A **very** short talk about Apache Kylin
Business Intelligence meets Big Data

Fabian Wilckens
EMEA Solutions Architect

The challenge today …

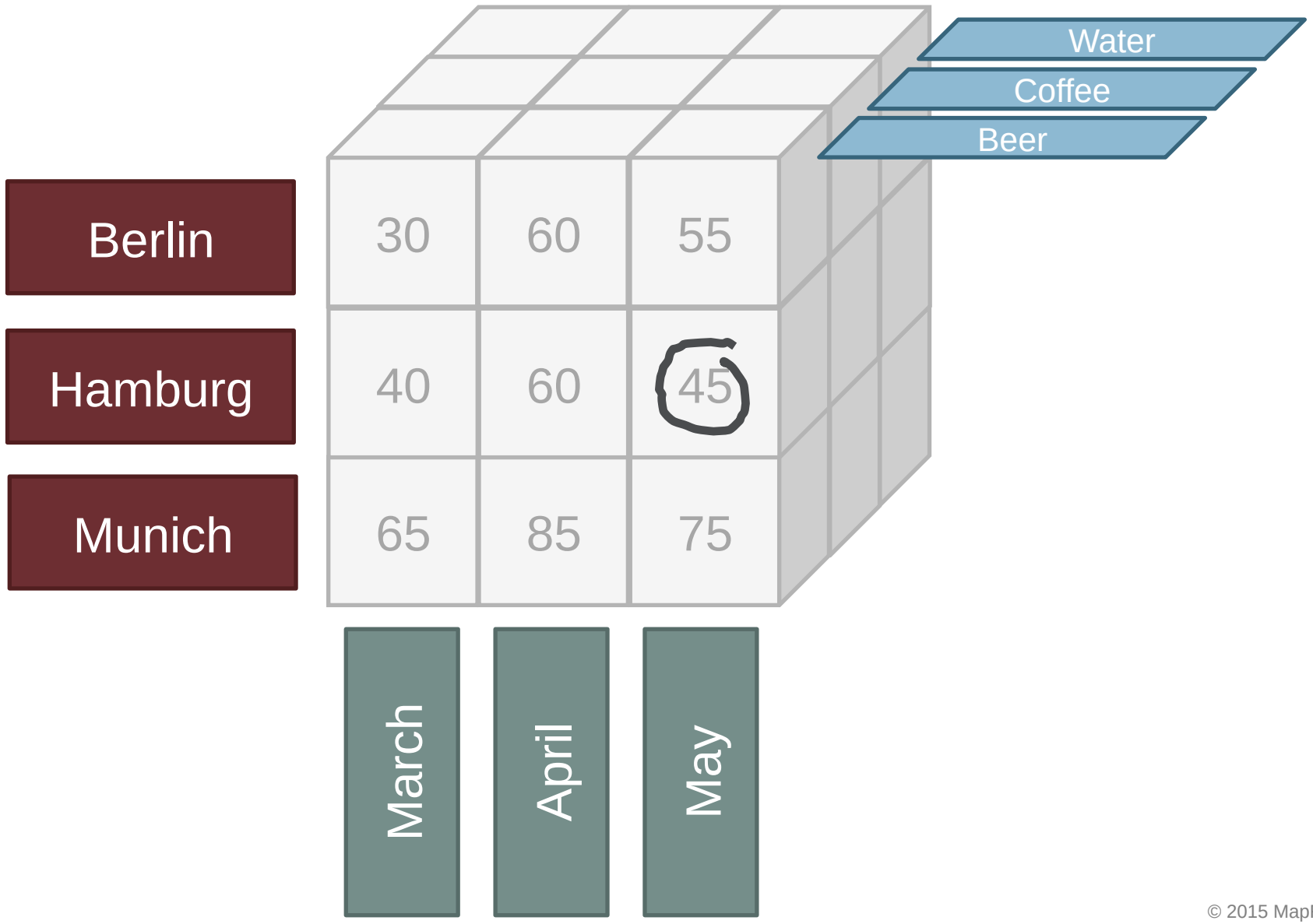
Very quickly: OLAP

Online Analytical Processing

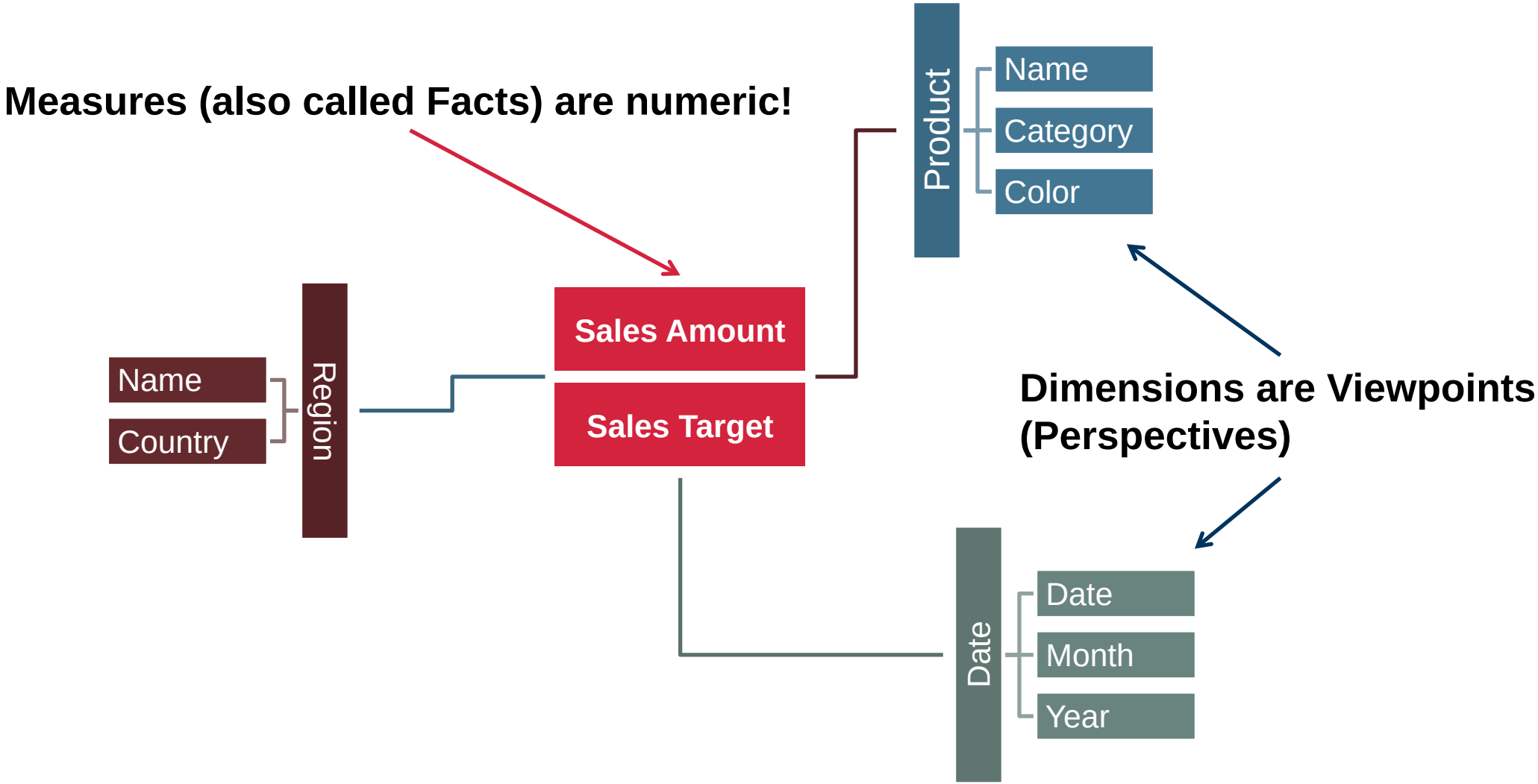
How many beers were ordered in Germany on a yearly basis?

How many beers were ordered in Germany on a yearly basis, broken down by months?

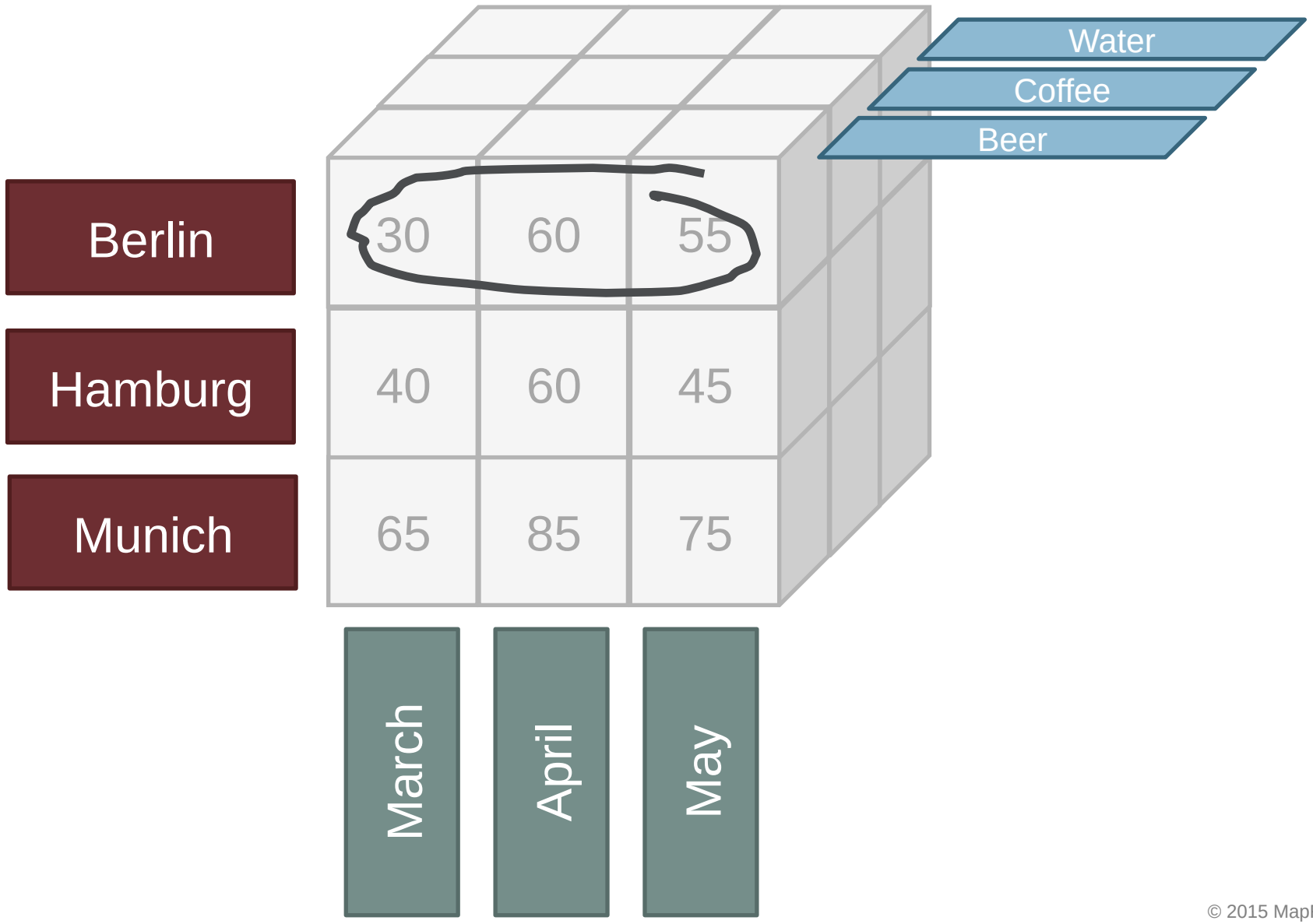
OLAP Cubes



Cubes: Measures and Dimensions

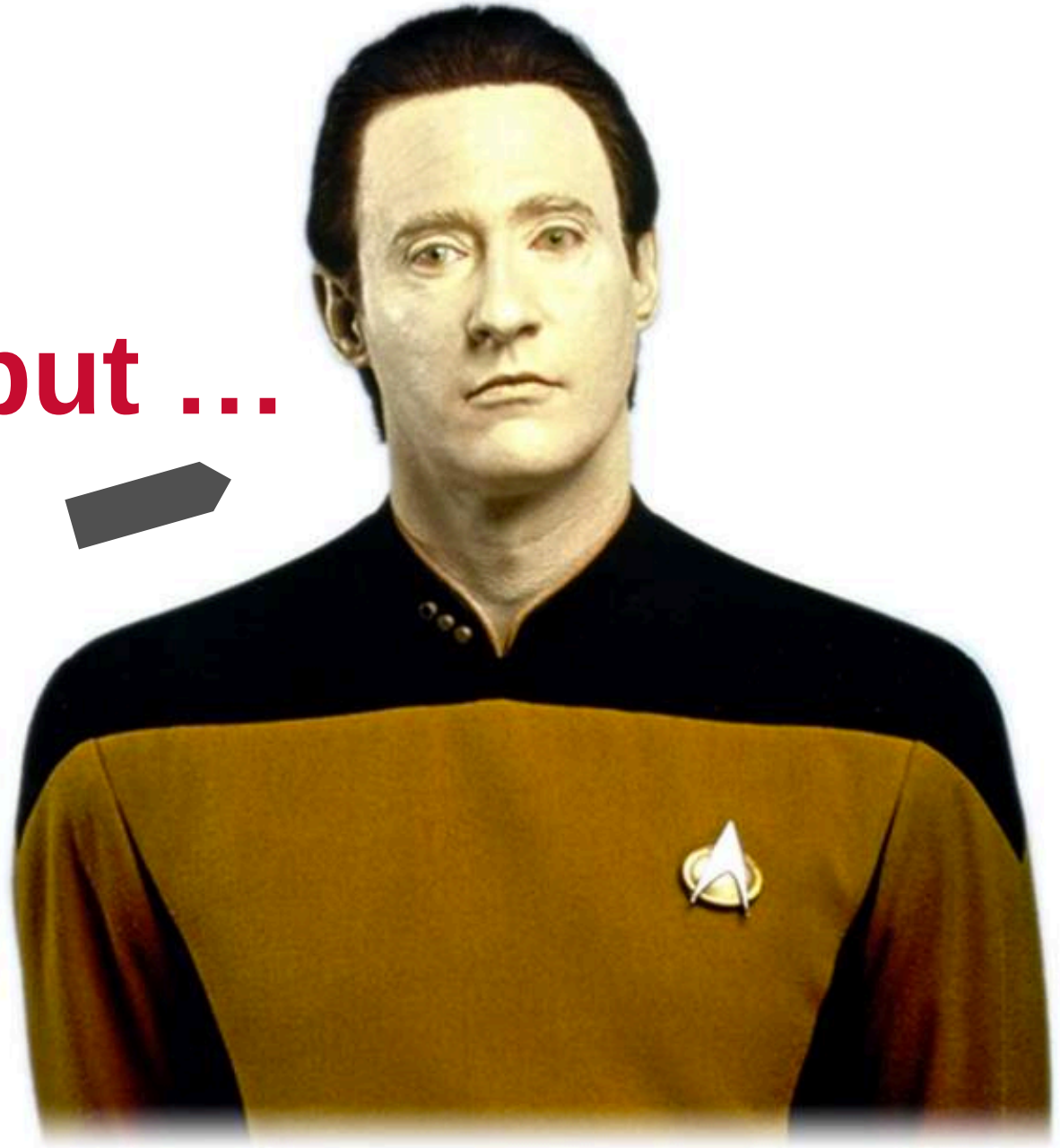


OLAP Cubes



That’s all great but …

How about **Big Data**



What is Kylin?

kylin / ˈkiːˈlɪn / 麒麟

--n. (in Chinese art) a mythical animal of composite form

Extreme OLAP Engine for Big Data

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

- Open Sourced on Oct 1st, 2014
- Accepted into incubation November, 2014
- Preparing for first Apache release

Goals

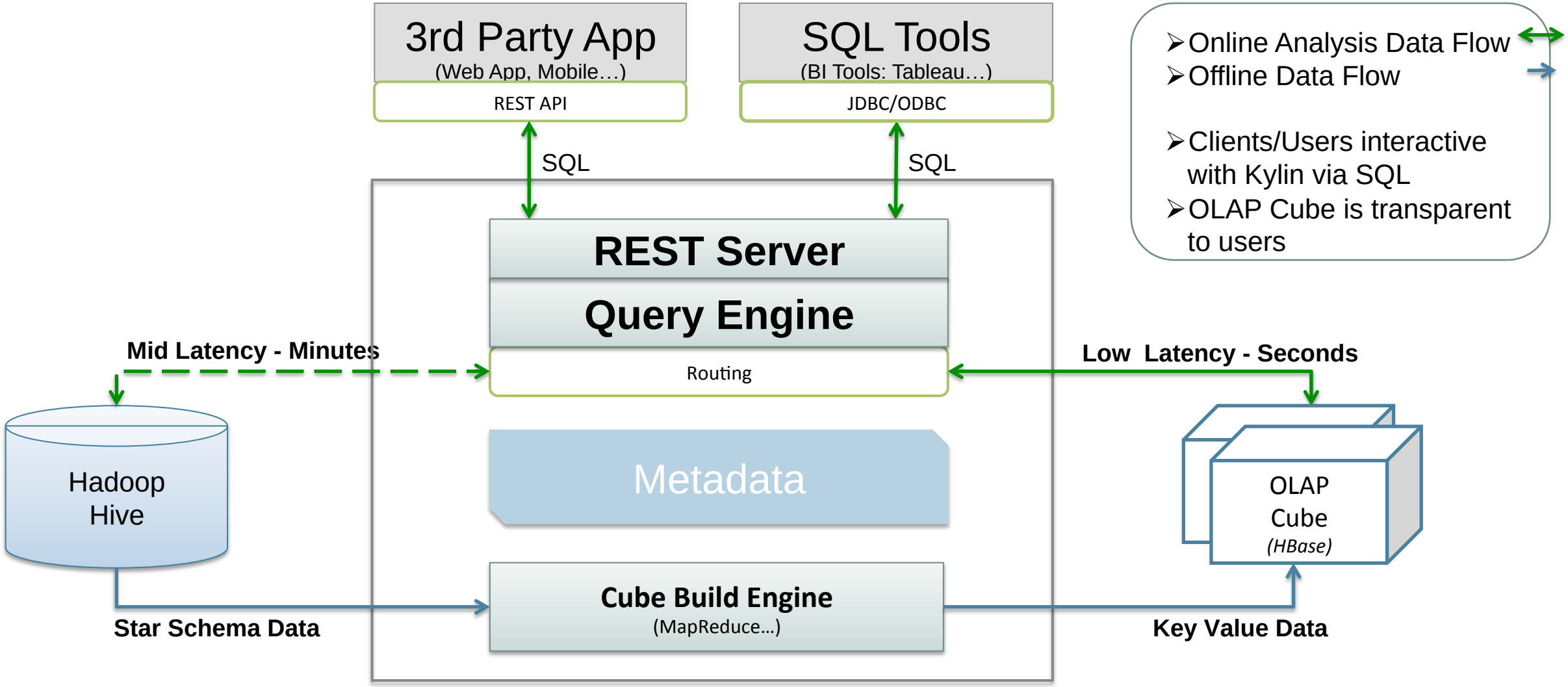
- Sub-second query latency on billions of rows
- ANSI SQL for both analysts and engineers
- Full OLAP capability to offer advanced functionality
- Seamless Integration with BI Tools

- Support for high cardinality and dimensionality
- High concurrency – thousands of end users
- Distributed and scale out architecture for large data volume

Kylin Depends on Hadoop Eco-system

- Hive
  - Input source, pre-join star schema during cube building
- MapReduce
  - Aggregate metrics during cube building
- HDFS
  - Store intermediate files during cube building
- HBase
  - Store and query data cubes
- Calcite
  - SQL parsing, code generation, optimization

Kylin Architecture Overview



Kylin Highlights

- ***Extremely Fast OLAP Engine at Scale***

Kylin is designed to reduce query latency on Hadoop for 10+ billions of rows of data to seconds

- ***ANSI SQL Interface on Hadoop***

Kylin offers ANSI SQL on Hadoop and supports most ANSI SQL query functions

- ***Seamless Integration with BI Tools***

Kylin currently offers integration capability with BI Tools like Tableau.

- ***Interactive Query Capability***

Users can interact with Hadoop data via Kylin at sub-second latency

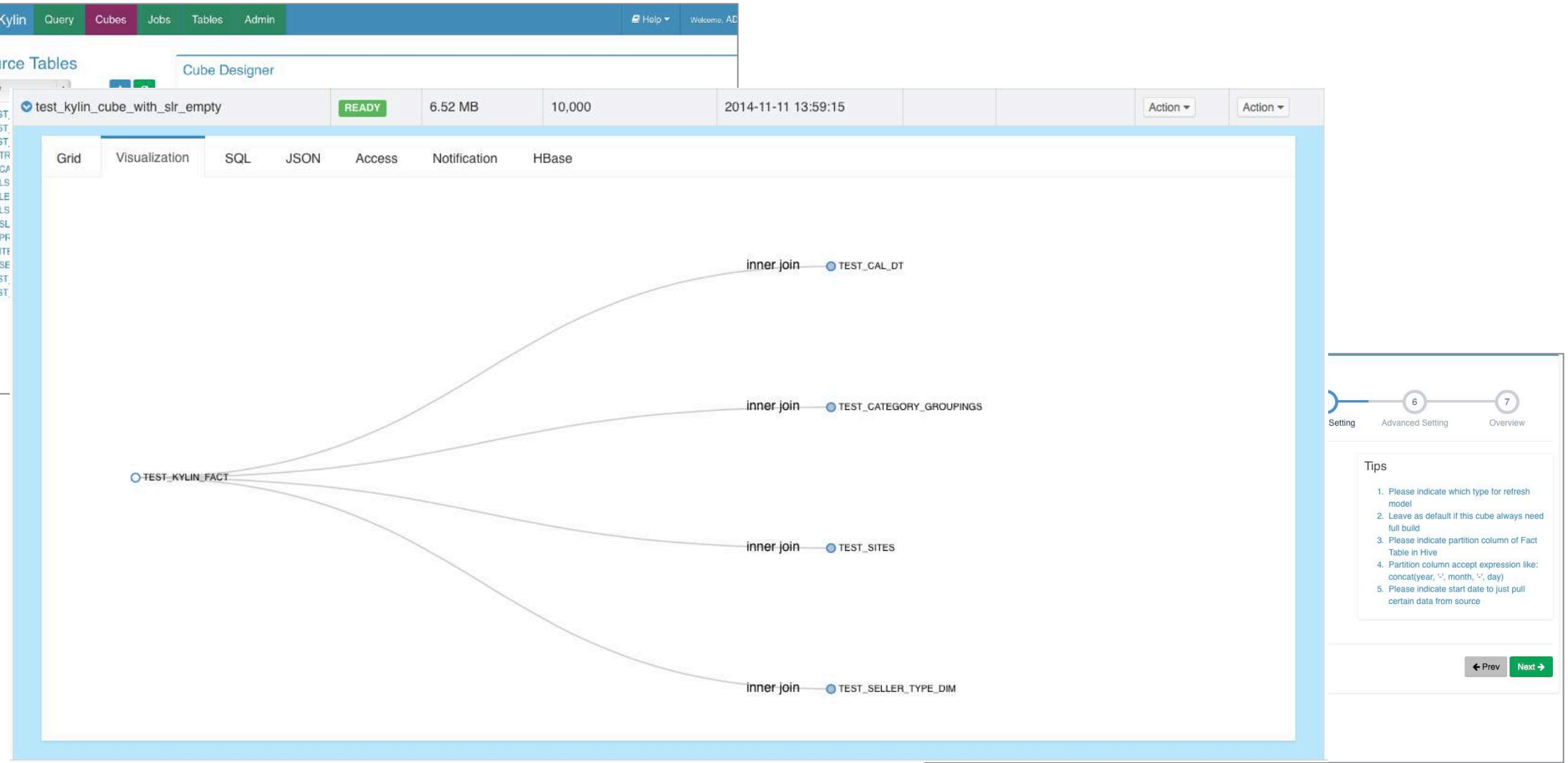
- ***MOLAP Cube***

User can define a data model and pre-build in Kylin with more than 10+ billions of raw data records

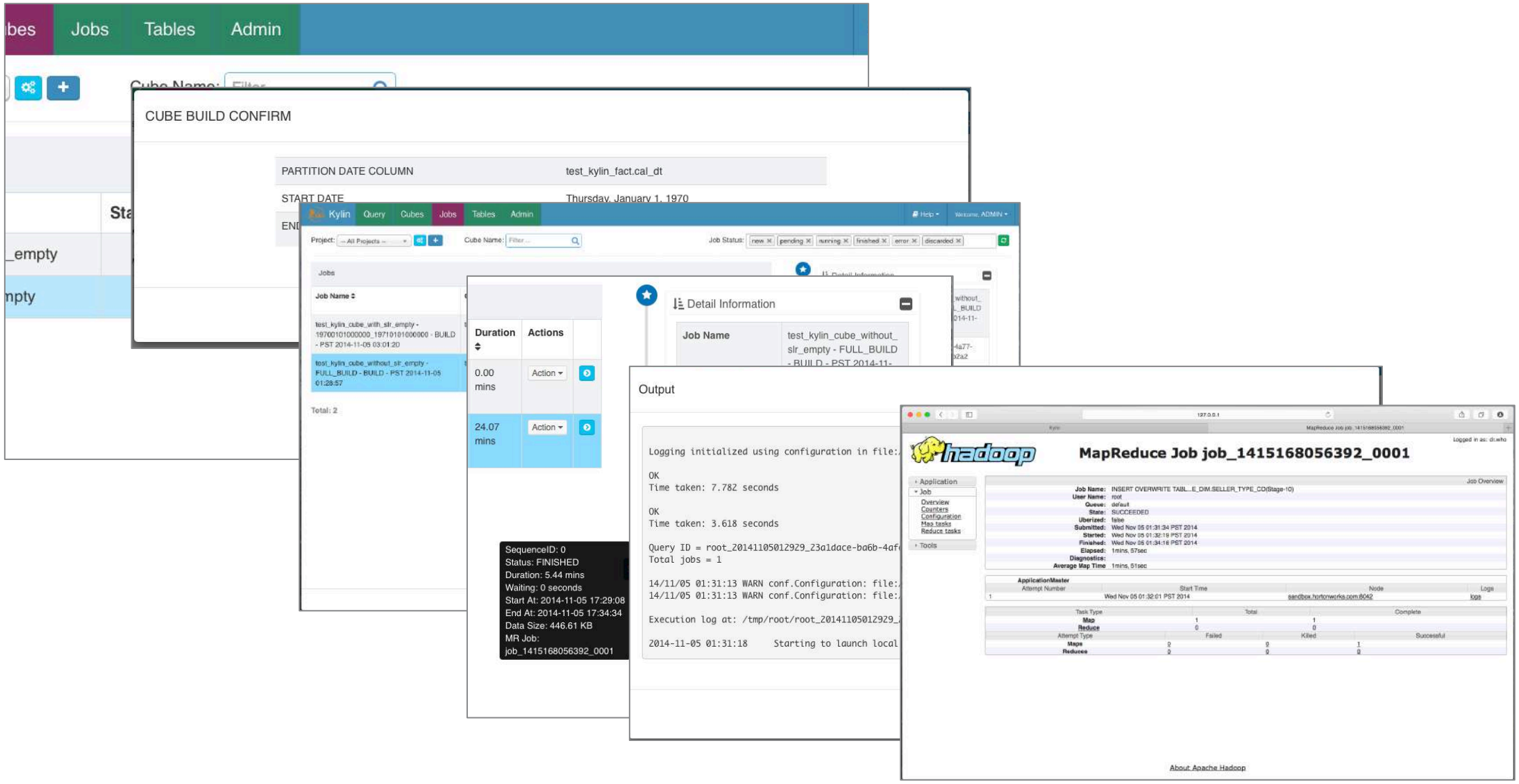
More Highlights

- Compression and Encoding Support
- Incremental Refresh of Cubes
- Approximate Query Capability for distinct Count (HyperLogLog)
- Leverage HBase Coprocessor for query latency
- Job Management and Monitoring
- Easy Web interface to manage, build, monitor and query cubes
- Security capability to set ACL at Cube/Project Level
- Support LDAP Integration

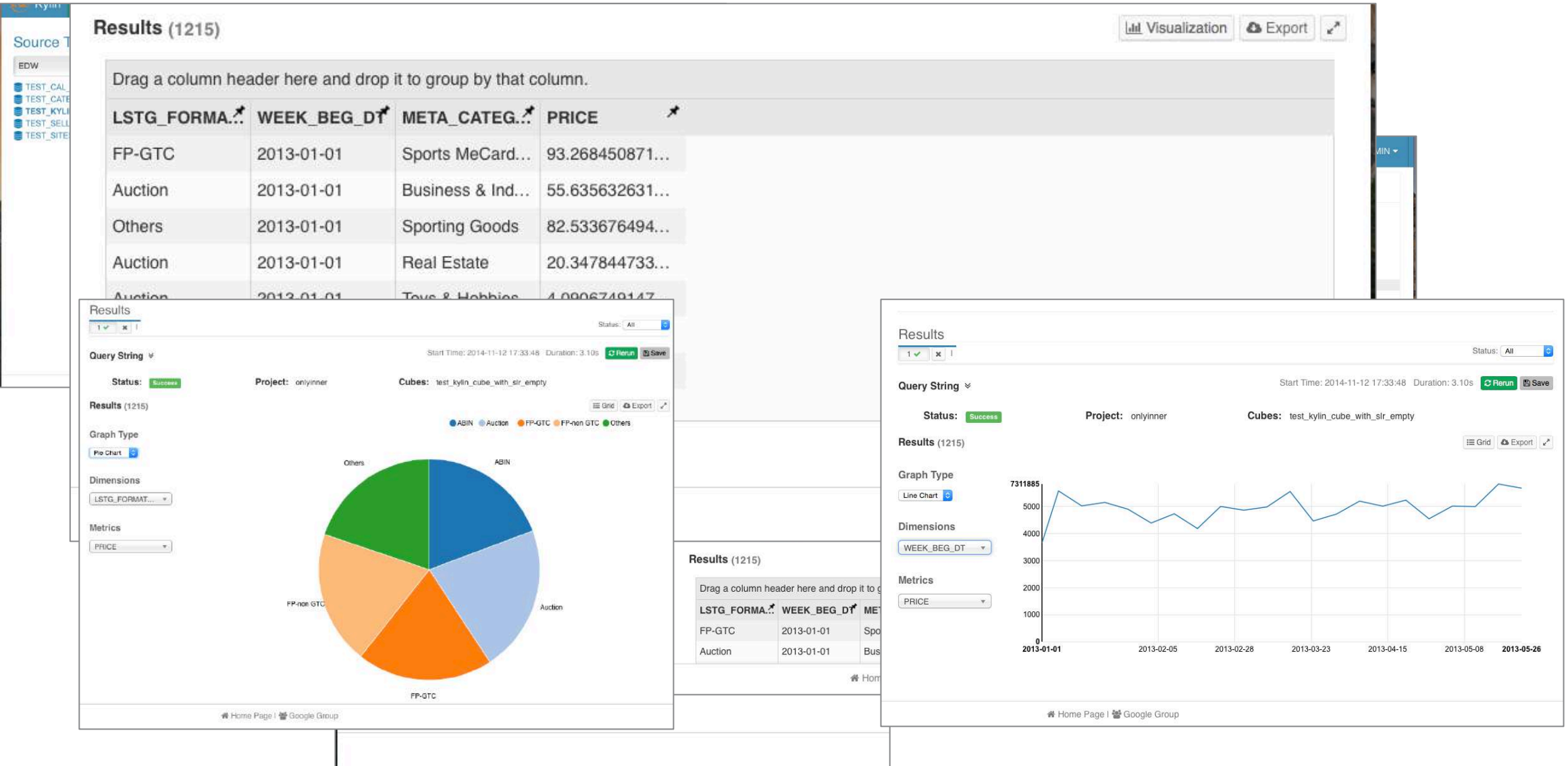
Cube Designer



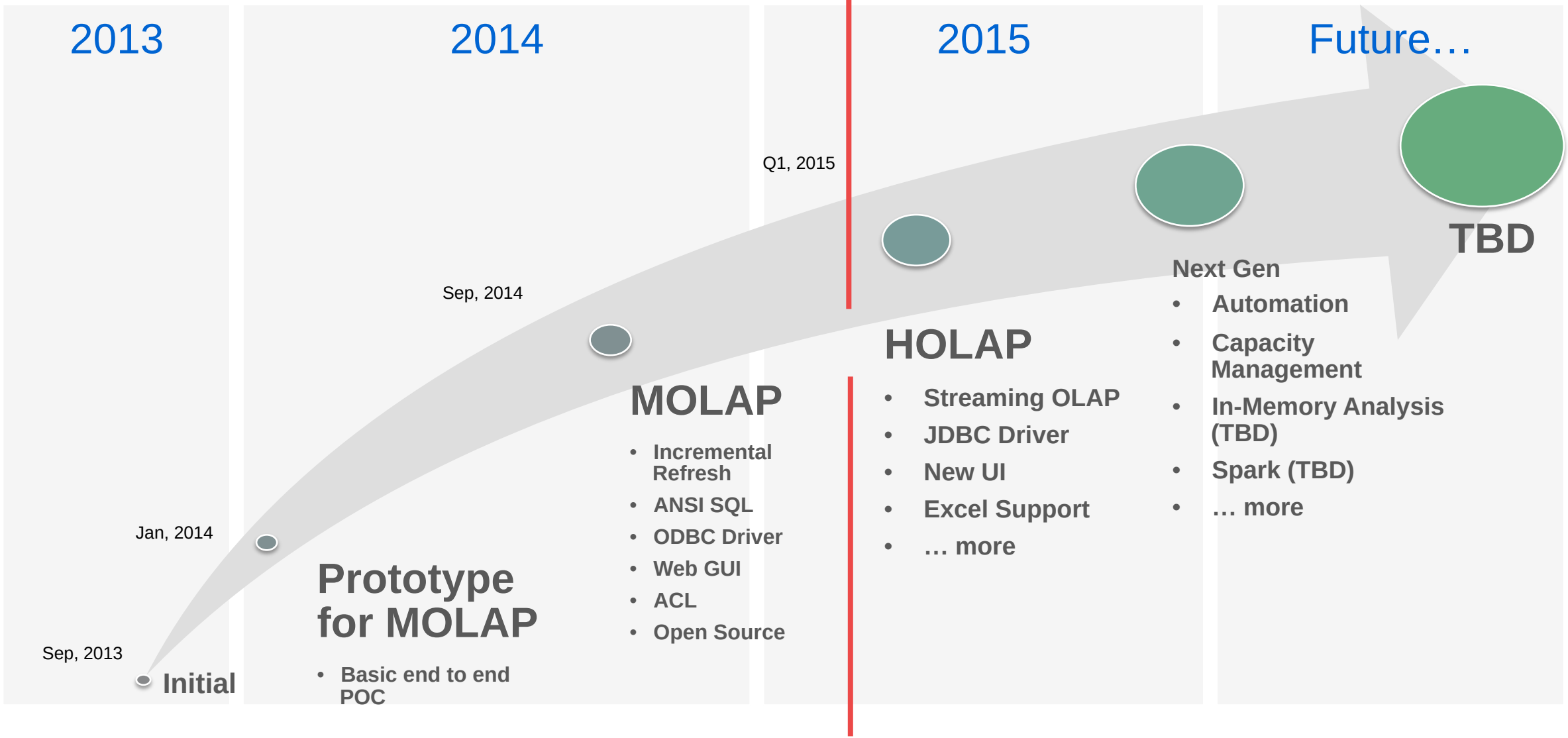
Job Management



Query and Visualization



Kylin History and Roadmap



Open Source

- Kylin Site:
  - http://kylin.io
- Twitter:
  - @ApacheKylin
- Source Code Repo:
  - https://github.com/KylinOLAP
- Google Group:
  - Kylin OLAP
- 微信
  - ApacheKylin

Fabian Wilckens
fwilckens@mapr.com