Building Enterprise OLAP on Hadoop for Financial Services Industry

Luke Han
luke@kyligence.io | @lukehq
Co-founder & CEO of Kyligence
Creator & VP of Apache Kylin
Microsoft Regional Director & MVP
About Kyligence

- Formed by creators of Apache Kylin in 2016
- Offers Enterprise and Cloud version of Apache Kylin
- Funding from Redpoint, Cisco, CBC and Shunwei
- Member of Microsoft Accelerator Shanghai 2017
- Dual HQ in Silicon Valley & Shanghai, China
Transition to Big Data...

How about your traditional data warehouse?
How about your existing OLAP/BI application?
Data Warehouse/OLAP in Financial Services Industry

- The biggest industry rely on DW/OLAP application
- Thousands applications build on top of EDW
- Experienced analysts with decade expertise
  ...in data...but not in technologies
Enterprise Data Warehouse Architecture

- Optimized for mission-critical analytics
- Well modeling
- Best practices of industry
- Thriving ecosystem
- Trained experts everywhere
But you are asked to...

- Migrate or build existing OLAP/BI app to Big Data
- Better performance…just because you have Big Data now
- Train yourself to learn MR/Spark/ML…and AI
OLAP: The Missing Part of Big Data

- Too many options
- Low performance
- Long learning curve
- Compatibility issue
- Technology vs Data

Presentation Visualization

Data Lake

Hive Impala Spark SQL Drill MapReduce Spark ...

Data Source
Apache Kylin: Bring OLAP back to Big Data

- MOLAP on Hadoop
- Simplified Data Modeling
- Optimized for aggregation query
- ANSI SQL
- Native on Hadoop
- On-Prem & In the Cloud
Kylin vs Hive: Star-Schema Benchmark

Apache Kylin vs. Apache Hive (lower is better)

* Based on 4 Nodes, 16 Core CPU, 96 GB Memory per node
### Global Users

<table>
<thead>
<tr>
<th>Internet</th>
<th>FSI</th>
<th>Telecom</th>
<th>Manufacturing</th>
<th>Others</th>
</tr>
</thead>
<tbody>
<tr>
<td>- eBay</td>
<td>- ABC</td>
<td>- China Mobile</td>
<td>- SAIC</td>
<td>- MachineZone</td>
</tr>
<tr>
<td>- Yahoo! Japan</td>
<td>- CCB</td>
<td>- China Telecom</td>
<td>- HUAWEI</td>
<td>- Glispa</td>
</tr>
<tr>
<td>- Baidu</td>
<td>- CMB</td>
<td>- Chine Unicom</td>
<td>- Lenovo</td>
<td>- Inovex</td>
</tr>
<tr>
<td>- Meituan</td>
<td>- CPIC</td>
<td>- AT &amp; T</td>
<td>- OPPO</td>
<td>- Adobe</td>
</tr>
<tr>
<td>- NetEase</td>
<td>- Citic Bank</td>
<td>- China Unionpay</td>
<td>- XIAOMI</td>
<td>- iFLYTEC</td>
</tr>
<tr>
<td>- Expedia</td>
<td>- China Unionpay</td>
<td></td>
<td>- VIVO</td>
<td></td>
</tr>
<tr>
<td>- JD.com</td>
<td>- HUATAI Securities</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- VIP.com</td>
<td>- Lufax</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- 360</td>
<td>- GUOTAI Securities</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Toutiao</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

500+ use cases in production global

Data collected from public information and kylin community.
Enterprise OLAP on Hadoop
Kyligence: Enterprise OLAP on Hadoop

Kyligence Solutions

Kyligence Analytics Platform (KAP)

KyAnalyzer
Agile BI

KyStudio
Model Designer

KyManager
Administrator Tool

KyStorage
Columnar Storage

Security
Cell Level ACL

Apache Kylin
Open Source
OLAP On Hadoop

On-Demand Deployment
On-Premises
Hybrid
In the Cloud
Kyiligence: Enterprise OLAP on Hadoop

Kyiligence Analytics Platform (KAP)

Query Pushdown: minutes latency

Cube Access: sub-second latency

Hive
Spark SQL
Impala

Data Exploration/Discovery

Intelligent Cubing by KAP

Mission Critical Analytics
Support
Data Exploration and Discovery
• Hive: 33 queries can’t support or run out of time
• KAP: all 99 queries supported
• Routine query between SQL on Hadoop and Apache Kylin
Speed Up
Mission Critical Analytics
Kyligence Studio: Data Modeling Designer

- Drag & Drop
- Smart Data Modeling
- Intelligent Optimization
Integrate with Business Intelligence tools
Seamless Integration with BI tools

- KyAnalyzer
- Tableau
- Power BI/Excel
- IBM Cognos
- MicroStrategy
- Superset
- Zeppenlin
- Saiku
- ...
Enhanced Security and Management

Cell Level ACL/SSO/LDAP/Kerberos…
Use Case: CPIC
CPIC: China Pacific Insurance (Group) Co., LTD

- Global Fortune 500 insurance company
- Top 2 insurance company in China
- $40+ billion revenue
- 8+ million customers
- 97,000+ employees
Challenges

• Legacy IBM Cognos + DB2 solution can’t support Big Data scenarios

• Long waiting time (minutes ~ hours for reporting)

• Low concurrency (100,000+ employees!)

• High cost
Journey of Kyligence Analytics Platform

1. 2016.12 ~ 2017.01 KAP POC: Performance Testing
   - Query Latency
   - Concurrency

2. 2017.01 ~ 2017.03 KAP POC: Compatibility
   - Cognos Connection
   - Cognos Syntax

3. 2017.03 ~ 2017.05 Development
   - Fixed Reports
   - Flexible Reports

4. 2017.05 ~ 2017.06 Go alive
   - All dataset aggregation and testing
   - Fixed Reports released

- No changes on Hadoop side
- No additional engineers required
- Most of work done by analysts
Benefits after Adopting Kyligence

• One-stop BI platform generates complicated reports
• Over 90% queries return within 3 seconds (including high-dimensional queries)
• Seamless integration with IBM Cognos, no change at front-end
• 2 KAP cubes replaced 2000+ IBM Cognos cubes
• Cost reduced significantly by adopting open source technology
"Kyligence enables us to find valuable insights faster from every insurance policy within seconds. Kyligence’s platform allows us to achieve more with less. Our lean management system has improved significantly."

-- Minchen Wu, Depute GM of IT, CPIC
China Construction Bank (CCB): 2nd Largest Bank in the World

**Fusion Big Data Platform**

“Apache Kylin is last piece of puzzle to serving data asserts management between legacy DW and new Big Data.”

-- Zhi Zhu, Vice Senior Manager of Tech Dept, CCB

- Open: Connect to Teradata/Greenplum and IBM Cognos/Saiku…
- Flexible: Self-Services for end users
- Efficiency: Speed up PC and Mobile analytics experience
Enterprise OLAP on Hadoop

Speed Up Mission Critical Analytics

Booth #855
luke@kyligence.io
http://kyligence.io