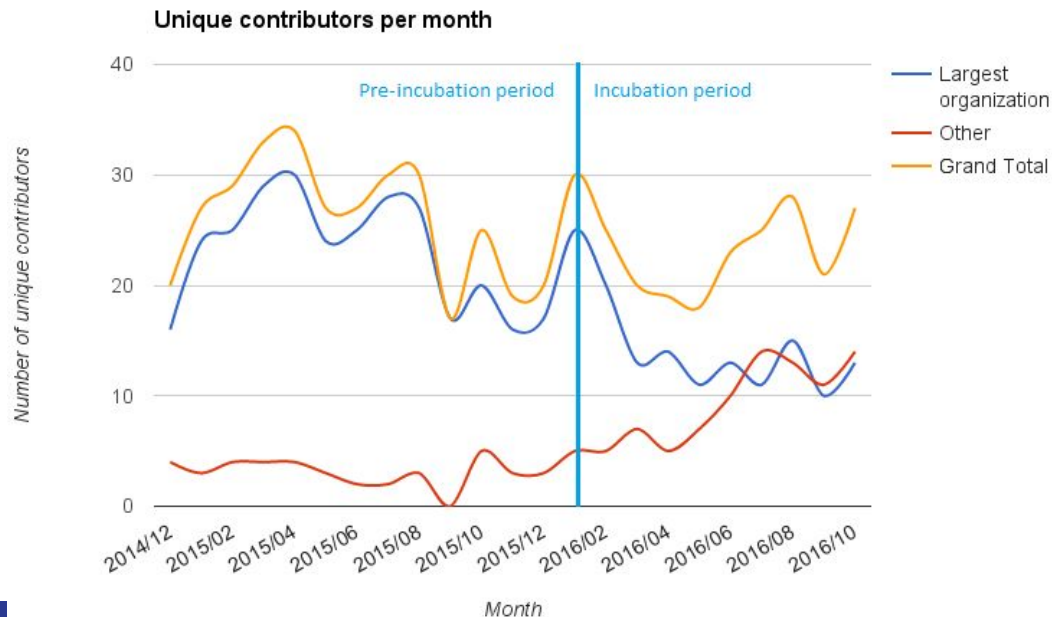
Beam: from incubation to TLP



- Incubation started in January '16
- TLP in January '17
- 1500 Pull Request during incubation
- 3 incubating releases
- 120+ individual contributors
- 3 new committers from new companies

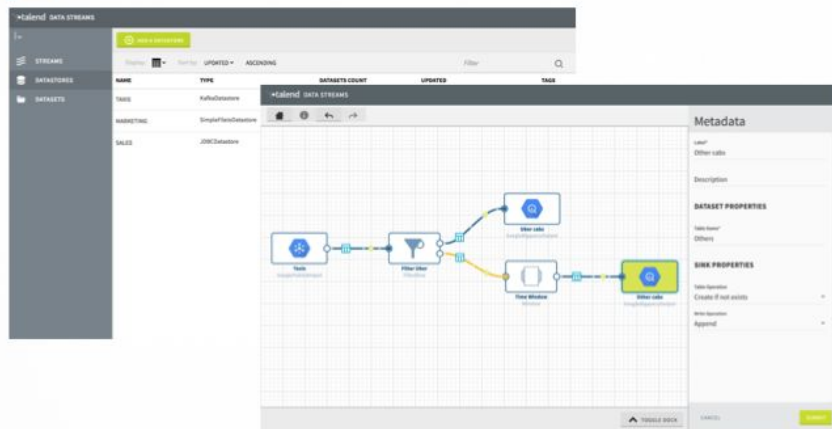
Apache Beam, ASF values

- Larger community
- More features provided by more people:
 - New IOs contributed by Talend
 - Spark runner improvements by PayPal & Talend
 - Apex runner
 - ...
- Extended use cases and users
 - AliBaba, eBay, ...

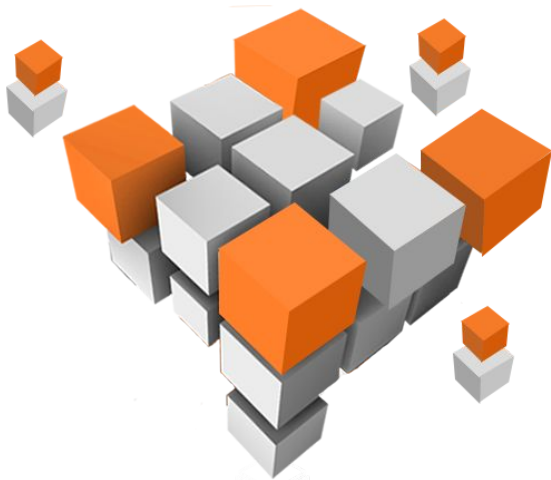


Talend products on top of Apache Beam

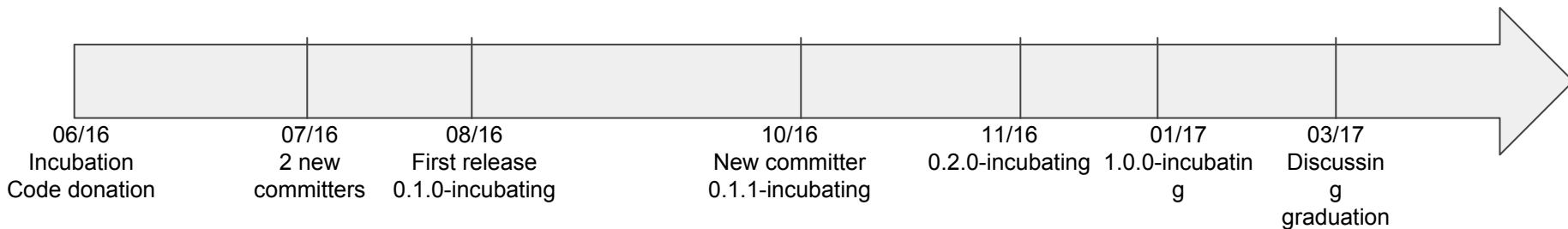
- Apache license allows this !
- DataPrep using Beam pipelines
- DataStream, an UI tool to generate Beam pipelines



Another example: Apache CarbonData !



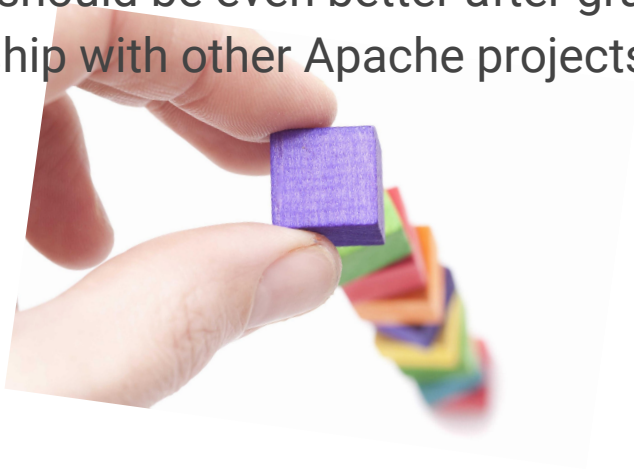
Another example: Apache CarbonData !



- Incubation started in June '16
- 650 Pull Requests
- Clearly increase of users & contributors on the mailing lists
- Heading to graduation !

Apache for Apache CarbonData

- More visible: well known and more popular project
- Growing community, with new ideas and contributions
 - Huawei contributors is now around 50%
 - Contributors from other companies is also 50%
- Attract new users (should be even better after graduation)
- Simplified relationship with other Apache projects



Thanks !

Special thanks to Liang & the CarbonData team !

You are awesome !



Q&A

