Kyligence

Simplify Data Analytics over the Cloud

Luke Han | luke.han@kyligence.io
- Co-founder & CEO of Kyligence Inc.

About Luke Han

- Luke Han
  - Co-founder & CEO at Kyligence
  - Co-creator and former PMC Chair of Apache Kylin
  - Apache Software Foundation (ASF) Member
  - Microsoft Regional Director
  - Former eBay Big Data Product Manager Lead

Kyligence = Kylin + Intelligence

AI-Augmented Data Warehouse

- Founded in 2016 by the original creators of Apache Kylin
- CRN - Top 10 Big Data Startups 2018 [US]
- Deloitte - China Rising Star 2019 [China]
- Dual HQ in Shanghai and San Jose

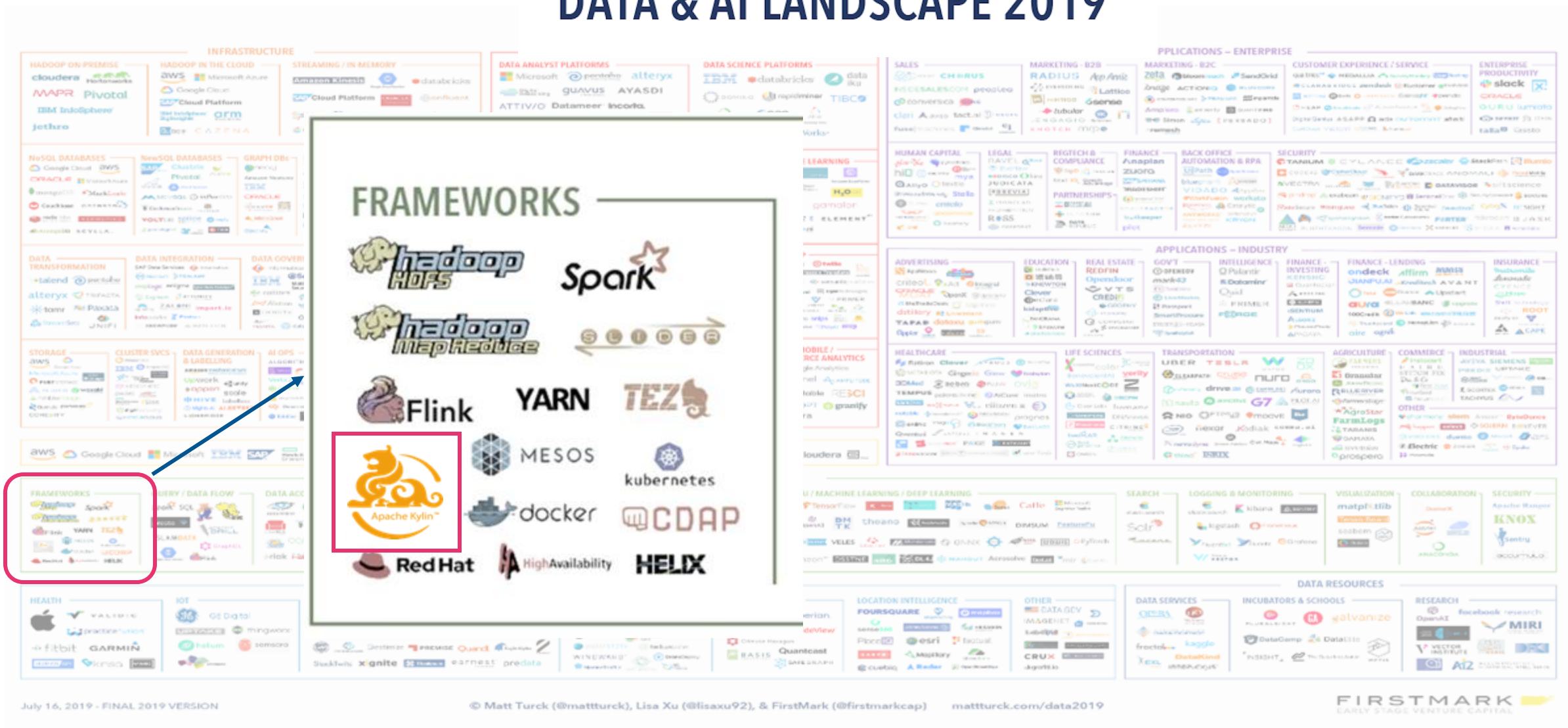
Founded Pre-A	Series A	Series B	Series C
2016	2017	2018	2019
RedPoint Cisco	CBC SHUNWEI	8Roads (Fidelity International)	Coatue

Kyligence is the only one who runs the same model as Databricks, Elastic in China based on an Apache Top Level Project

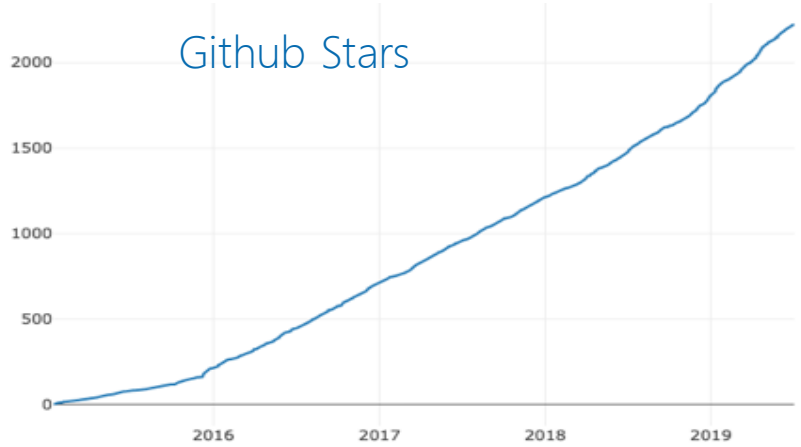
elasticsearch

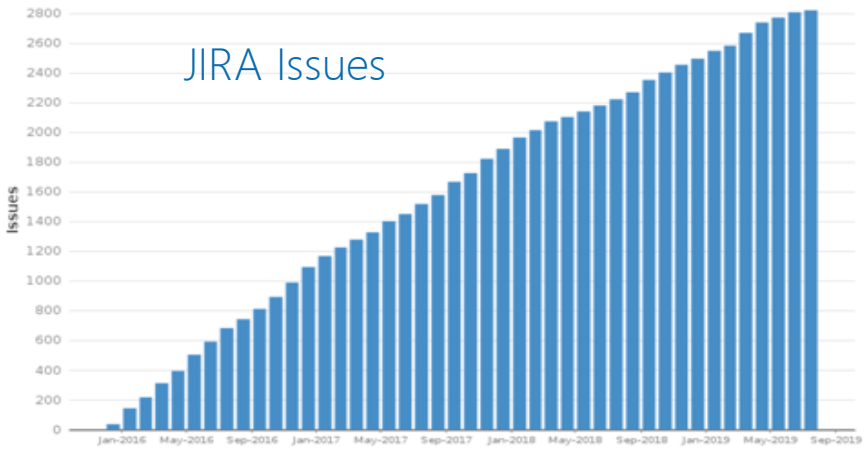
Trusted by Global Future 500

Apache Kylin – Leading Open Source OLAP for Big Data



Apache Kylin Community & Adoptions





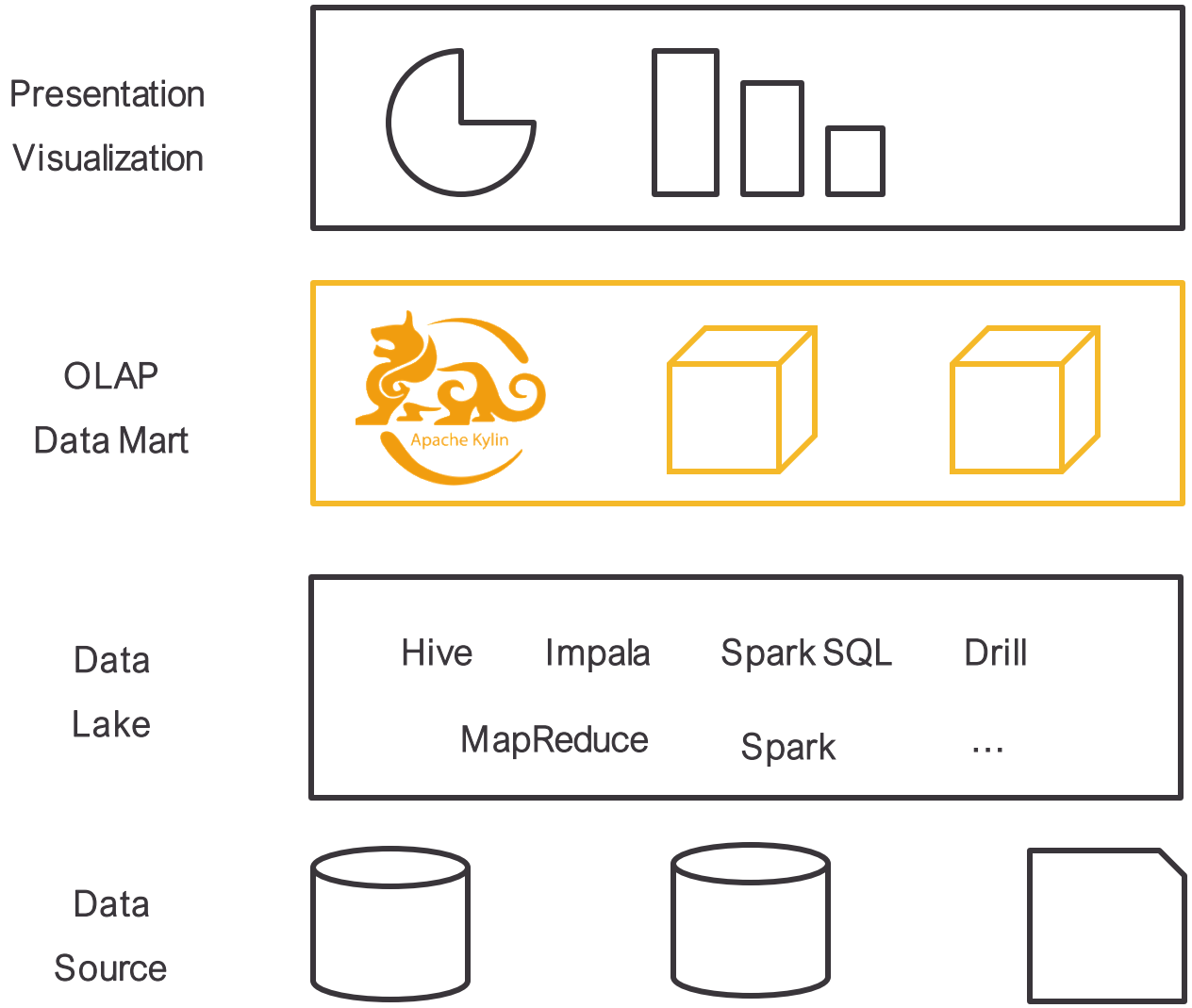
Apache Kylin™

1000+ Global Adoptions

Leading Open Source OLAP

ebay
iCloud
amazon
Microsoft
YAHOO! JAPAN
Walmart
cisco
DECATHLON
蚂蚁金服 ANT FINANCIAL
Baidu 百度
Ctrip
滴滴
中信银行 CHINA CITIC BANK
中国农业银行 AGRICULTURAL BANK OF CHINA
中国移动 China Mobile
中国电信 CHINA TELECOM
美团
平安科技 PINGAN TECHNOLOGY
趣
OLX
iQIYI 爱奇艺

Apache Kylin –managing your Golden Data over Data Lake



- Semantic Layer
- Speed up Analytics using Pre-Calculation
- ANSI SQL Interface
- High Concurrency and High Performance
- Batch & Streaming together

Use Case: IBM Cognos Replacement

2 Kyligence Cubes vs 1200 IBM Cognos Cubes

Challenge

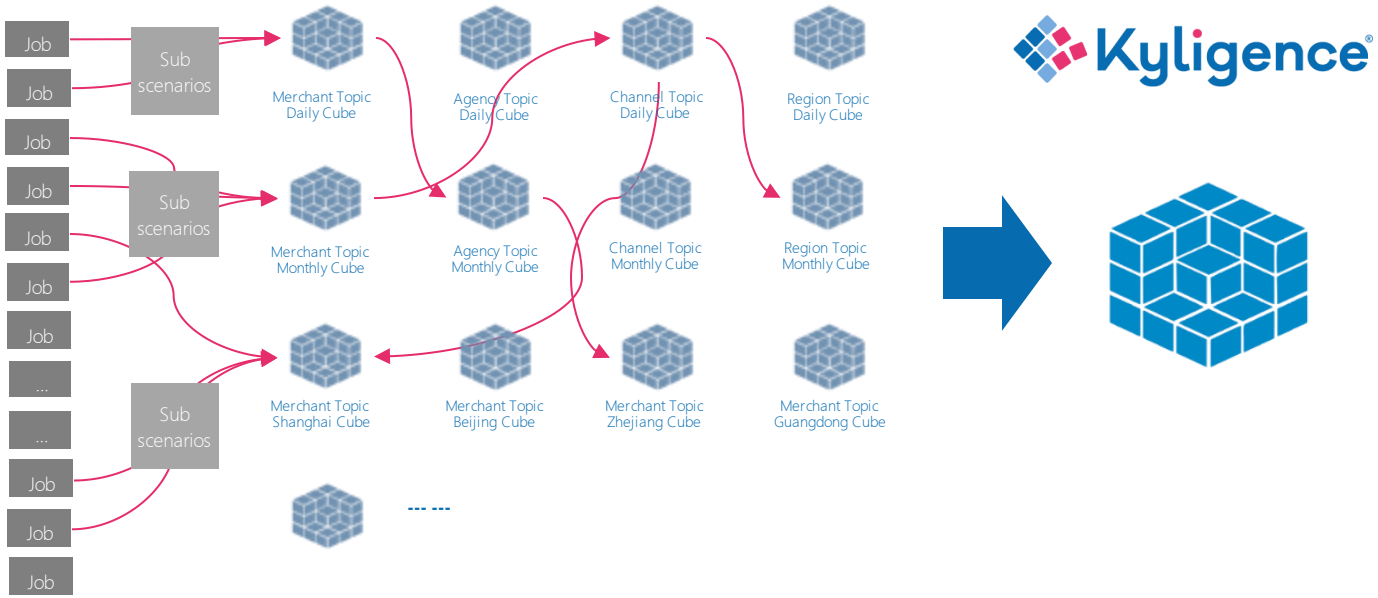
- 1200+ existing cubes to manage
- 1000+ jobs to maintain
- Time-to-insight over 4 days

Solution

- Using Kyligence replaced IBM Cognos backend but continue keep Cognos Reporting to interactive with Kyligence

Result

- 1000x improved maintenance efficiency
- 10x faster and more stable analytics performance
- Time-to-Insight less then 4 hours

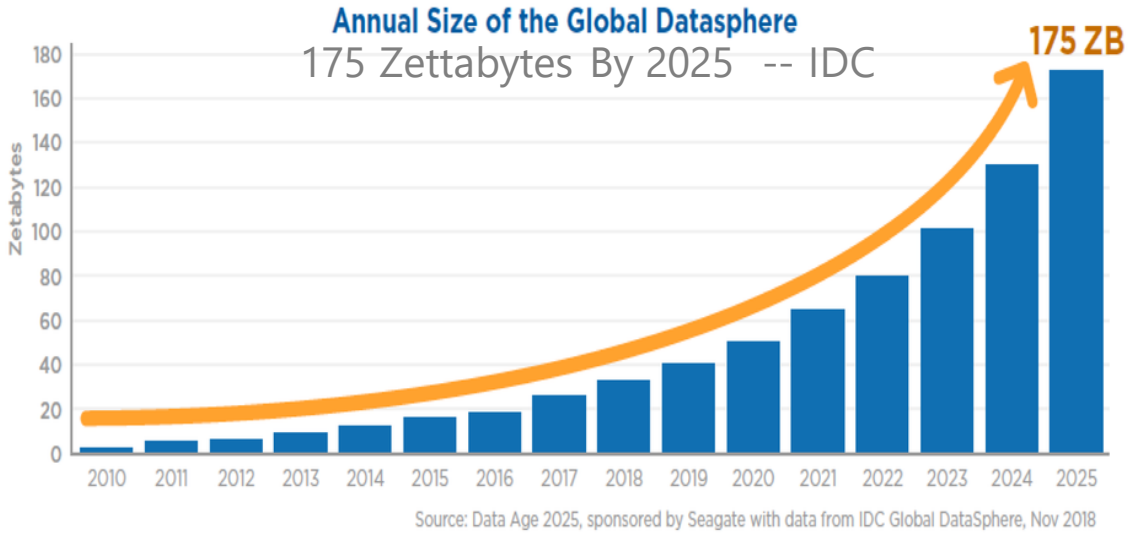


1200+ IBM Cognos Cubes, 1000+ ETL Jobs with dependencies

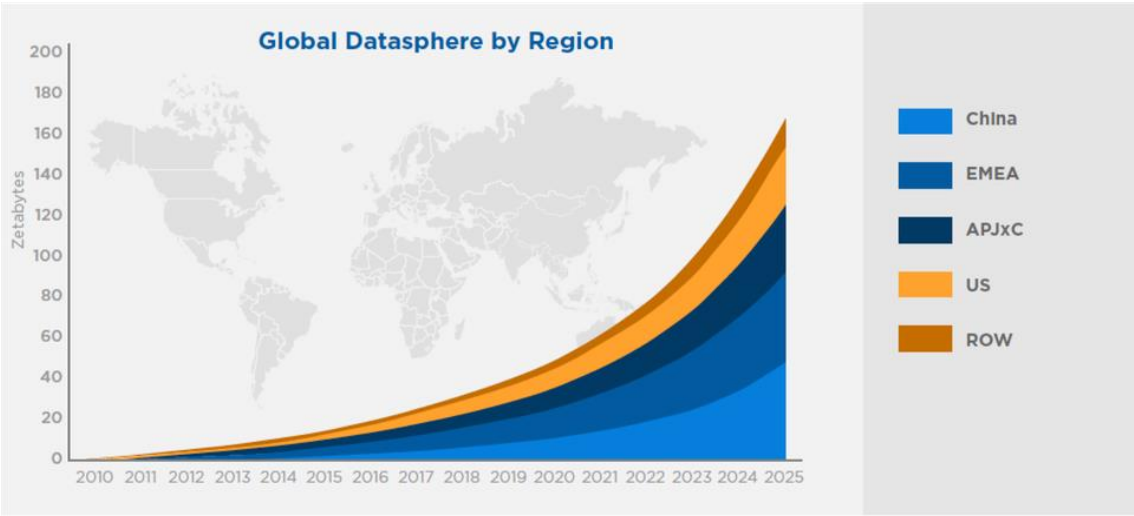
2 Cubes, 1 ETL Job

Data is next Oil

The world's most valuable resource is no longer oil, but data. —"The Economist"



China's Datasphere is expected to grow 30% on average over the next 7 years and will be the largest Datasphere of all regions by 2025 --IDC



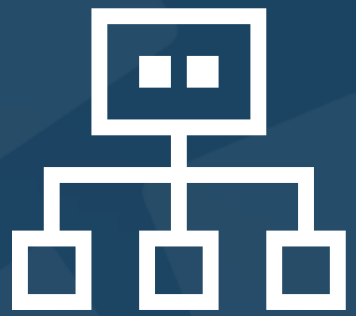

and, Data is moving to Cloud

But, the Chaos Happens Again!!!

snowflake amazon web services Azure amazon REDSHIFT kafka Spark ORACLE elasticsearch MySQL

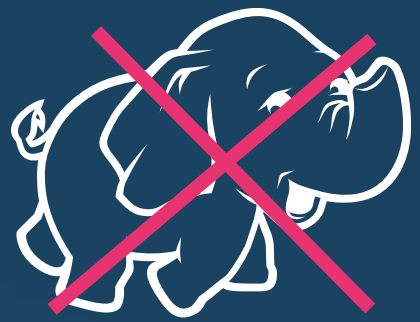
Kyligence

What's missing there?

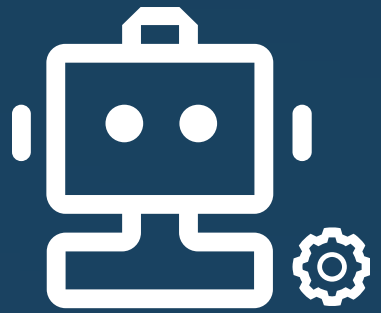


Intelligent
Semantic Layer

High Concurrency ,
High Performance



No More
Hadoop on Cloud



Automation
and Automation

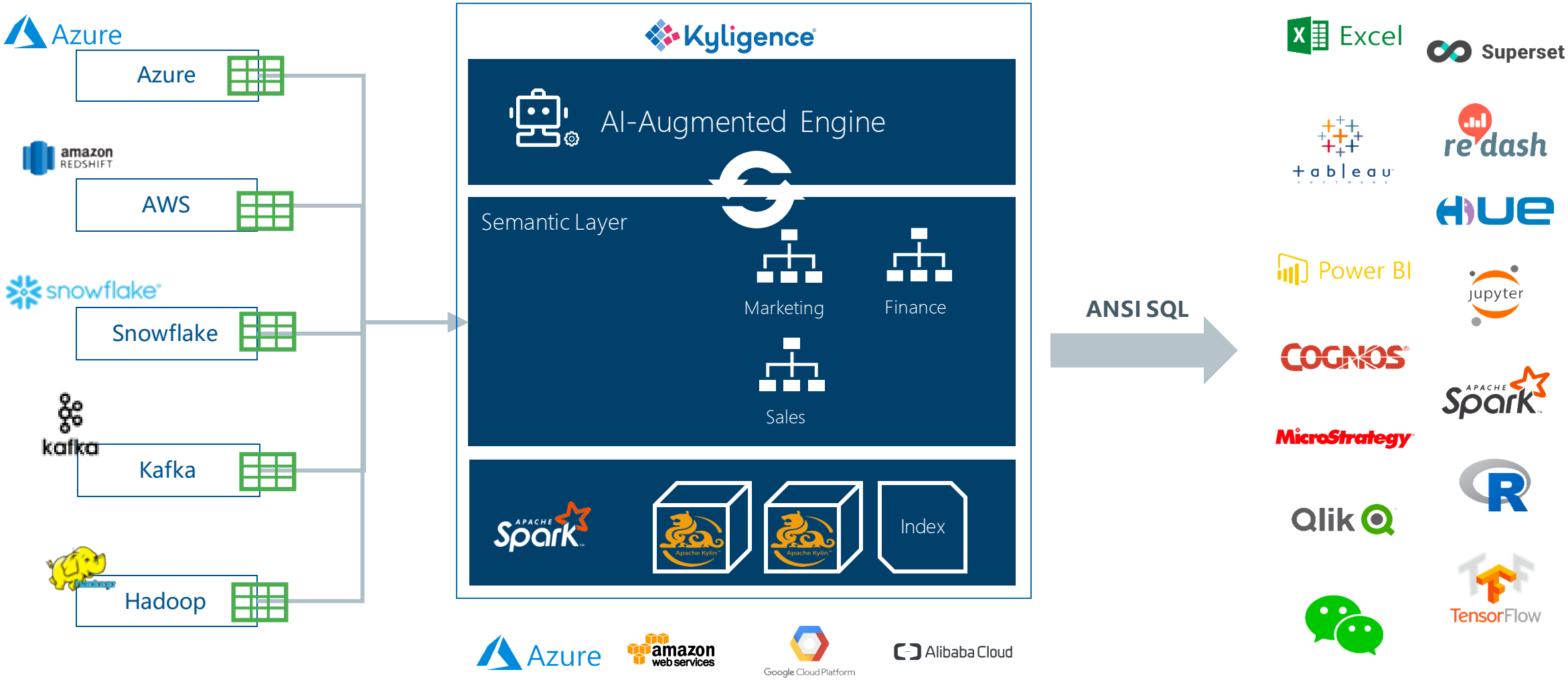
What's the must to have of Big Data over the Cloud

Cloud Native

Elastic Sizing

Lower TCO

Kyligence Cloud – Simplify Data Analytics over the Cloud



Spark Native – Reduced Hadoop Overhead in the Cloud

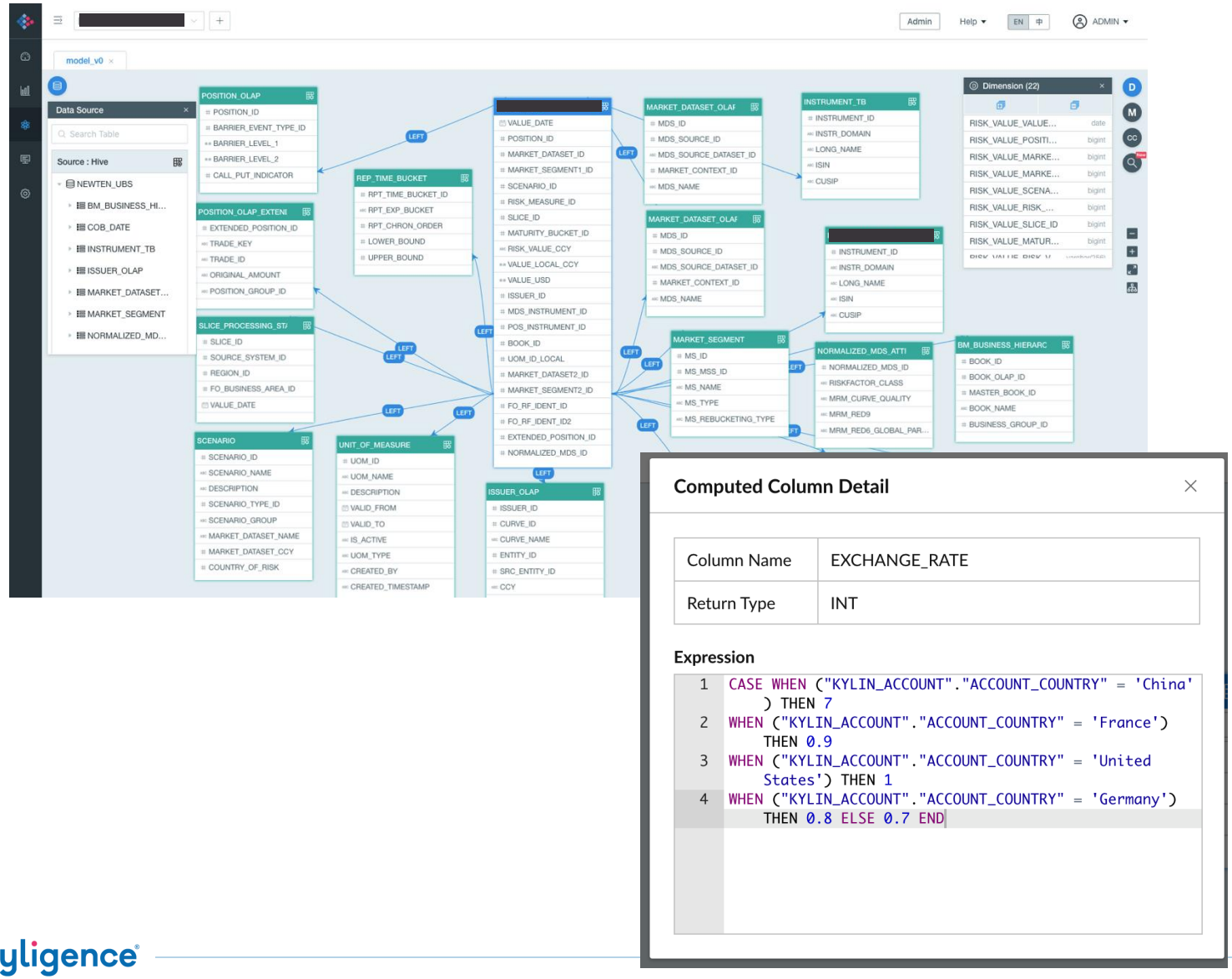
1. Read
...from S3/ADLS/Snowflake...

2. Build
...Build Cube & Index

3. Store
...Persistent #2 Data into Cloud Storage: S3/ADLS

4. Serve
...Execute SQL parallelly with Cloud Storage

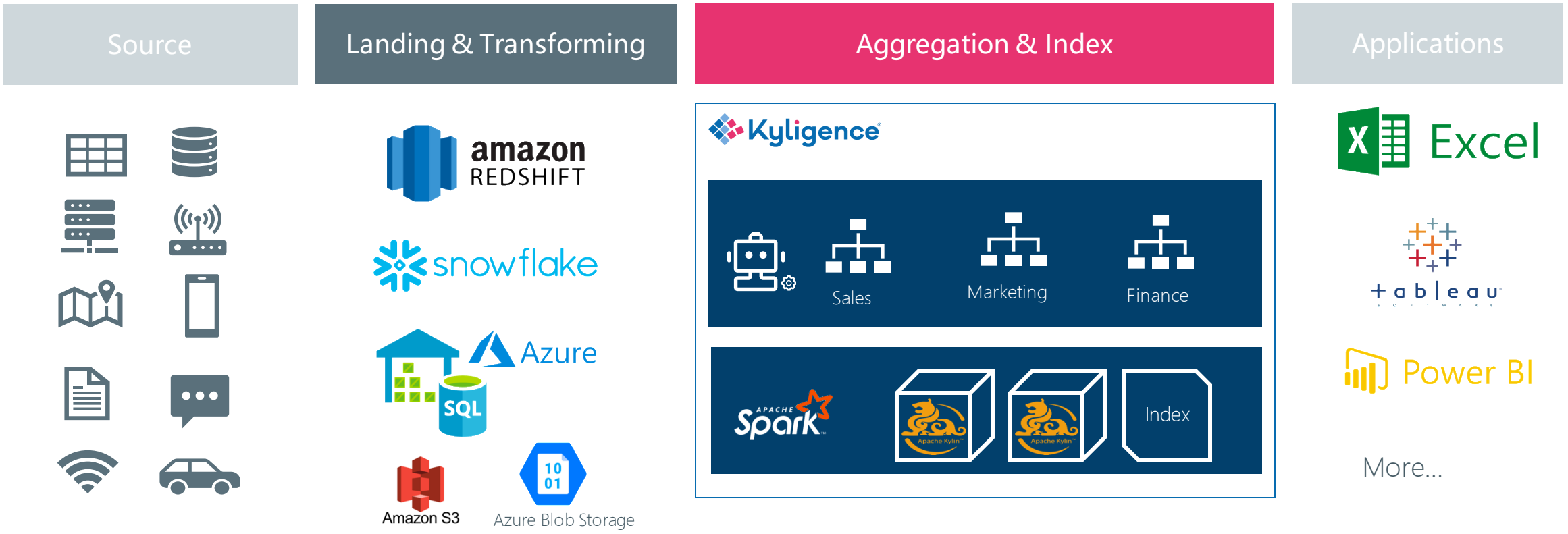
Semantic Layer – The Key to Govern Data



- The Single Source of Truth
- Business KPIs, Metrics, Hierarchies, etc.
- Synchronize across BI tools
- Flexible business calculations without coding:

---- YTD/MTD/Many-to-Many

Semantic Layer & OLAP for Cloud DW



Demo – Semantic Layer for Snowflake

...removed to reduce file size

Your car is self-driving…and it is produced automatically…

How about your Data?

What will happen when you arrived office with your auto pilot Tesla?

The Reality!

Rows of women office workers entering data using keypads…

- Data Scientists vs SQL Monkey
- Automation is the Key
- Bias of taste – which is better: Hadoop MPP, Cloud…

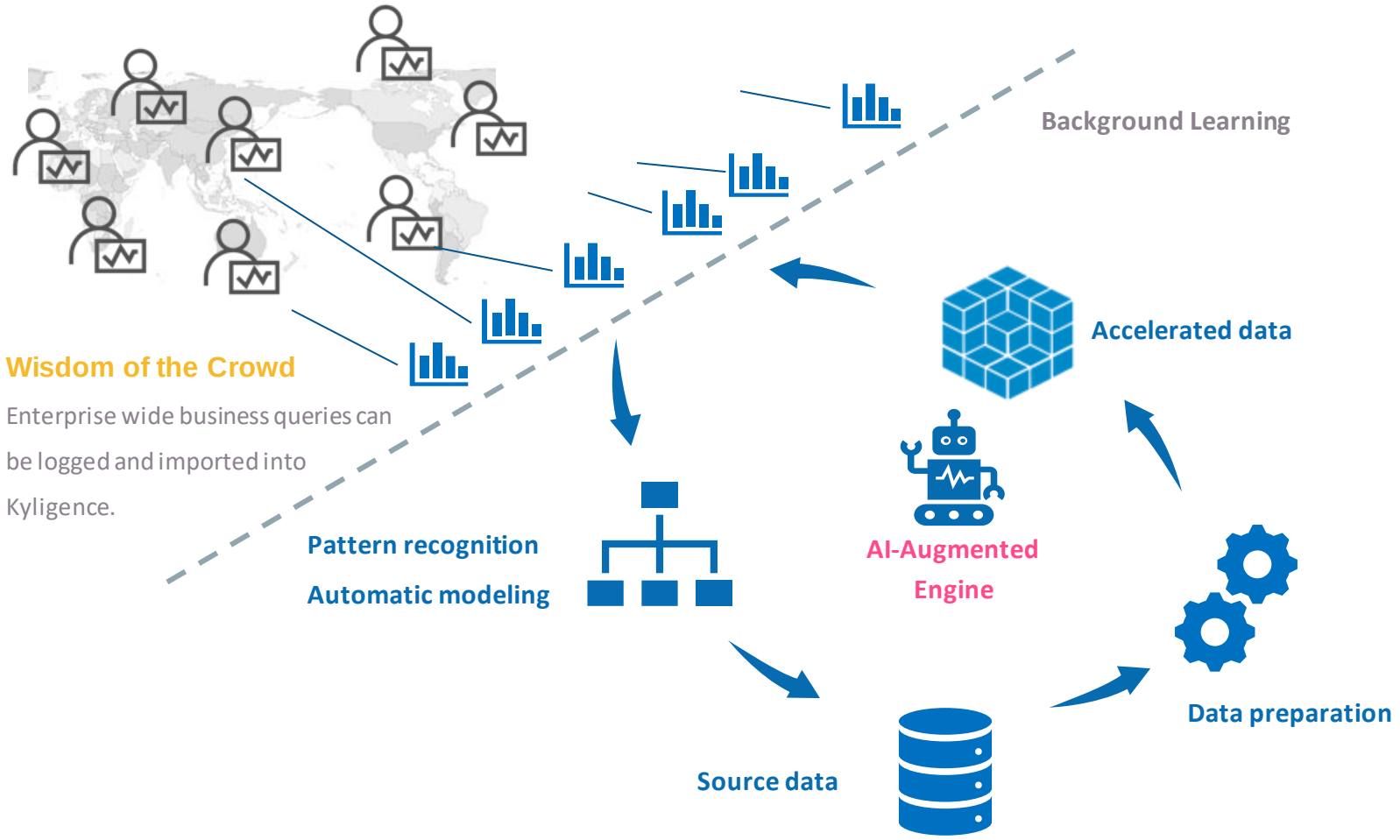
Automation is the Key

Augmented Analytics Is the Future of Data and Analytics

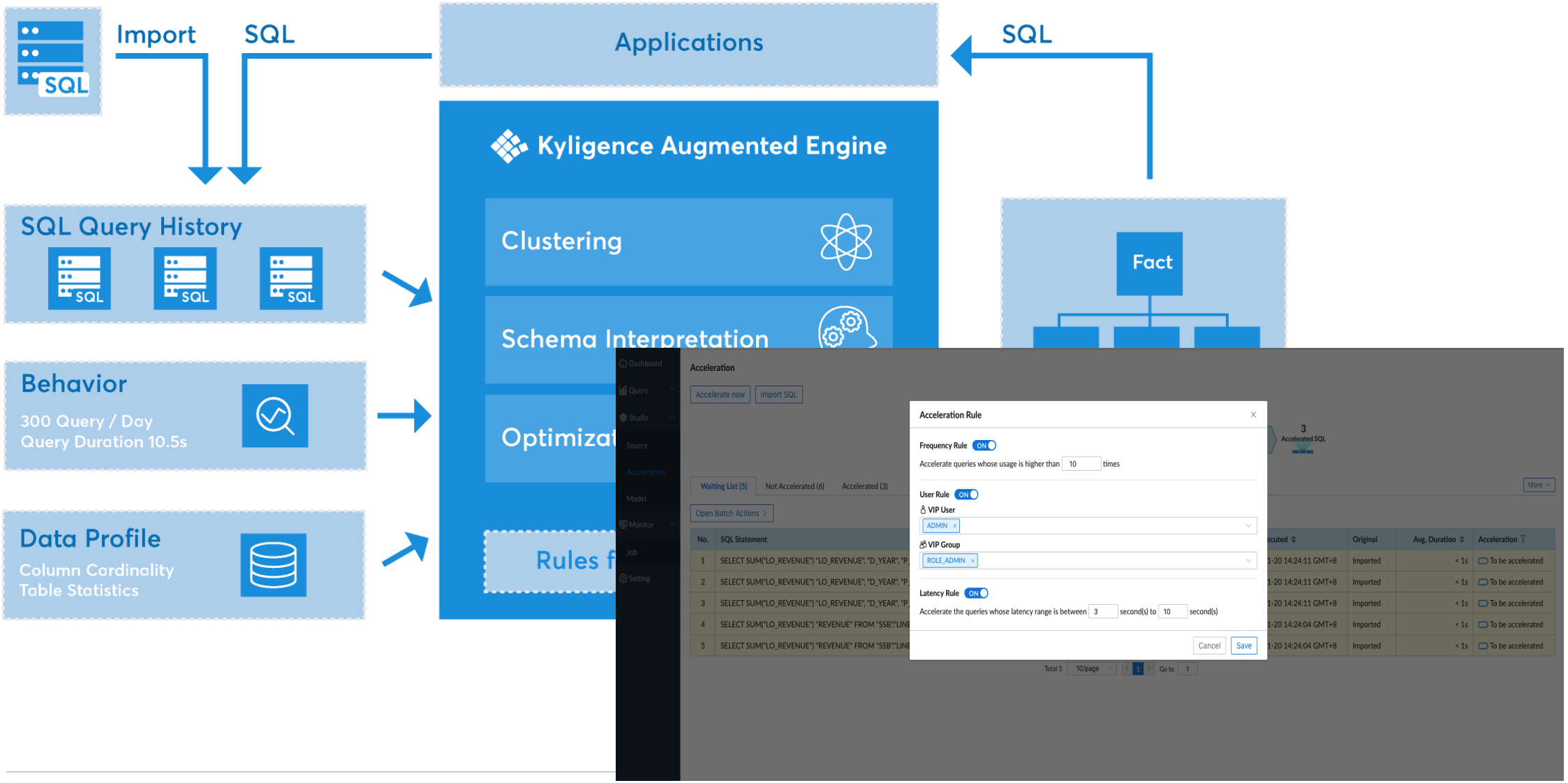
Published: 27 July 2017 | Analysts: Carlie Idoine, Cindi Howson, Rita Sallam

"...an approach that automates insights using machine learning and natural-language generation, marks the next wave of disruption in the data and analytics market. Data and analytics leaders should plan to adopt augmented analytics as platform capabilities mature."

AI-Augmented – Automation is the Future



AI-Augmented Engine – Learn from your analytics history

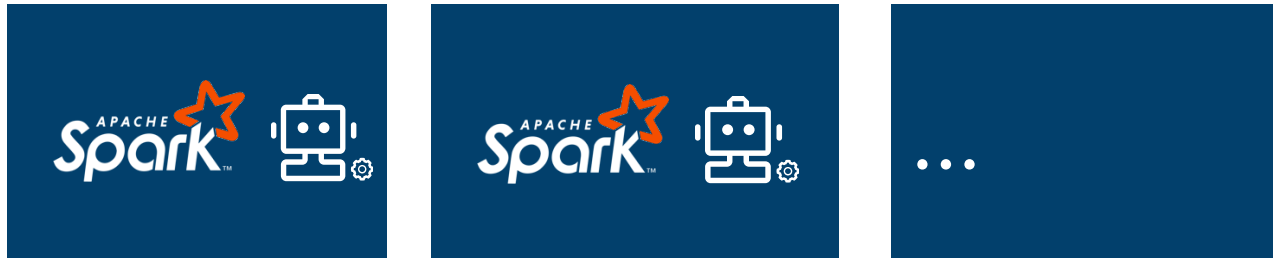


AI-Augmented Engine - Demo

...removed to reduce file size

High Concurrency – High Performance at Scale

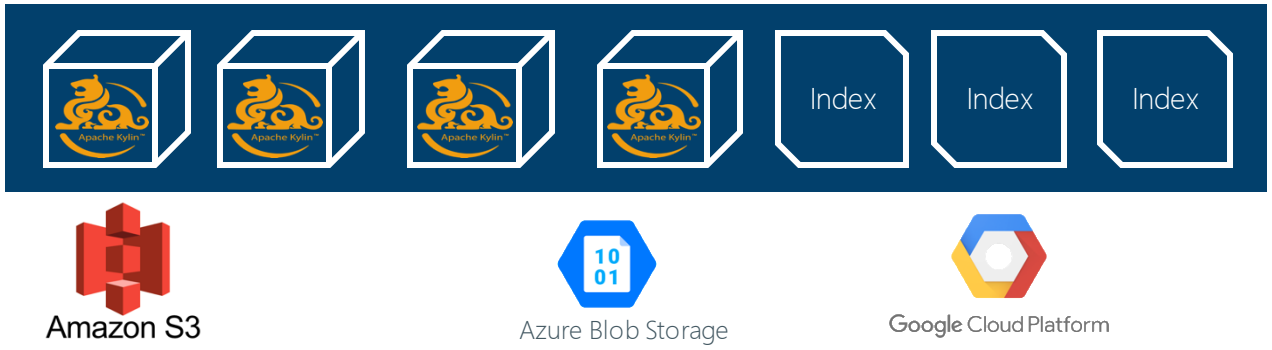
Kyligence Computing Node at Scale



Scale Out Kyligence Nodes for High Concurrency

- Provision in the fly
- Only query node
- One node support 50+ concurrency

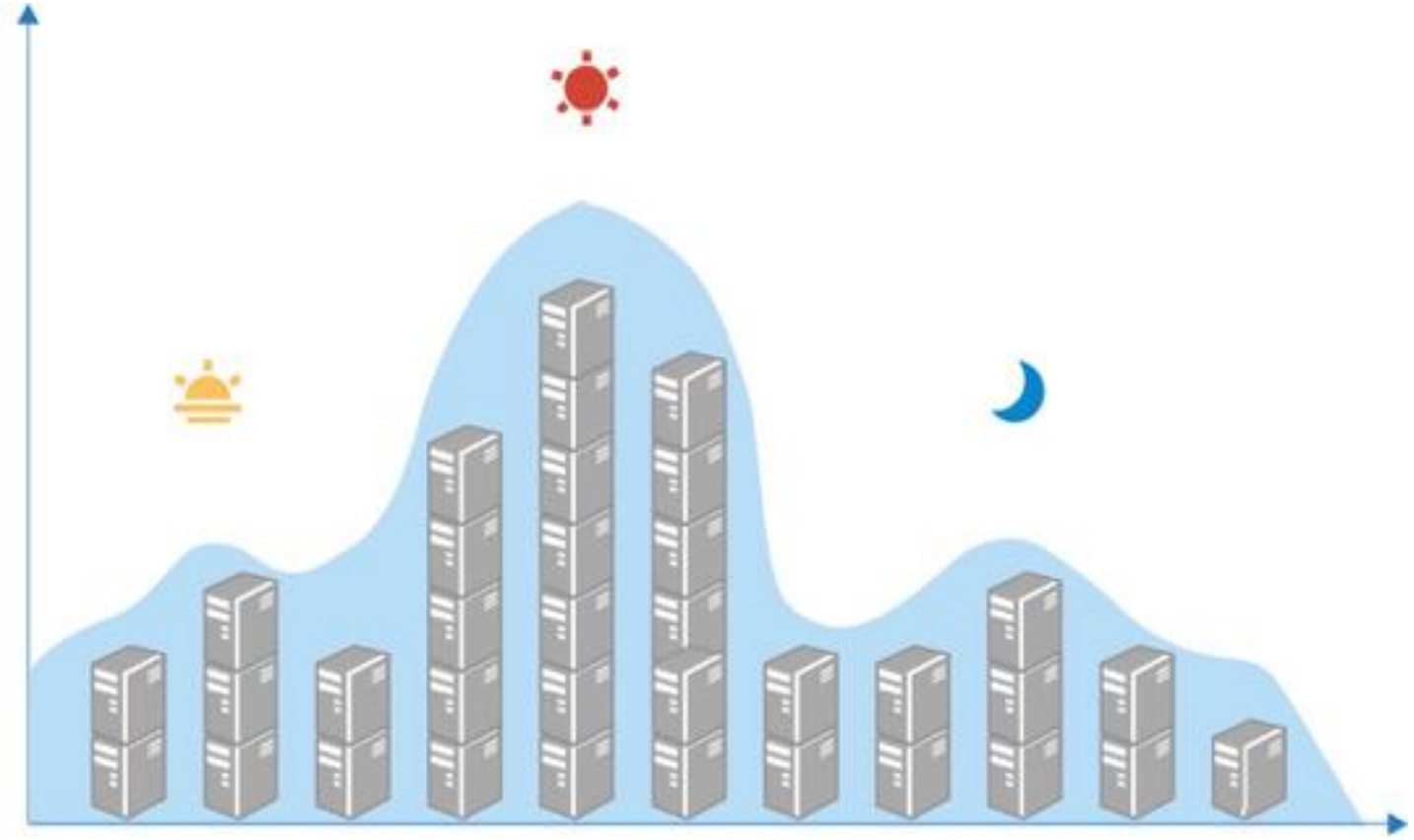
Kyligence Data stored in Cloud Storage



Scale to store PB size Data

- Read only
- Lower cost

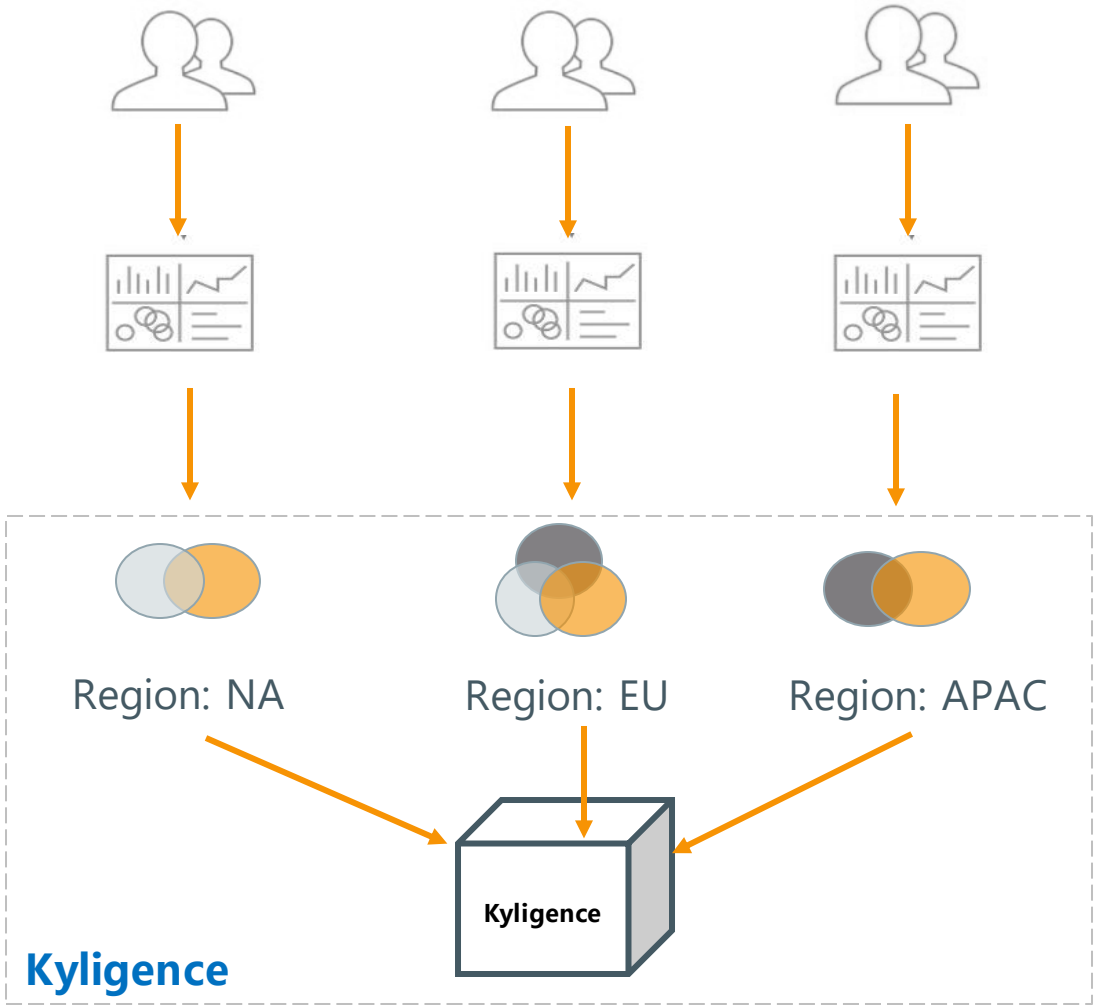
Elastic Scaling – Handle Peak Time Automatically



- Less computation and storage resource utilized
- Dynamic on-demand cluster resizing
- Use spot instances
- Efficiently planning for data growing

Security – Support Cell Level ACL

- Column/Row/Cell Level ACL
- LDAP/SSO
- Kerberos/HTTPS
- Source table ACL sync
- SDK



Interactive with billions of rows in Excel

Interactive with Excel on billions of rows

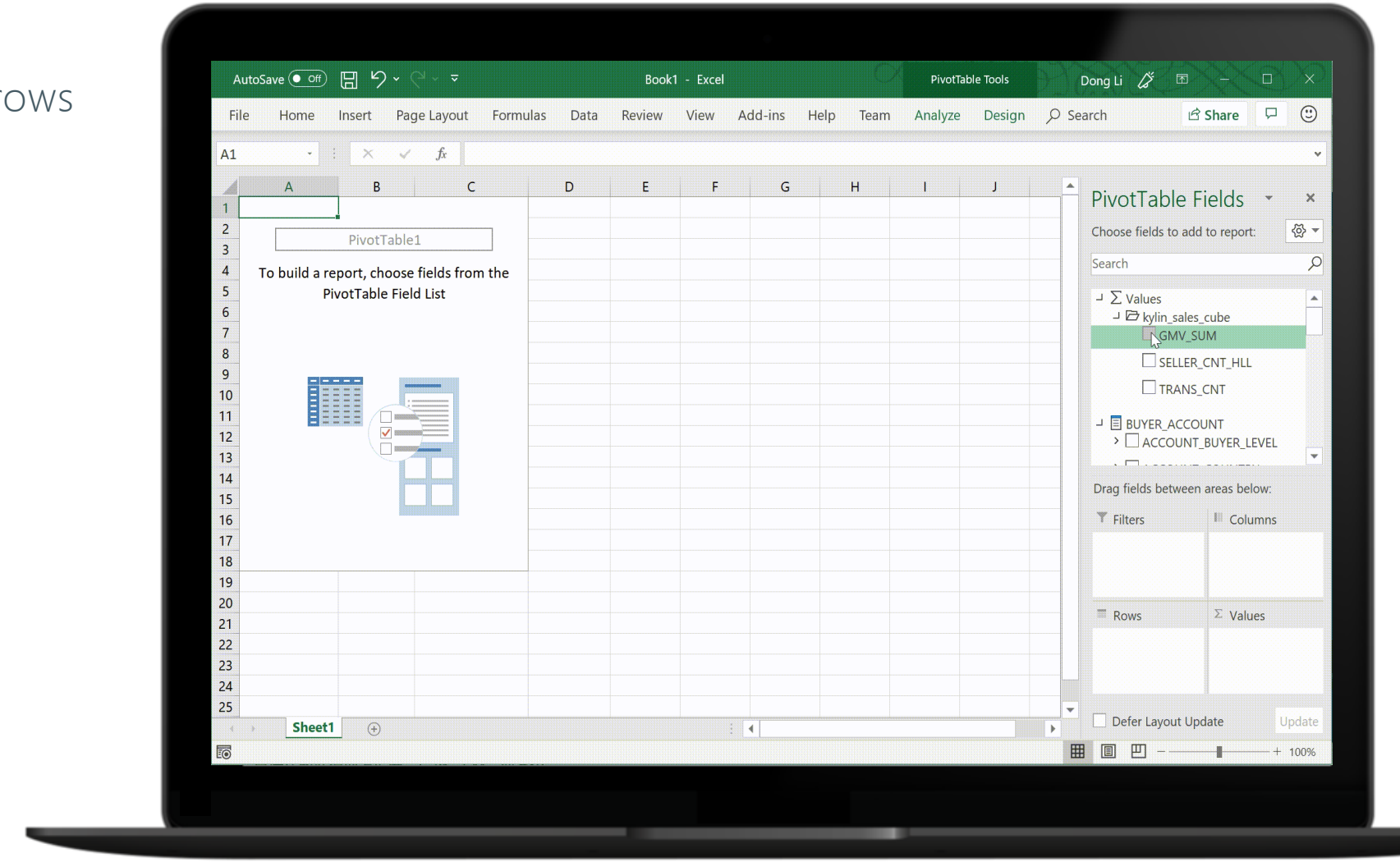
- Perfect work with the Kyligence semantic layer
- Seamlessly integration with analysts favorite BI tools

Power BI

Excel

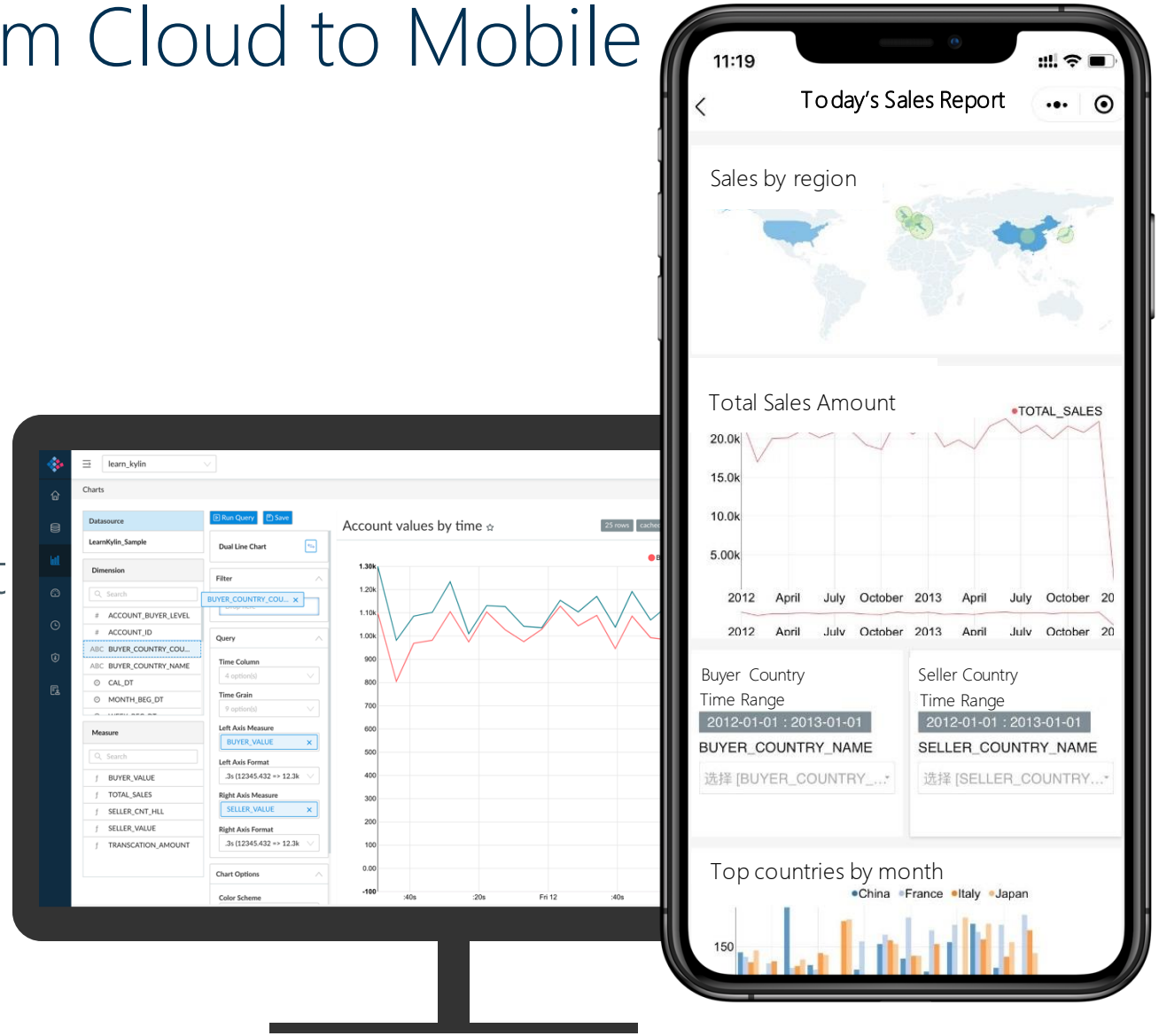
+ableau

Qlik



WeChat Application – from Cloud to Mobile

- Collaboration via WeChat
- NO Coding
- No iOS/Android Development
- Anytime, anywhere
- Auto Responsive
- *beta version released*



Use Case: MS SSAS Migration to Azure

Challenge

- 300 billion rows of data
- 8 billion rows loaded/day
- Service unavailable usually
- Lack of data access control

One of the World's Largest Cubes

Result

- 1200+ dimensions in 1 cube
- Support Project/Column/Cell level data access control

Lower TCO

- $300/day saved for cloud infrastructure cost
- Less than 1 hour to build cube

Take away

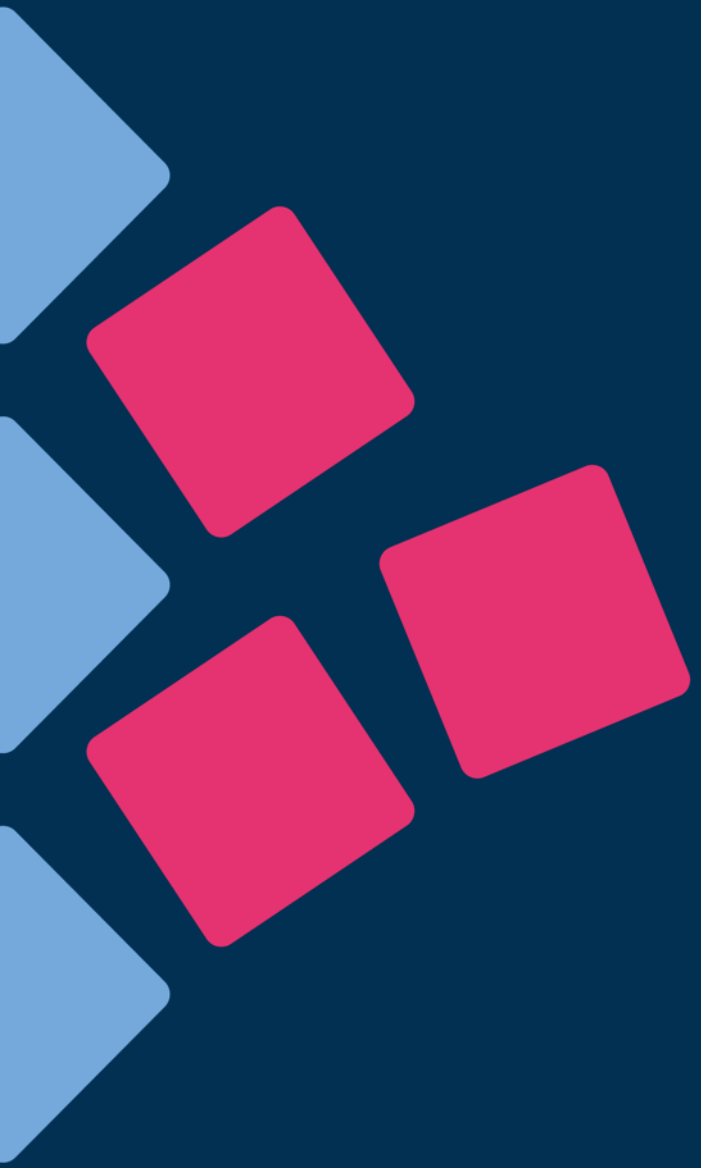
Try here:

https://cloud.kyligence.io

Kyligence Cloud Simplified Data Analytics in the Cloud

- Semitic Layer – the key for data governance and management
- AI-Augmented Engine – automation is the future of data analytics
- Cloud Native - Cloud is the destination of data and analytics
- Spark Native – Reduced Hadoop Overhead in the Cloud
- High Concurrency but Lower TCO

Kyligence



Unleash Big Data Productivity

Luke Han | luke.han@Kyligence.io