ebay inc™

QCon

Apache Kylin Open Source Journey

韩卿 | Luke Han

Co-Creator & PMC Member

lukehan@apache.org

Agenda

- About Apache Kylin
- Kylin Open Source Journey
- Apache Incubating
- Build Community and Ecosystem
- The Good, The Bad and The Ugly
- Q&A

About Apache Kylin (麒麟)

http://kylin.io

Extreme OLAP Engine for Big Data

Kylin is an open source Distributed Analytics Engine that provides SQL interface and multi-dimensional analysis (OLAP) on Hadoop supporting extremely large datasets

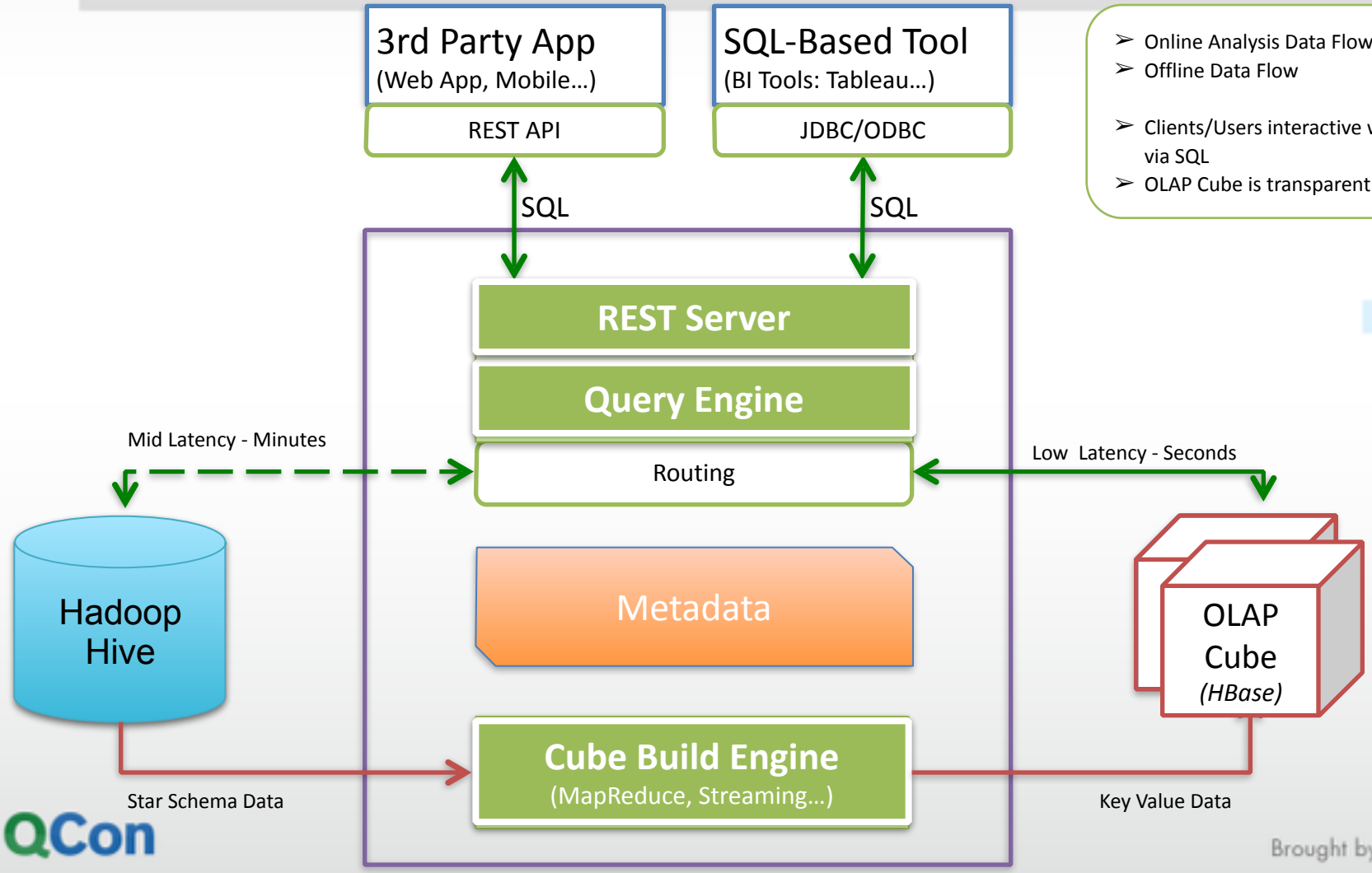
- First Apache Project open sourced by eBay Inc.
- First Apache Project fully contributed from eBay CCOE

- Open Sourced on Oct 1st, 2014
- Be accepted as Apache Incubator Project on Nov 25th, 2014
  - Apache Kylin is an effort undergoing incubation at The Apache Software Foundation (ASF), sponsored by Incubator.

Technical Challenges

- Huge volume data
  - Table scan
- Big table joins
  - Data shuffling
- Analysis on different granularity
  - Runtime aggregation expensive
- Map Reduce job
  - Batch processing

Apache Kylin Architecture

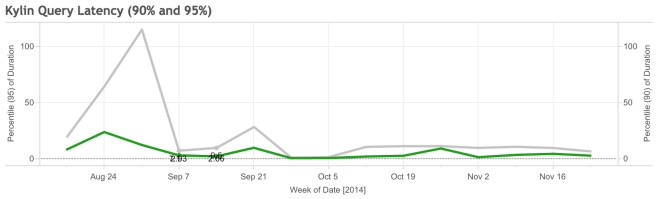


Features

- Extremely Fast OLAP Engine **at scale**
- **ANSI SQL** Interface on Hadoop
- Seamless Integration with BI Tools, like **Tableau**
- **Interactive** Query Capability
- **MOLAP** Cube
- Compression and Encoding Support
- **Incremental** Build of Cubes
- **Approximate** Query Capability for Distinct Count (HyperLogLog)
- Leverage HBase **Coprocessor** for query latency
- Job Management and Monitoring
- User friendly **Web GUI** for manage, build, monitor and query cubes
- Security capability to set ACL at Cube/Project Level
- Support LDAP Integration

- **Streaming Support Coming soon!**

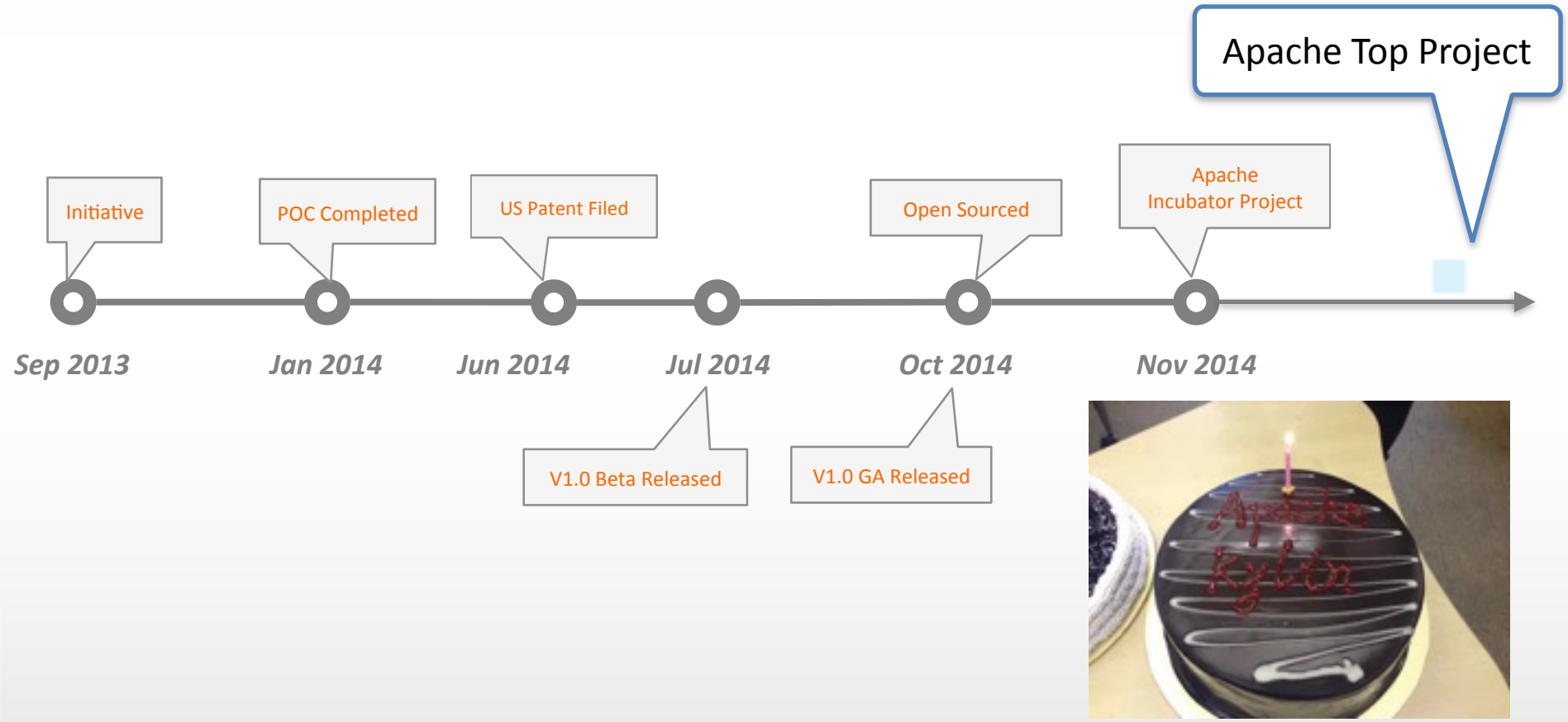
90%tile queries <5s



Agenda

- About Apache Kylin
- Kylin Open Source Journey
- Apache Incubating
- Build Community and Ecosystem
- The Good, The Bad and The Ugly
- Q&A

Kylin Open Source Journey



Ready for Open Source

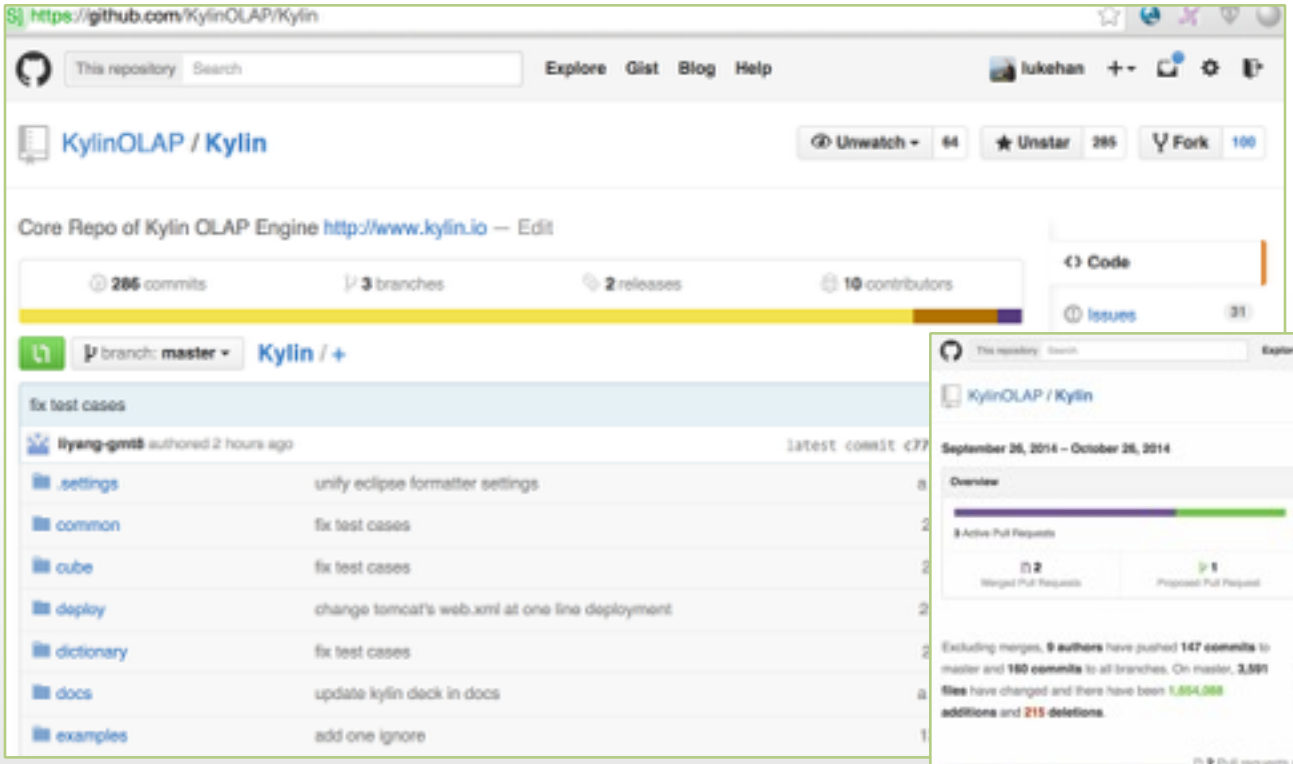
- Open Source from Day One
- Internal vs External
- Intellectual Property
- Legal
- Domain
- License
  - Apache/MIT/BSD/GPL…
- Team

Patent

- Why?
- How?
- Patent vs Open Source

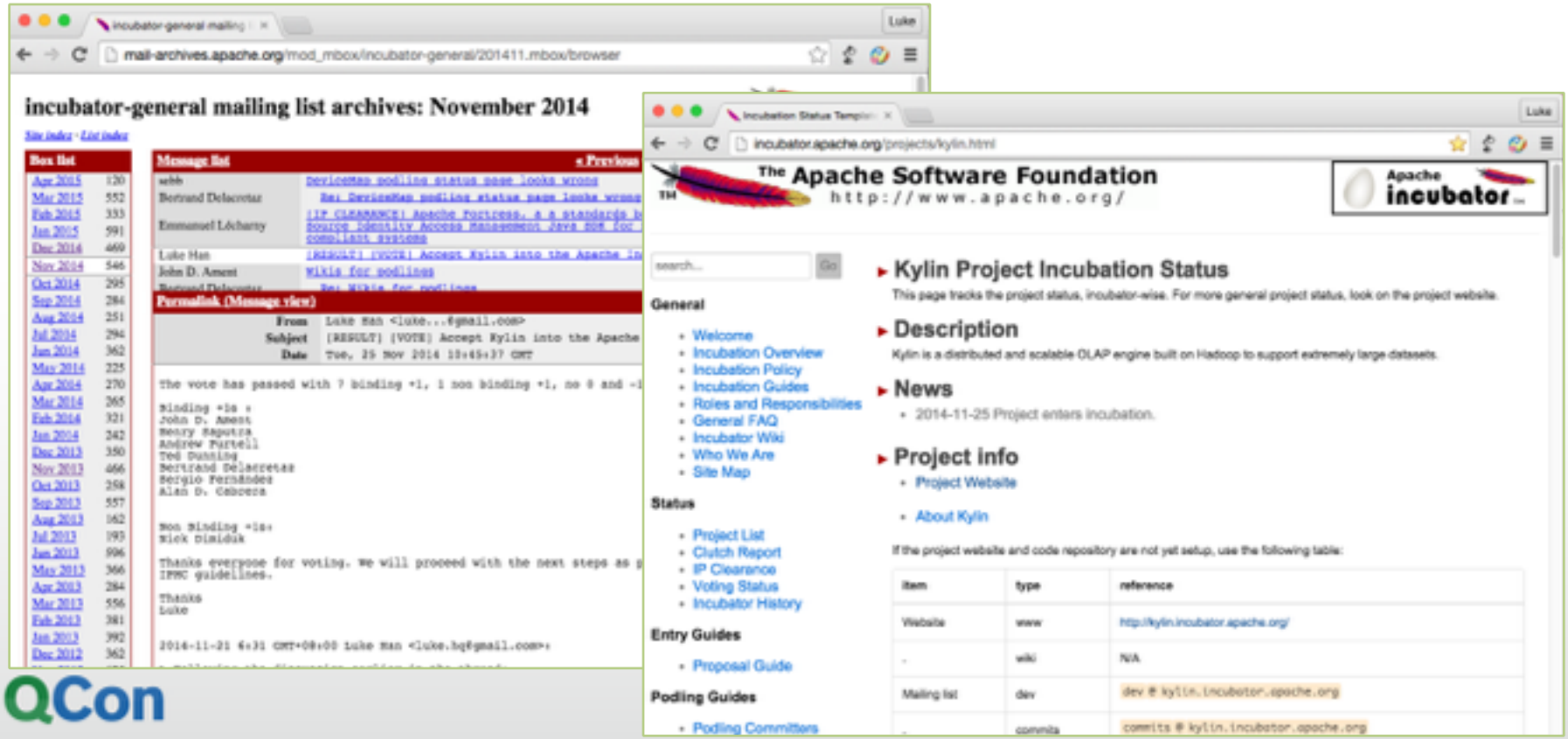
Phase I: Open Source on Github

- Code pushed to github.com on Oct 1st, 2014



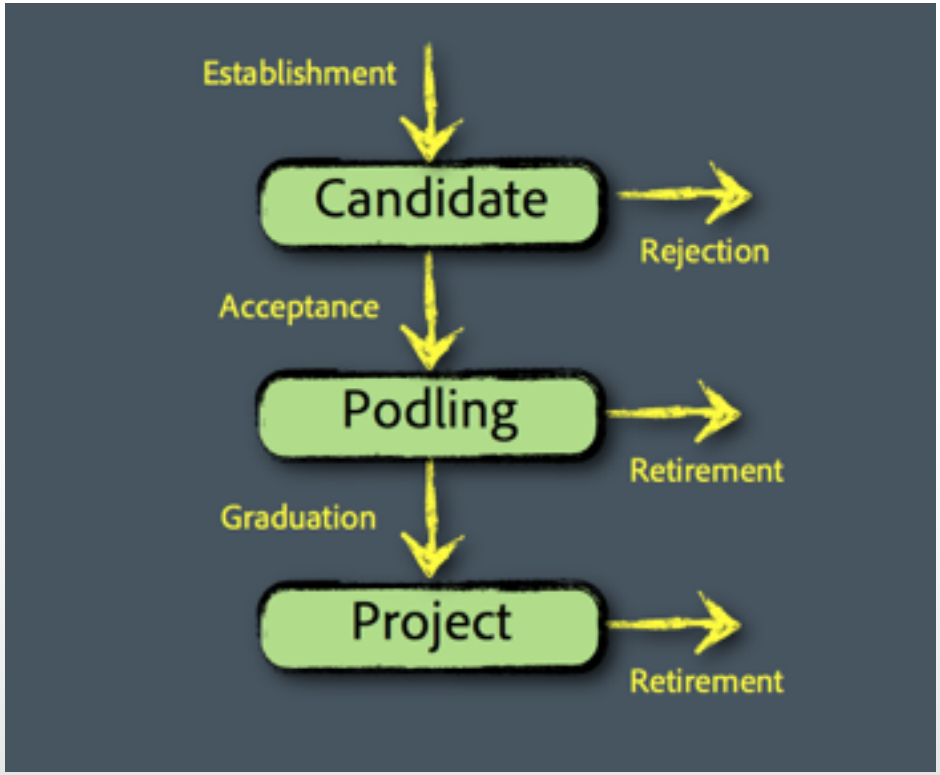
Phase II: Apache Incubator

- Be accepted as Apache Incubator Project on Nov 25th, 2014

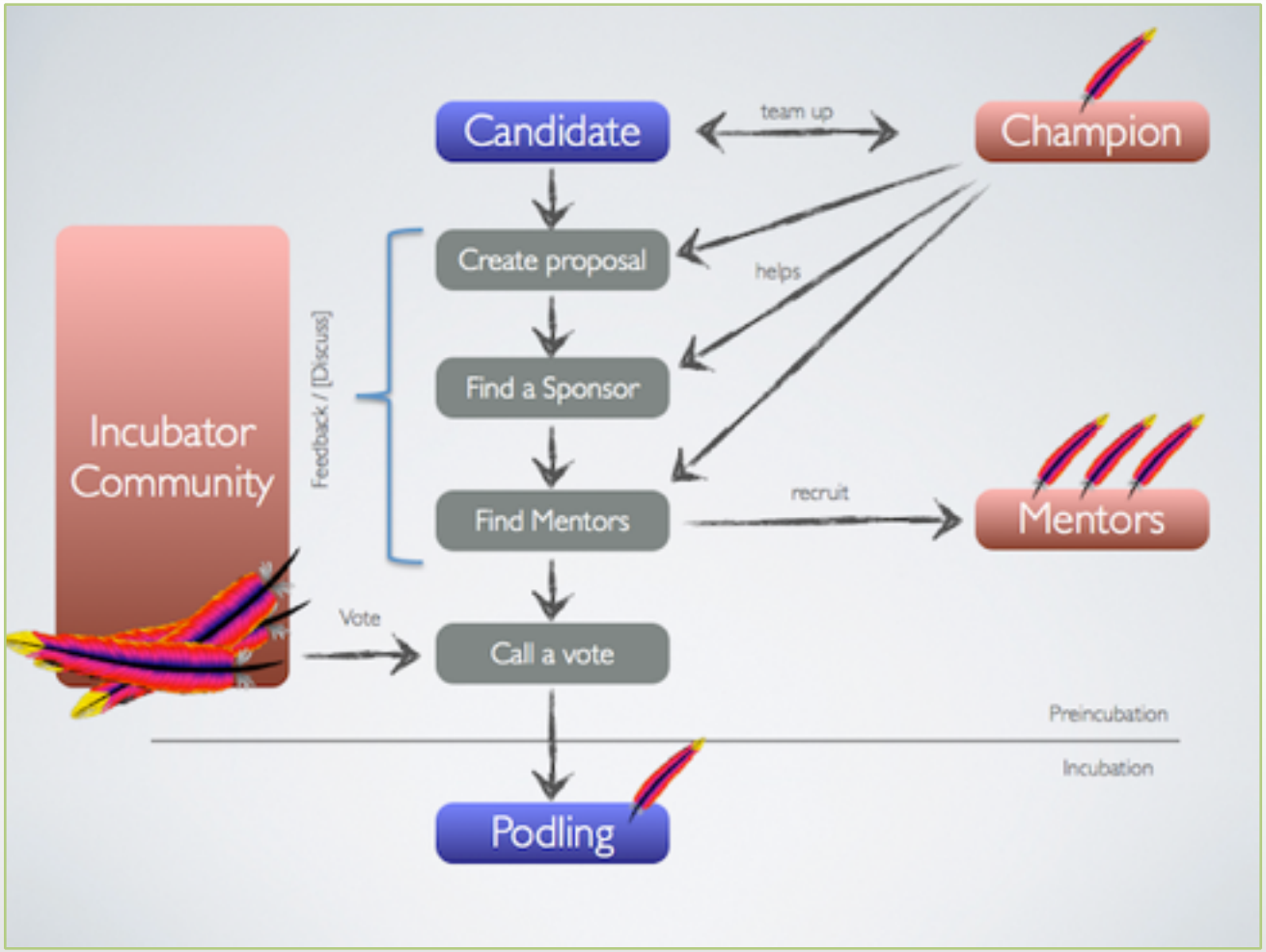


Why & How Apache?

- Hadoop Ecosystem Home
- Branding
- Community
- The Apache Way

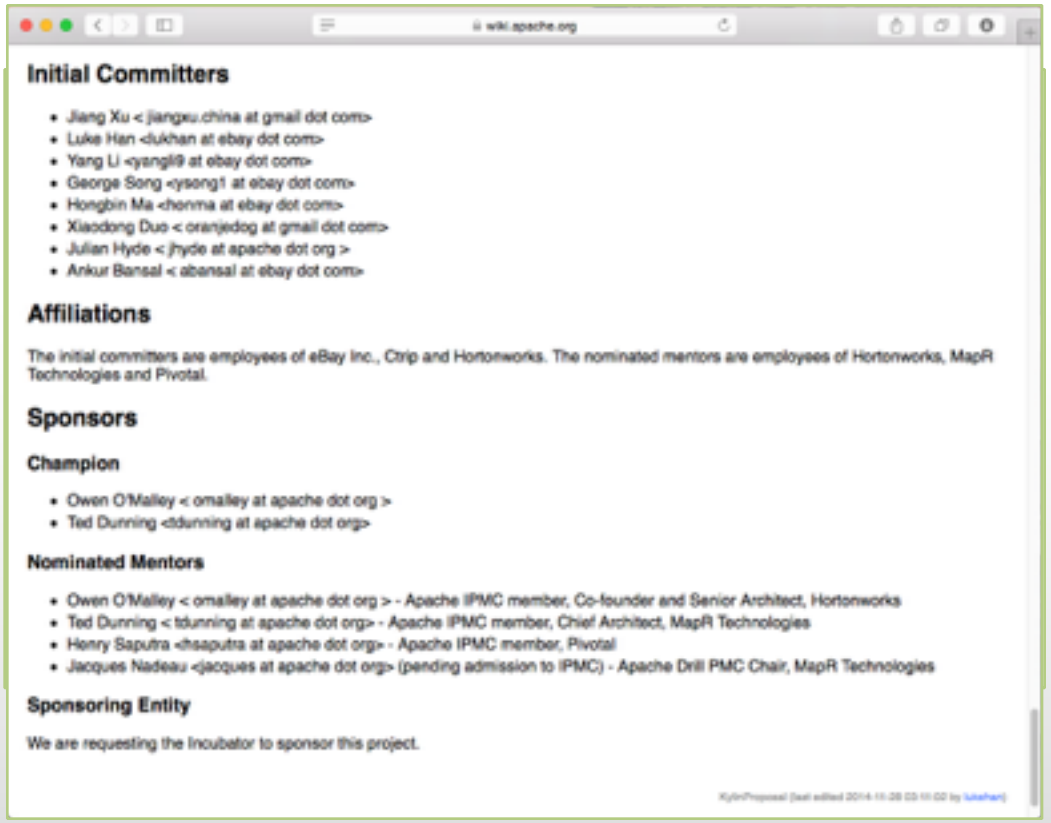


Incubation Progress



Incubator Project Proposal

- IPMC & PPMC
- Mentors and Champion
- Committers



wiki.apache.org

Initial Committers

- Jiang Xu < jiangxu.china at gmail dot com>
- Luke Han <lukhan at ebay dot com>
- Yang Li <yangli9 at ebay dot com>
- George Song <ysong1 at ebay dot com>
- Hongbin Ma <honma at ebay dot com>
- Xiaodong Duo < oranjedog at gmail dot com>
- Julian Hyde < jhyde at apache dot org >
- Ankur Bansal < abansal at ebay dot com>

Affiliations

The initial committers are employees of eBay Inc., Ctrip and Hortonworks. The nominated mentors are employees of Hortonworks, MapR Technologies and Pivotal.

Sponsors

Champion

- Owen O'Malley < omalley at apache dot org >
- Ted Dunning <tdunning at apache dot org>

Nominated Mentors

- Owen O'Malley < omalley at apache dot org > - Apache IPMC member, Co-founder and Senior Architect, Hortonworks
- Ted Dunning < tdunning at apache dot org> - Apache IPMC member, Chief Architect, MapR Technologies
- Henry Saputra <hsaputra at apache dot org> - Apache IPMC member, Pivotal
- Jacques Nadeau <jacques at apache dot org> (pending admission to IPMC) - Apache Drill PMC Chair, MapR Technologies

Sponsoring Entity

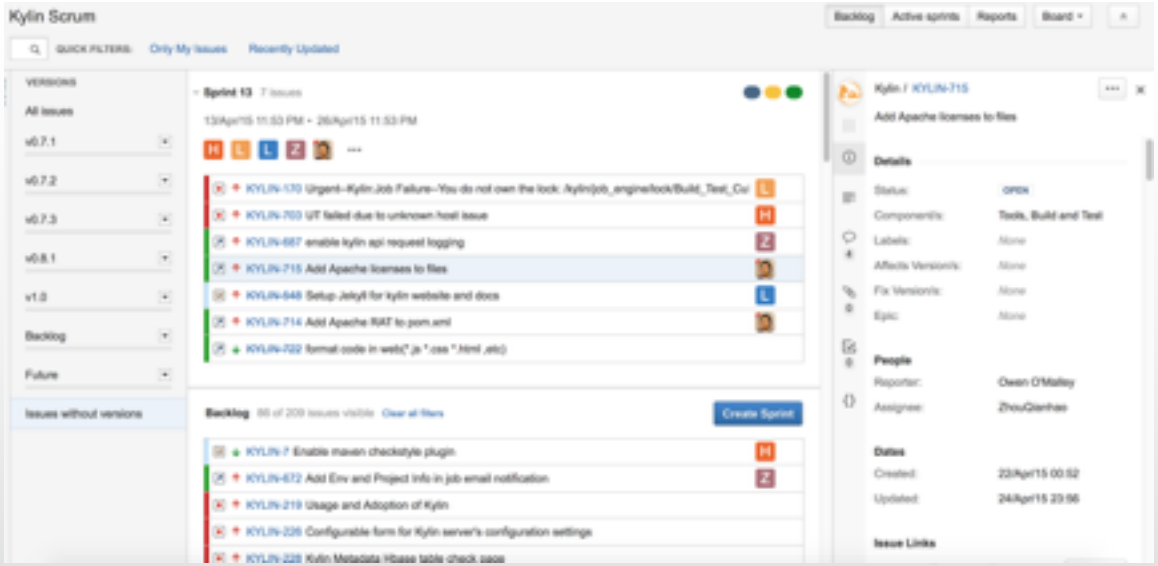
We are requesting the Incubator to sponsor this project.

Agenda

- About Apache Kylin
- Kylin Open Source Journey
- Apache Incubating
- Build Community and Ecosystem
- The Good, The Bad and The Ugly
- Q&A

Infrastructure Setup

- Mailing List
  - Private@
  - Dev@
- Source Code Repo
  - git & svn
  - Migration
- Website
- JIRA
- Wiki



IP Clearance & Release

- Kylin for brand name?
- Apache License
  - GPL Dependency?
- Apache Release
  - README, LICENSE, NOTICS, DECLIARMER
  - Source Headers
  - Licensing of dependencies
  - Binaries

Team onboard Apache Way

- Community then Code
- Mailing list discussions
- Vote
- Code Quality and Style
- JIRA for each issue, feature
- Merge Pull Request
- Recruiting contributor/committer

How to contribute?

- Join mailing list:
  - dev@kylin.incubator.apache.org
- Create JIRA or Leave Comments
- Pull Request/Patch to Apache Github Mirror

Graduate to Top Project

- Diversity
- Complete (and sign off) tasks documented in the status file
- Ensure suitability for project name and product name
- Demonstrate ability to create Apache releases
- Demonstrate community readiness
- Ensure that mentors and the IPMC have no remaining issues

Ready to Apache?

The Apache Software Foundation

Community-led development since 1999.

Agenda

- About Apache Kylin
- Kylin Open Source Journey
- Apache Incubating
- Build Community and Ecosystem
- The Good, The Bad and The Ugly
- Q&A

Build Community and Ecosystem

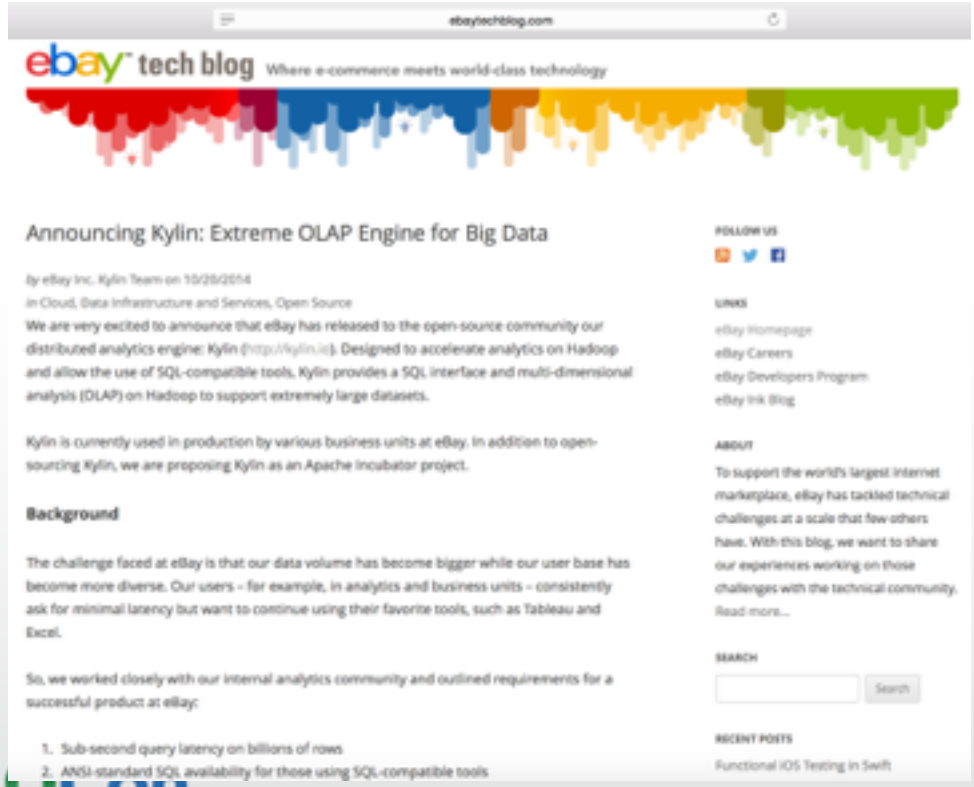
- What's community?
- How to grow community?
- Community than Code!

Marketing - Website

- http://kylin.io
  - Hosted on github.io (Github Pages)
  - Hosted on Apache Infra Server
    - http://kylin.incubator.apache.org

Marketing - Blog

- Publish via eBay Tech Blog to gain focus from industry
- http://www.ebaytechblog.com/2014/10/20/announcing-kylin-extreme-olap-engine-for-big-data

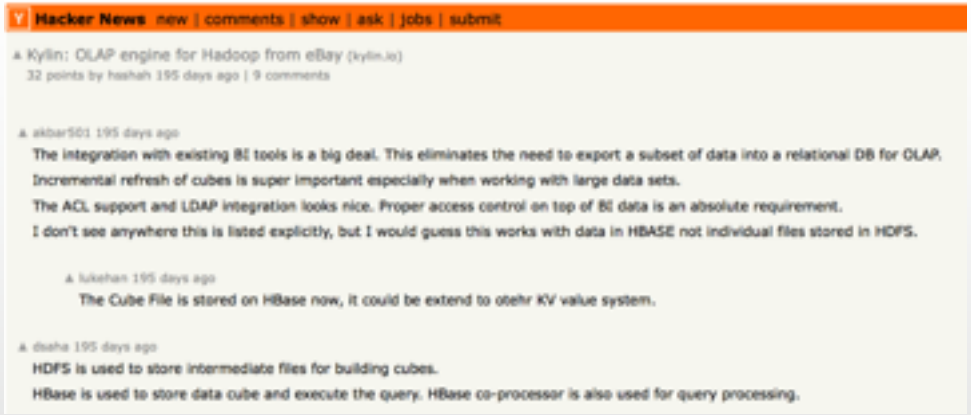


"Like arch-rival Amazon.com, the soon-to-split eBay Inc. is something of an oddity in that it hasn't historically been a big contributor to the open-source community. But the e-commerce pioneer hopes to change that with the release of the source-code for a homegrown online analytics processing (OLAP) engine that promises to speed up Hadoop while also making it more accessible to everyday enterprise users."

-- siliconangle.com

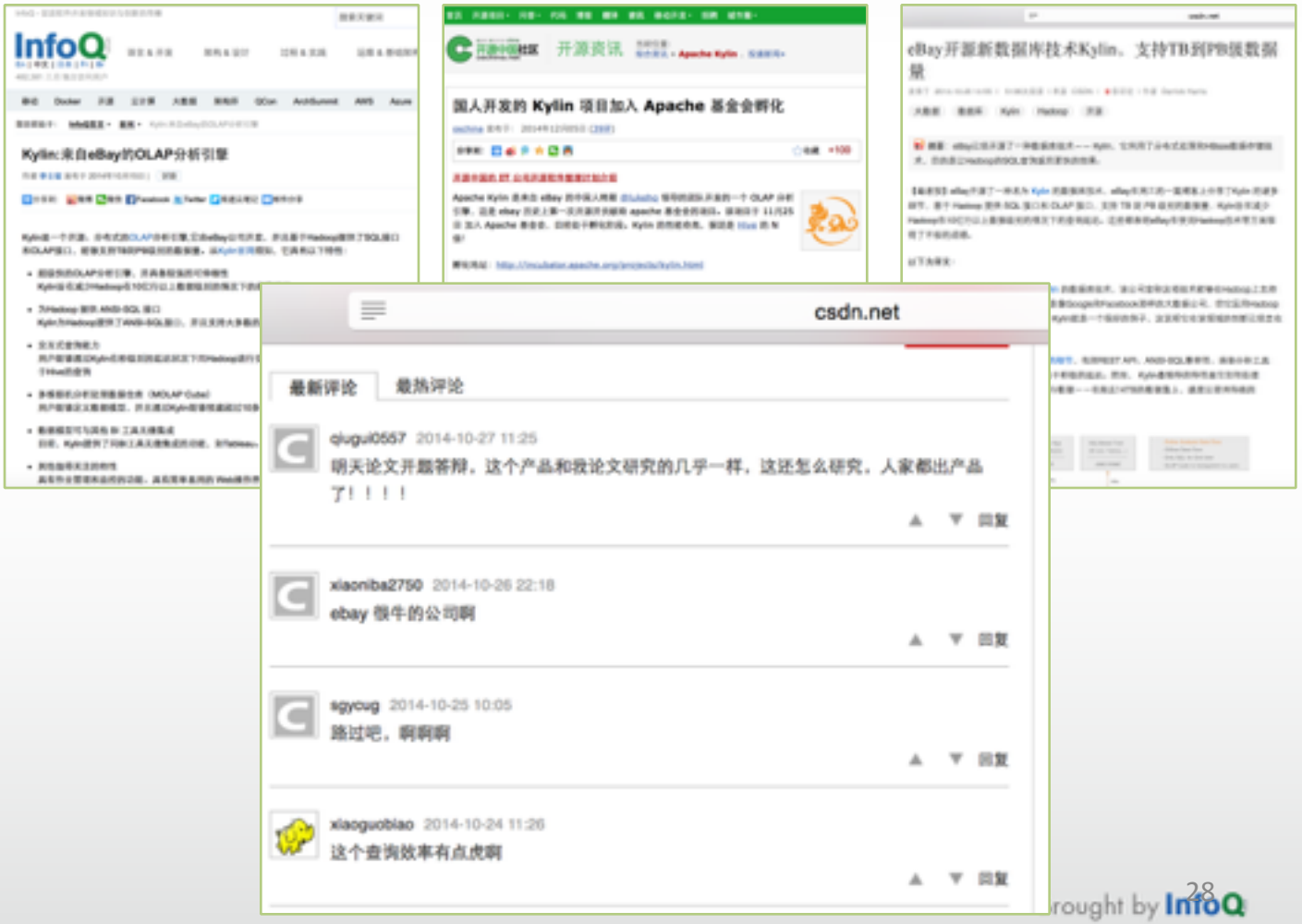
Marketing – Social Media

- Github
  - KylinOLAP
- Twitter
  - @ApacheKylin
- HackNews
- Facebook
  - Page: kylin.io
- LinkedIn
  - Group: Kylin
- WeChat(微信)
  - ApacheKylin
- …



Marketing - Media

- InfoQ
- CSDN
- OSChina
- …



Build Community – Mailing List

Build Community – Meetup

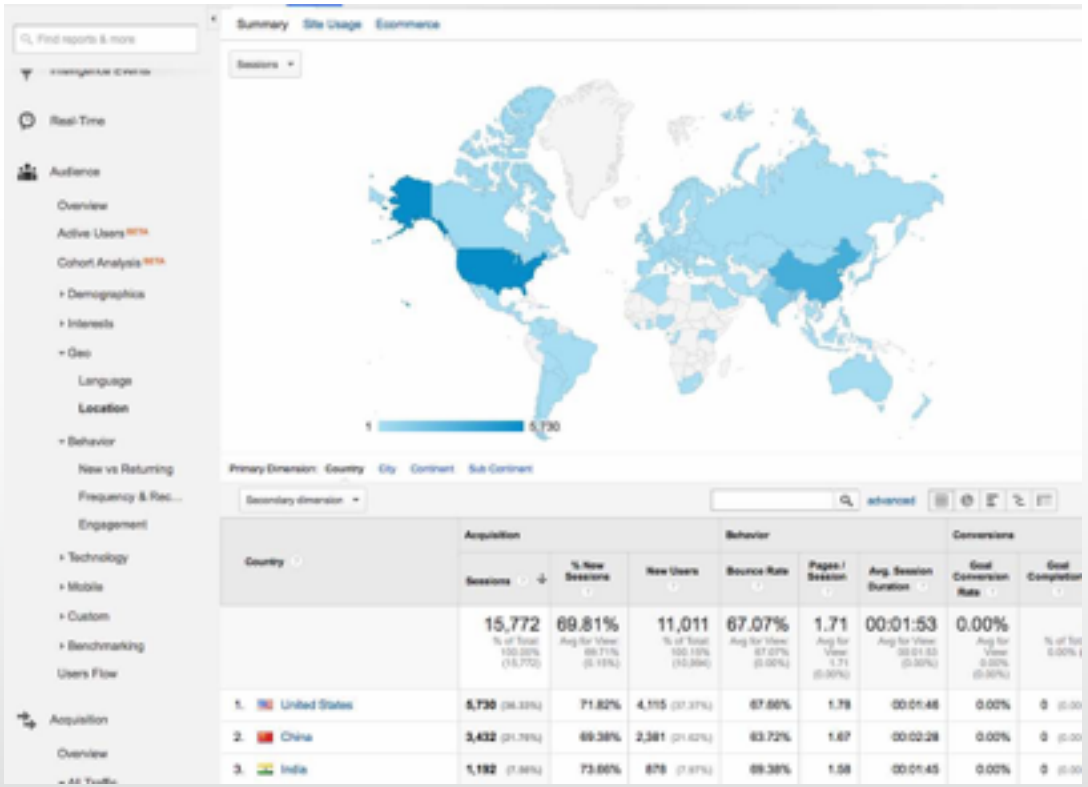
- Hive Meetup Bay Area, Dec 2014
- Apache Kylin Meetup Bay Area, Dec 2014
- Apache Kylin Tech Talk @AWS Seattle, Dec 2014
- Apache Kylin Meetup Beijing, Dec 2014
- Spark Meetup Bay Area, March 2015
- Kylin Meetup in China, coming soon
- …

Build Community – Conference

- Big Data Summit Shanghai, Oct 2014
- Big Data Technology Conference Beijing, Dec 2014
- Database Technology Conference Beijing, April 2015
- Hadoop Summit Europe, April 2015
- **QCon Beijing, April 2015**
- Strata+Hadoop World London, May 2015
- HBaseCon San Francisco, May 2015
- Hadoop Summit San Jose, June 2015
- …

Know your community

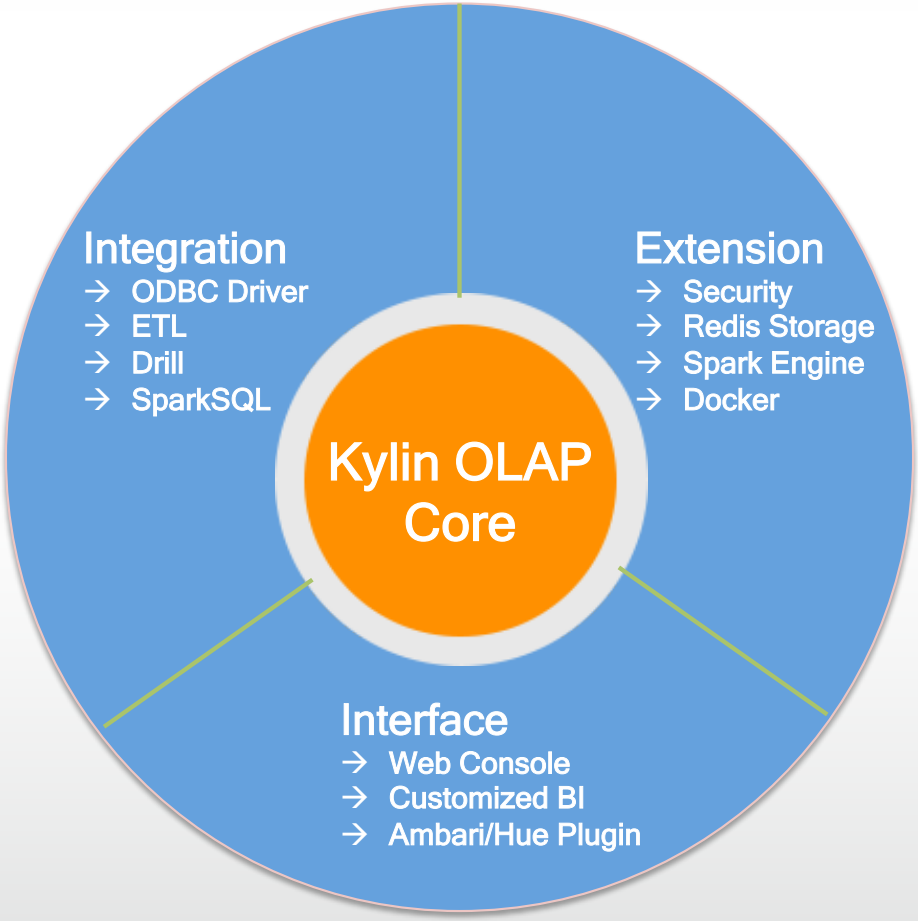
- Google Analytics
- Github Statistics
- Mailing List
- WeChat
- …



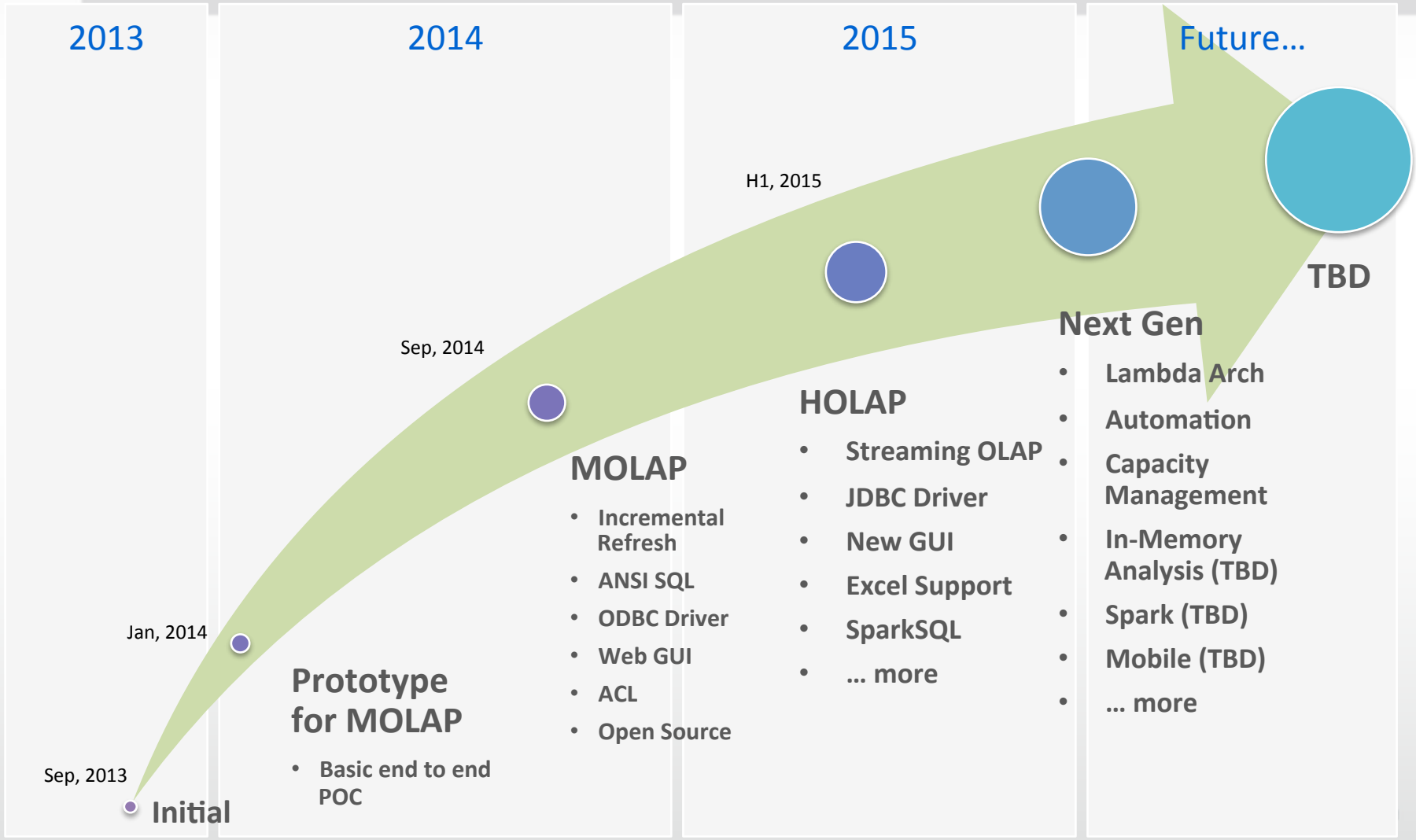
QCon

Apache Kylin Ecosystem

- Kylin Core
  - Fundamental framework of Kylin OLAP Engine
- Extension
  - Plugins to support for additional functions and features
- Integration
  - Lifecycle Management Support to integrate with other applications like BI tools
- Interface
  - Allows for third party users to build more features via user-interface atop Kylin core



Apache Kylin Evolution Roadmap



Excellence of Engineering

Team Philosophy

Recruit best people

Done is better than perfect

Do academic research

Explain design in simple words

Everyone does dirty work

You write first version, I write second one

Debate, Decision & Delivery

Agenda

- About Apache Kylin
- Kylin Open Source Journey
- Apache Incubating
- Build Community and Ecosystem
- The Good, The Bad and The Ugly
- Q&A

The Good

- 知名度
- 个人成长
- 团队文化
- 项目质量
- 成就感
- 和牛人做邻居

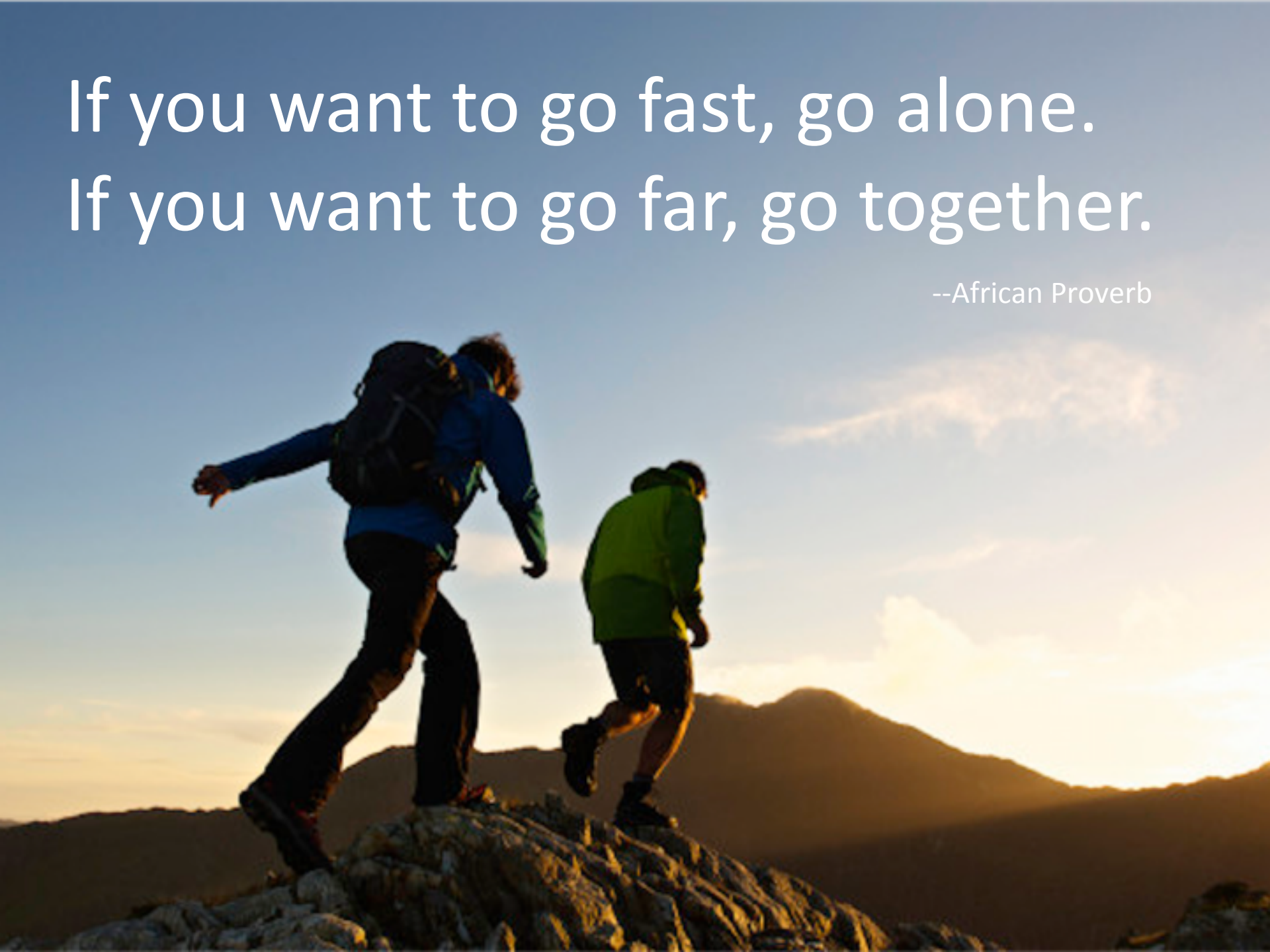
全世界都在注视着你和你的代码！

The Bad

- 开发效率降低
- 内部项目进度vs外部支持和问题
- 业余时间
- Roadmap and Features from external

The Ugly

- 开源不等于免费
- 请尊重开源作者
- Ask question with right way
- 山寨


If you want to go fast, go alone.
If you want to go far, go together.
--African Proverb

Apache Kylin

- Kylin Site:
  - http://kylin.incubator.apache.org
  - http://kylin.io
- Twitter:
  - @ApacheKylin
- WeChat(微信)
  - ApacheKylin

www.infoq.com/cn

InfoQueue

@InfoQ

infoqchina

软件
正在改变世界！