



ebay inc<sup>tm</sup>

QCon

# Apache Kylin Open Source Journey

韩卿 | Luke Han

Co-Creator & PMC Member

[lukehan@apache.org](mailto: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*

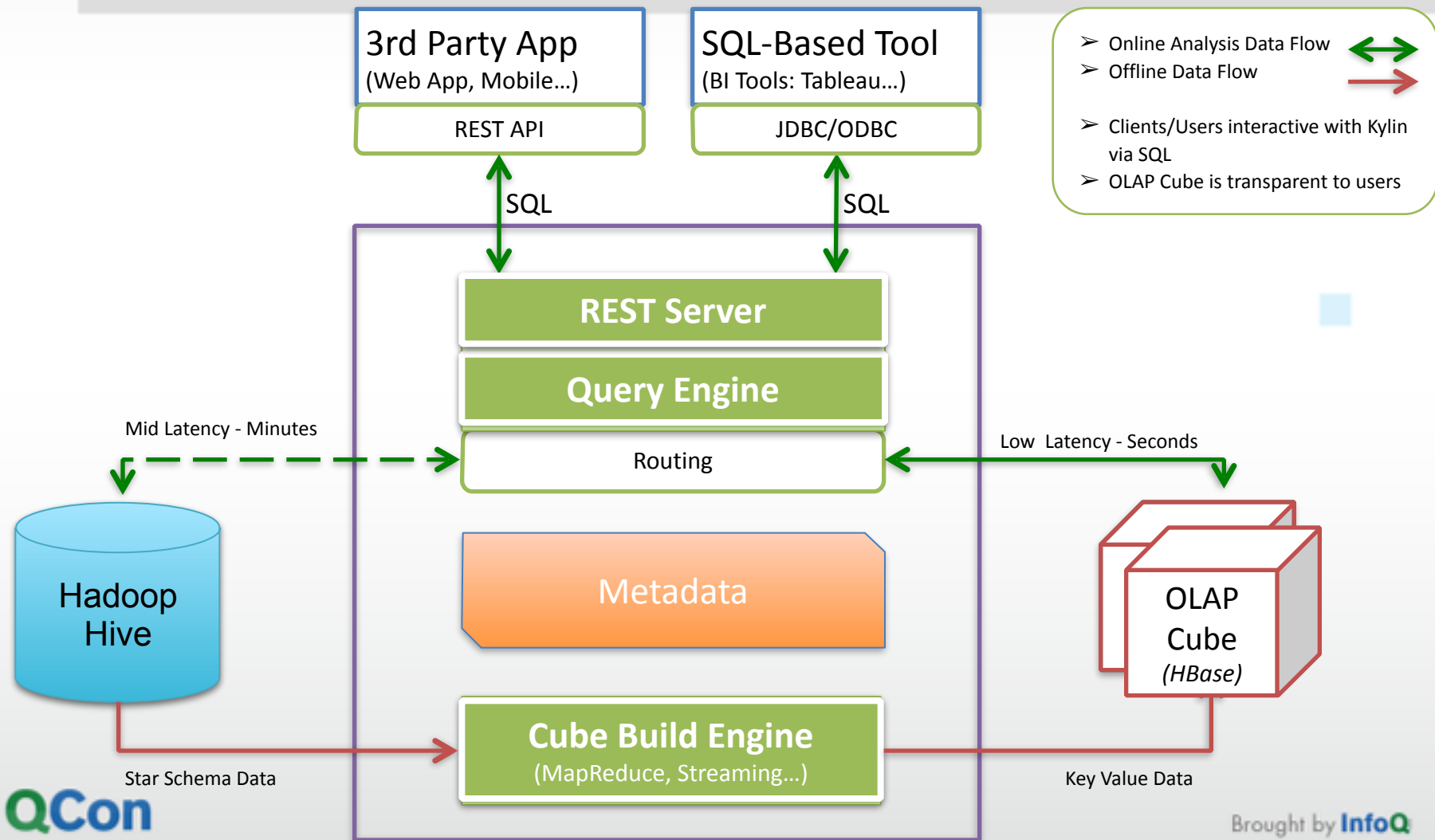
- First Apache Project open sourced by eBay Inc.
- First Apache Project fully contributed from eBay CCOE
- Open Sourced on Oct 1st, 2014
- Be accepted as Apache Incubator Project on Nov 25th, 2014
  - Apache Kylin is an effort undergoing incubation at The Apache Software Foundation (ASF), sponsored by 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

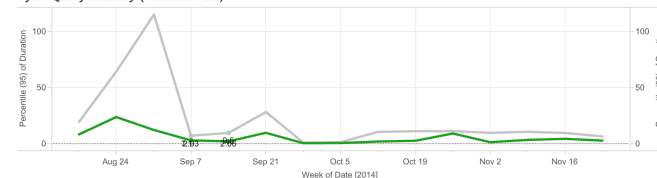


# Features

- 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

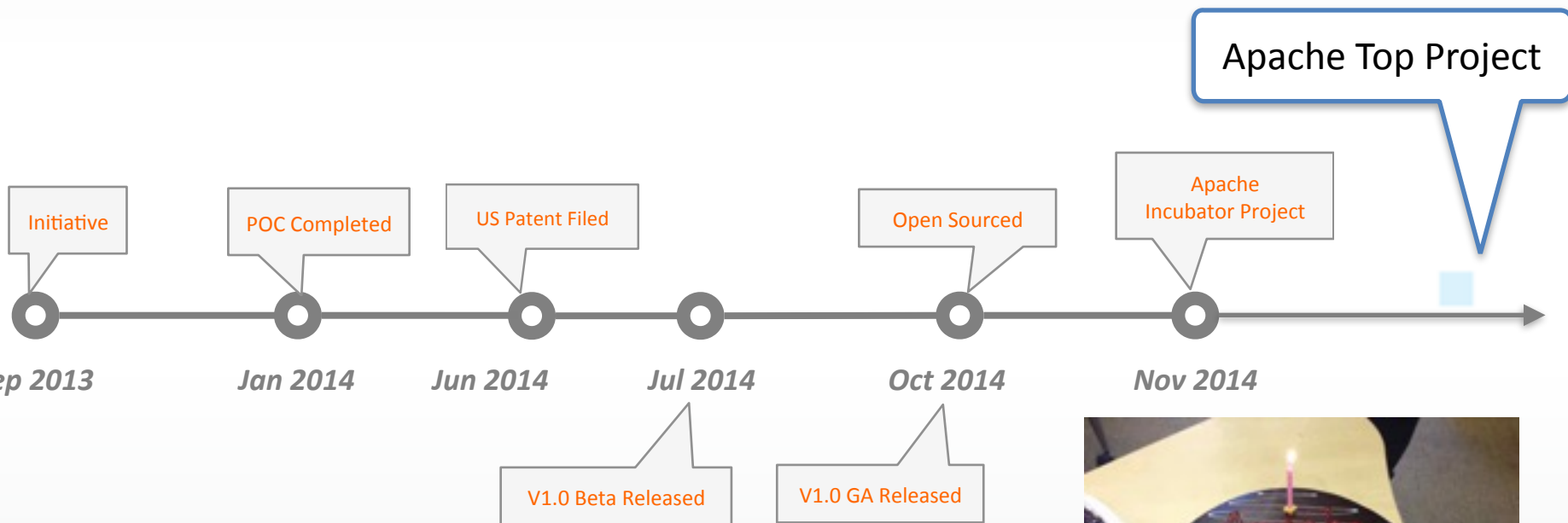
Kylin Query Latency (90 and 95%)



# 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

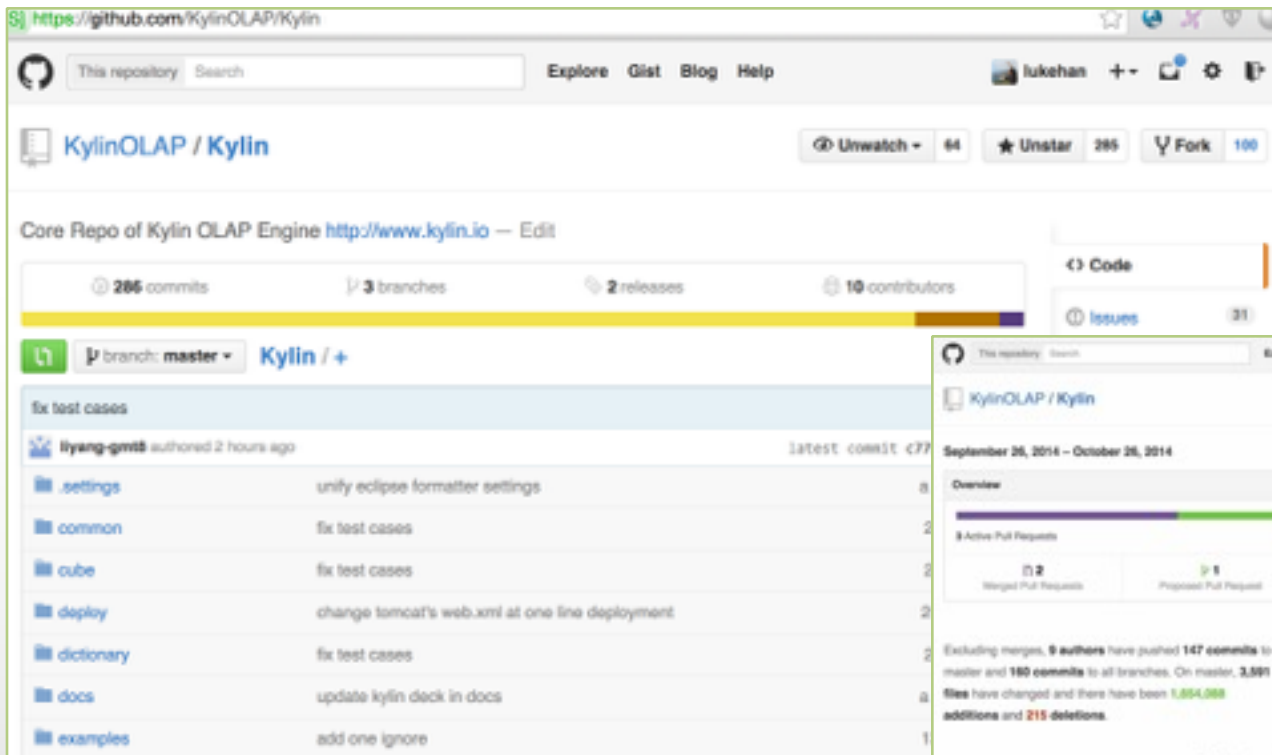
# Patent

- Why?
- How?
- Patent vs Open Source



# Phase I: Open Source on Github

- Code pushed to github.com on Oct 1<sup>st</sup>, 2014



https://github.com/KylinOLAP/Kylin

This repository Search Explore Gist Blog Help

KylinOLAP / Kylin

Core Repo of Kylin OLAP Engine <http://www.kylin.io> — Edit

286 commits 3 branches 2 releases 10 contributors

branch: master Kylin / +

File	Commit Message	Author	Time
fix test cases	fix test cases	liyang-gmt8	2 hours ago
.settings	unify eclipse formatter settings	liyang-gmt8	2 hours ago
common	fix test cases	liyang-gmt8	2 hours ago
cube	fix test cases	liyang-gmt8	2 hours ago
deploy	change tomcat's web.xml at one line deployment	liyang-gmt8	2 hours ago
dictionary	fix test cases	liyang-gmt8	2 hours ago
docs	update kylin deck in docs	liyang-gmt8	2 hours ago
examples	add one ignore	liyang-gmt8	2 hours ago



KylinOLAP / Kylin

September 26, 2014 – October 26, 2014

Overview

9 Active Pull Requests 17 Active Issues

2 Merged Pull Requests 1 Proposed Pull Request 8 Closed Issues 9 New Issues

Excluding merges, 9 authors have pushed 147 commits to master and 180 commits to all branches. On master, 3,591 files have changed and there have been 1,854,088 additions and 215 deletions.

2 Pull requests merged by 2 people

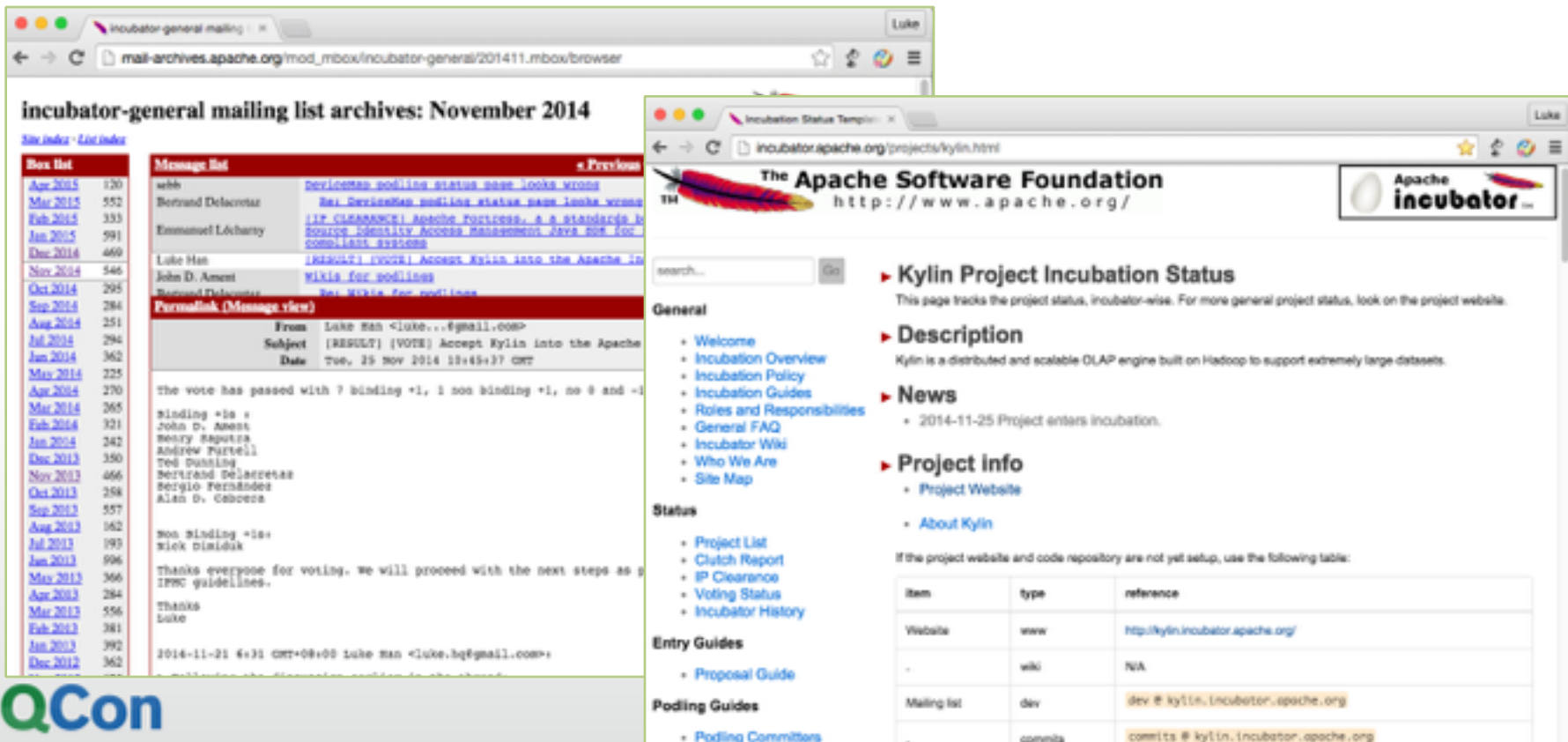
Merged #20 combine rowkey & aggregation to advanced\_setting 3 days ago

Merged #11 Fix Workflows type to README 3 days ago

1 Pull request proposed by 1 person

# Phase II: Apache Incubator

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



The screenshot shows two browser windows. The left window displays the 'incubator-general mailing list archives: November 2014' from mail-archives.apache.org. It includes a 'Box list' with dates and message counts, and a 'Message list' showing a vote on accepting Kylin into the Apache Incubator. The right window shows the 'The Apache Software Foundation' website for the 'Kylin Project Incubation Status'. It features a search bar, navigation links, and a table of project status items.

**incubator-general mailing list archives: November 2014**

Box list

Date	Count
Apr 2015	120
Mar 2015	552
Feb 2015	133
Jan 2015	591
Dec 2014	469
Nov 2014	546
Oct 2014	295
Sep 2014	284
Aug 2014	251
Jul 2014	294
Jun 2014	362
May 2014	225
Apr 2014	270
Mar 2014	365
Feb 2014	321
Jan 2014	242
Dec 2013	350
Nov 2013	466
Oct 2013	258
Sep 2013	557
Aug 2013	162
Jul 2013	193
Jun 2013	996
May 2013	366
Apr 2013	284
Mar 2013	556
Feb 2013	381
Jan 2013	392
Dec 2012	362

Message list

From: Luke Han <luke...@gmail.com>  
Subject: [RESULT] [VOTE] Accept Kylin into the Apache Incubator  
Date: Tue, 25 Nov 2014 13:45:17 GMT

The vote has passed with 7 binding +1, 1 non binding +1, no 0 and -1

binding +1s:  
John D. Anest  
Henry Saputra  
Andrew Purtell  
Ted Dunning  
Bertrand Delarretas  
Sergio Fernandez  
Alan D. Cabrera

non binding +1s:  
Nick Dimadis

Thanks everyone for voting. We will proceed with the next steps as per IPWC guidelines.

Thanks  
Luke

2014-11-21 4:31 GMT+08:00 Luke Han <luke.hq@gmail.com>:

**The Apache Software Foundation**  
http://www.apache.org/

**Kylin Project Incubation Status**  
This page tracks the project status, incubator-wise. For more general project status, look on the project website.

**Description**  
Kylin is a distributed and scalable OLAP engine built on Hadoop to support extremely large datasets.

**News**  
• 2014-11-25 Project enters incubation.

**Project info**  
• Project Website  
• About Kylin

If the project website and code repository are not yet setup, use the following table:

Item	type	reference
Website	www	http://kylin.incubator.apache.org/
-	wiki	N/A
Mailing list	dev	dev@kylin.incubator.apache.org
-	commits	commits@kylin.incubator.apache.org

**General**

- Welcome
- Incubation Overview
- Incubation Policy
- Incubation Guides
- Roles and Responsibilities
- General FAQ
- Incubator Wiki
- Who We Are
- Site Map

**Status**

- Project List
- Clutch Report
- IP Clearance
- Voting Status
- Incubator History

**Entry Guides**

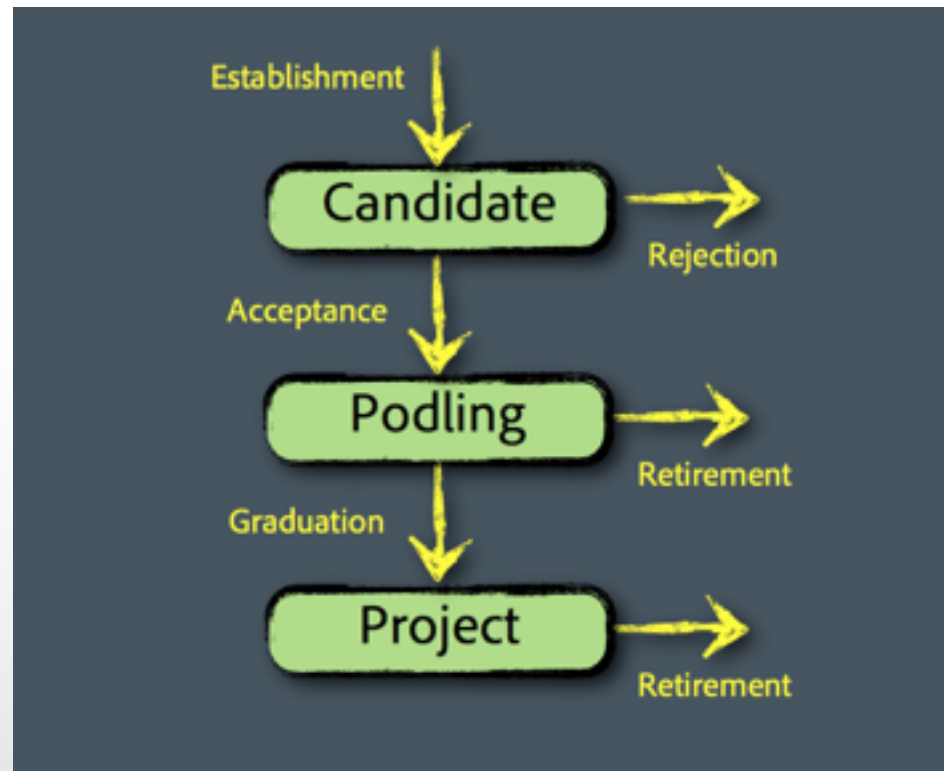
- Proposal Guide

**Podling Guides**

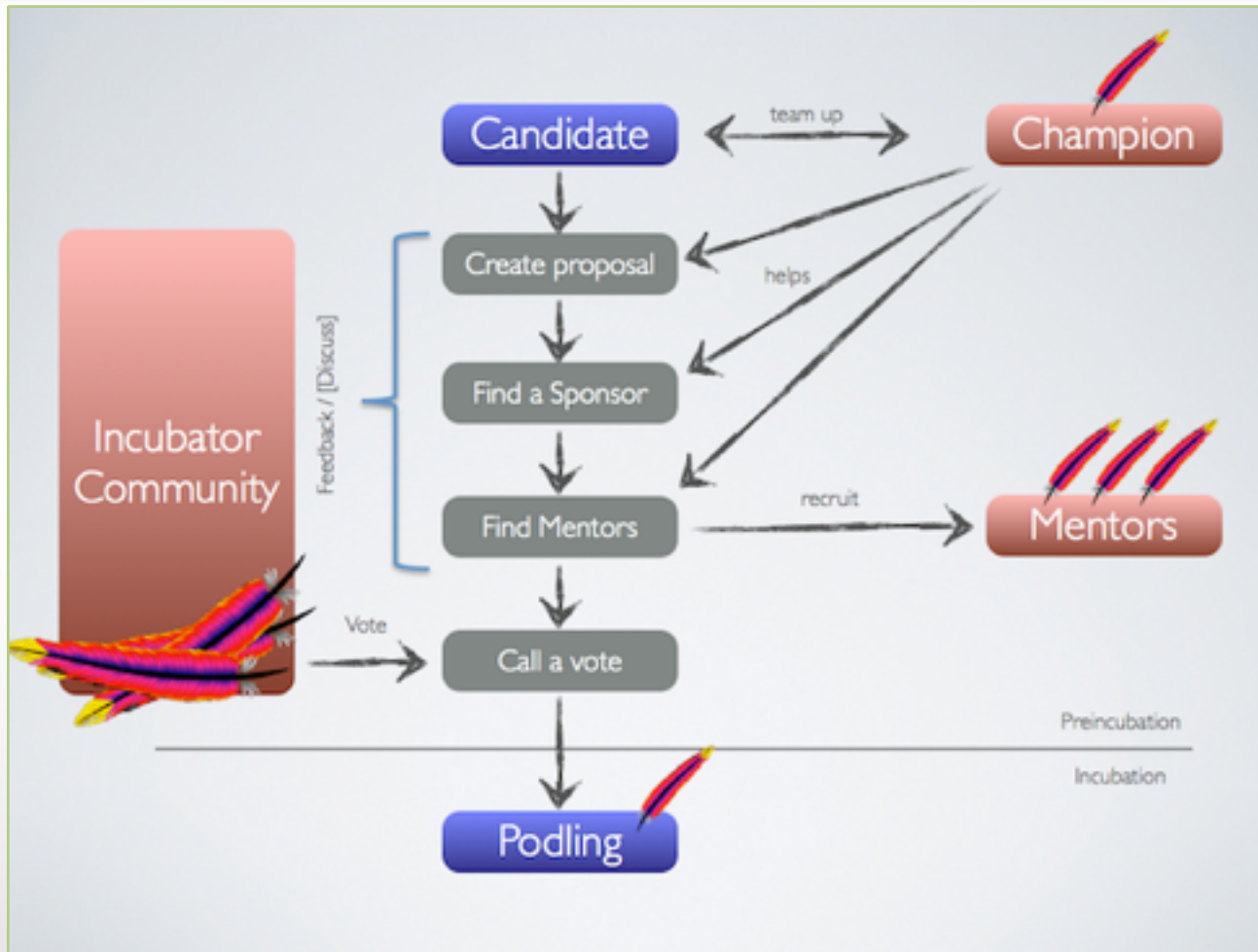
- Podling Committers

# Why & How Apache?

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

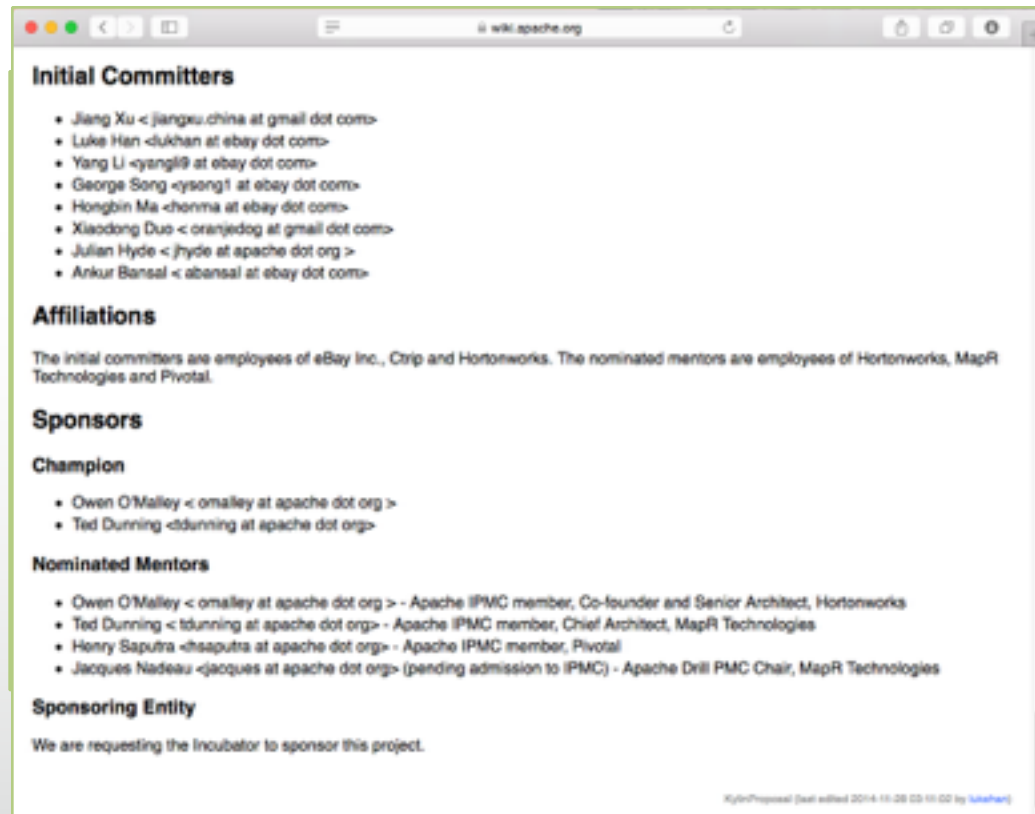


# Incubation Progress



# Incubator Project Proposal

- IPMC & PPMC
- Mentors and Champion
- Committers



The screenshot shows a web browser window displaying the Apache Incubator Project Proposal document. The document is titled "Initial Committers" and lists the following members:

- Jiang Xu <jiangxu.china at gmail dot com>
- Luke Han <dukhan at ebay dot com>
- Yang Li <lyangli9 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>

The document also includes sections for "Affiliations", "Sponsors", "Champion", "Nominated Mentors", and "Sponsoring Entity".

**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.

Kyle/Proposed [last edited 2014-11-28 03:10:02 by LukeHan]

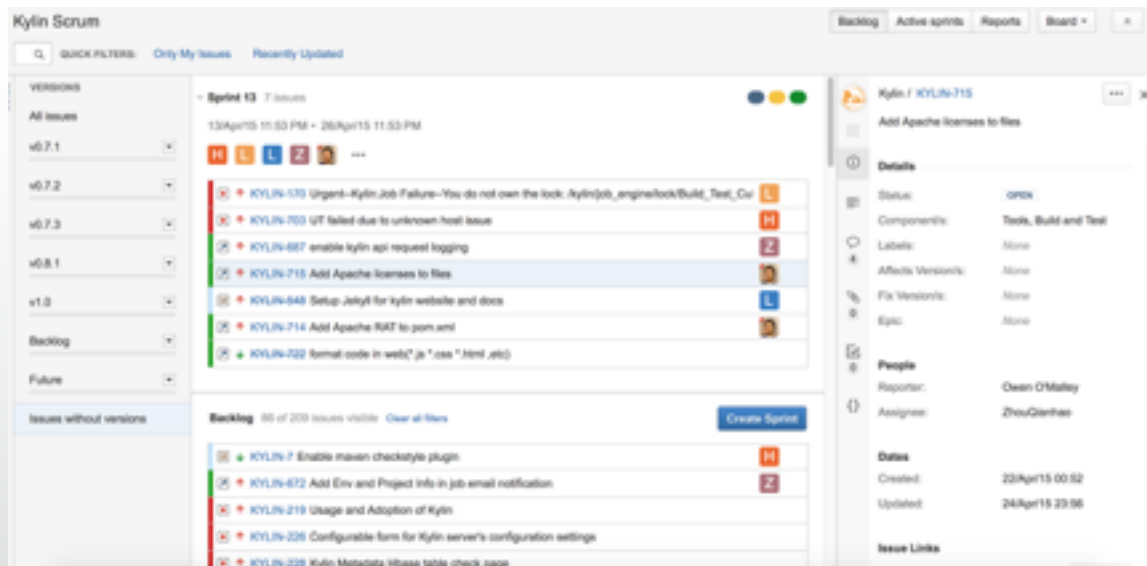
# Agenda

- 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, NOTICES, 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/commmitter

# How to contribute?

- Join mailing list:
  - [dev@kylin.incubator.apache.org](mailto: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!



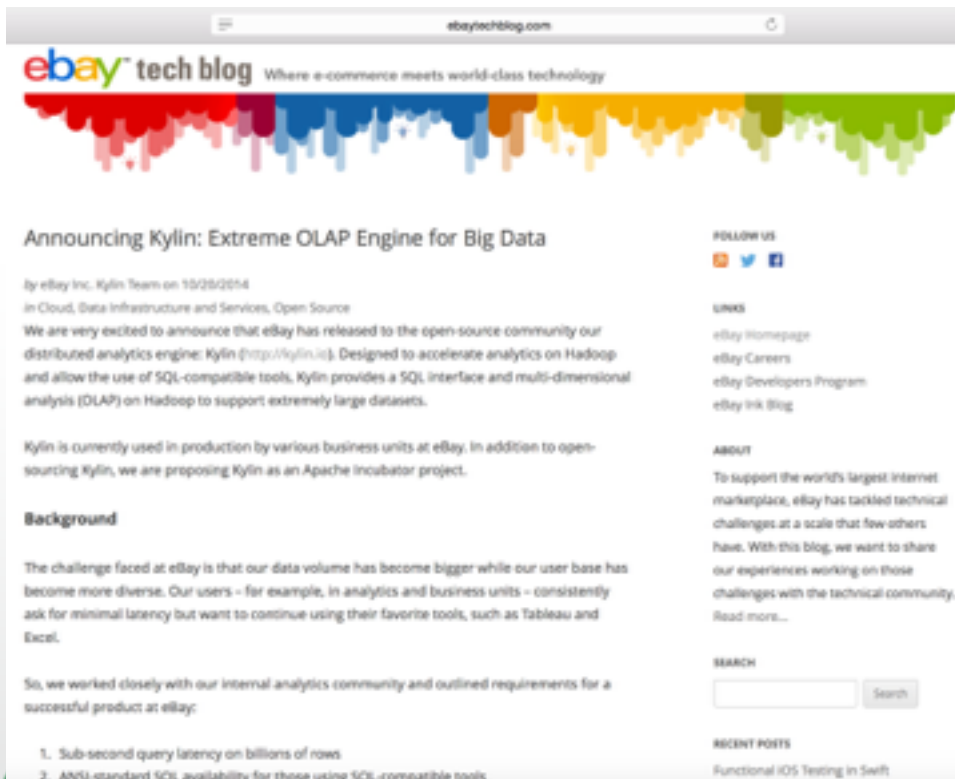
# Marketing - Website

- <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>



The screenshot shows the eBay Tech Blog interface. The header features the 'ebay tech blog' logo with the tagline 'Where e-commerce meets world-class technology'. The main article is titled 'Announcing Kylin: Extreme OLAP Engine for Big Data' and is dated 'by eBay Inc. Kylin Team on 10/20/2014'. The article text states: 'In Cloud, Data Infrastructure and Services, Open Source We are very excited to announce that eBay has released to the open-source community our distributed analytics engine: Kylin (http://kylin.taobao.com). Designed to accelerate analytics on Hadoop and allow the use of SQL-compatible tools, Kylin provides a SQL interface and multi-dimensional analysis (OLAP) on Hadoop to support extremely large datasets.' Below the text, there is a 'Background' section starting with 'The challenge faced at eBay is that our data volume has become bigger while our user base has become more diverse...' and a 'SEARCH' bar. On the right side, there are 'FOLLOW US' links for Twitter and Facebook, and a 'LINKS' section with links to 'eBay Homepage', 'eBay Careers', 'eBay Developers Program', and 'eBay Ink Blog'.



*"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](http://kylin.io)
- LinkedIn
  - Group: Kylin
- WeChat(微信)
  - ApacheKylin
- ...



# Marketing - Media

- InfoQ
- CSDN
- OSChina
- ...



# Build Community – Mailing List



Re: 关于关联表的cube创建和查询 4 messages in org.apache.kylin.dev

From	Sent On	Attachments
wanghaifei	Mar 30, 2015 7:41 pm	
Ted Dunning	Mar 30, 2015 11:38 pm	
Li Yang	Apr 4, 2015 5:29 pm	
Shi, Shaofeng	Apr 6, 2015 3:38 am	

Subject: Re: 关于关联表的cube创建和查询  
From: Ted Dunning (ted\_dunning@gmail.com)  
Date: Mar 30, 2015 11:38:25 pm  
List: org.apache.kylin.dev

The standard language of Apache mailing lists is English. I do know that a lot of contributors speak Chinese, however, so could somebody contribute a translation in addition to answering this question?

2015-03-30 19:41 GMT-07:00 wanghaifei <wang\_hf@jd.com>:

你好，  
我在测试cube时遇到几个问题。

这里有个例子：有个事实表 fact1 和对照表 dim，我对他们进行关联。

建立了如下关联关系：

建立的维度如下：

问题1：  
在最后一步保存cube时出现这个错误：Cannot find rowkey column PROVINCE\_ID in cube-  
但是，我在建立维度时去掉对照表的省id，cube就创建成功。  
但，这会造成无法在cube查询时，进行关联操作，只能进行单表查询(个人观点)，不知道是不是建cube出问题，还是查询sql 写法的问题？

问题2：  
这里成功运行完cube以后出来的表示，用的hive元表中的表名，而非特定的cube名称(只有在执行查询时，才会出现对应的cube)，这就不支持由文件路径？

# 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
- ...



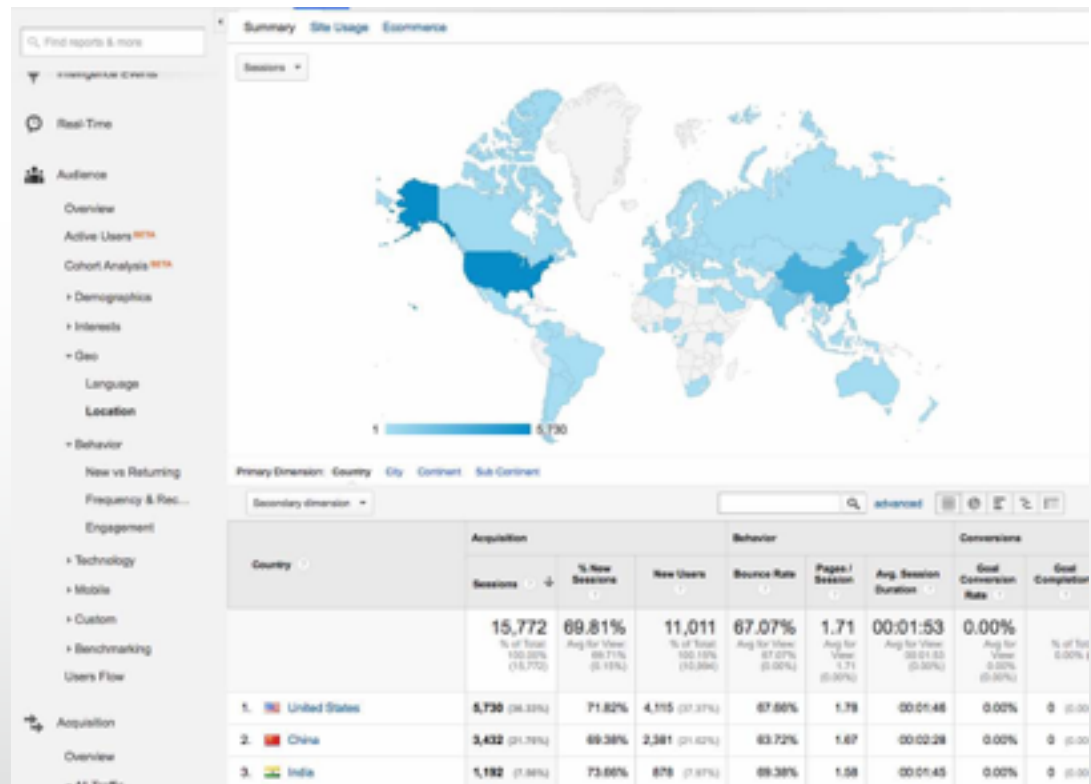
# Build Community – Conference

- 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
- ...





# Apache Kylin Ecosystem

- **Kylin Core**

- Fundamental framework of Kylin OLAP Engine

- **Extension**

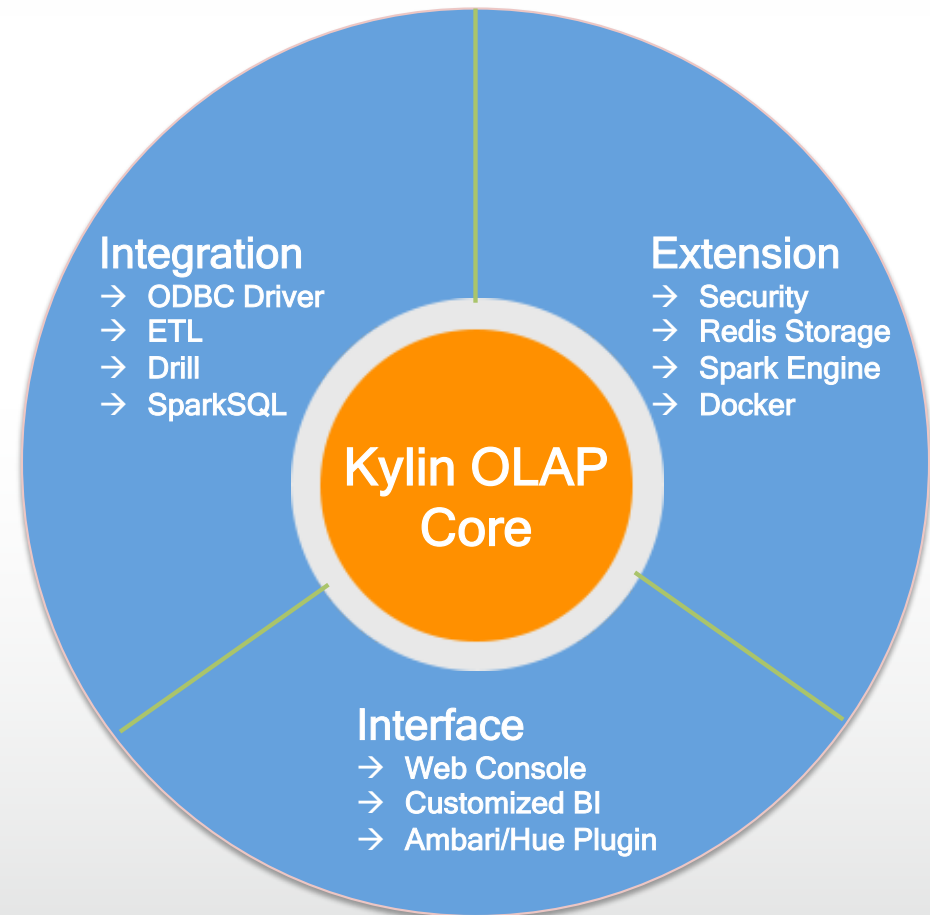
- Plugins to support for additional functions and features

- **Integration**

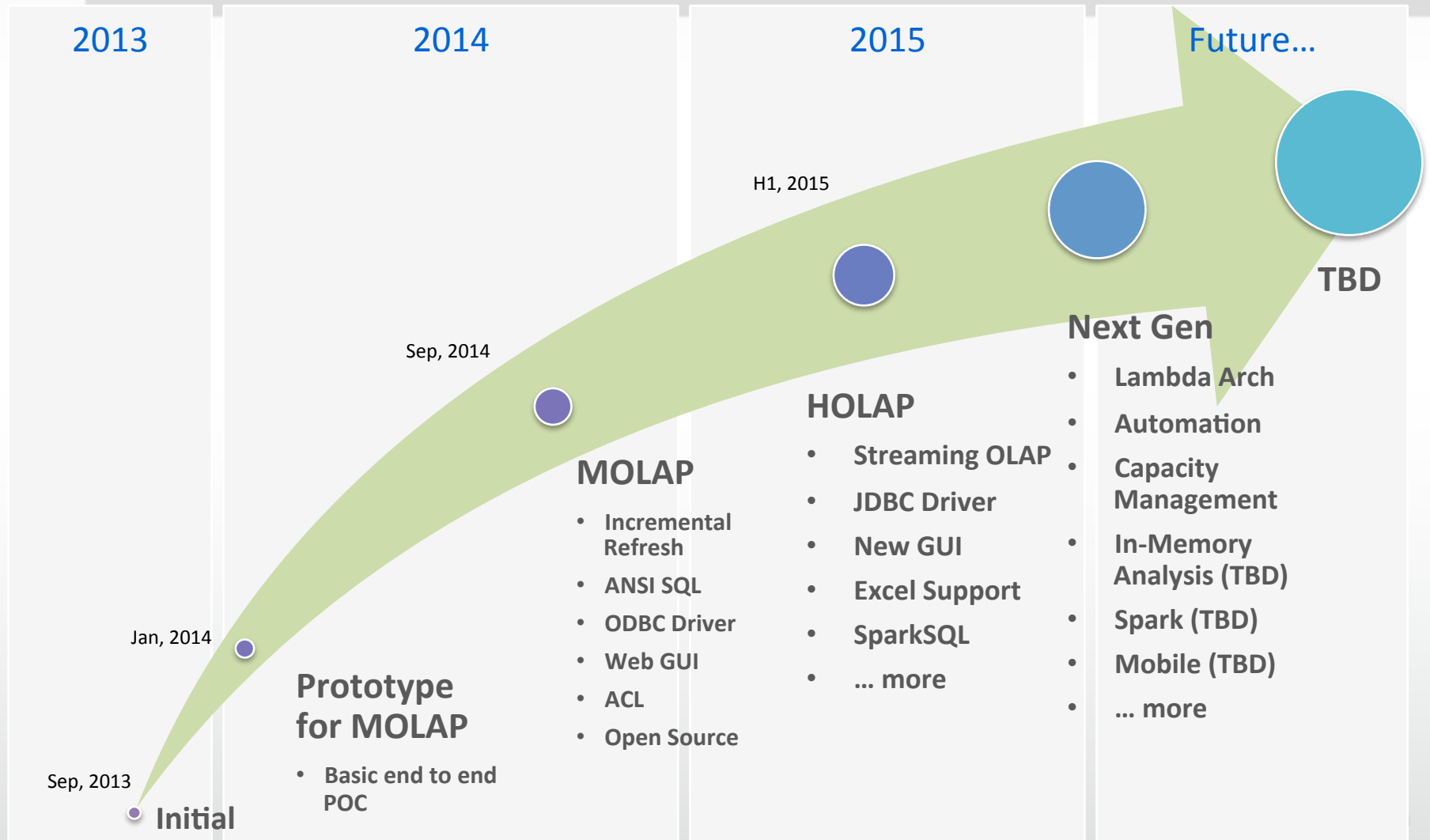
- Lifecycle Management Support to integrate with other applications like BI tools

- **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





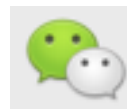
# Apache Kylin

- Kylin Site:
  - <http://kylin.incubator.apache.org>
  - <http://kylin.io>
- Twitter:
  - [@ApacheKylin](https://twitter.com/ApacheKylin)
- WeChat(微信)
  - ApacheKylin





@InfoQ



infoqchina

软件  
正在改变世界!