

A large, abstract graphic consisting of multiple overlapping, wavy lines in shades of blue and cyan, creating a sense of motion and data flow across the upper half of the slide.

Kyligence Introduction

MicroStrategy Partnership
Saswata Sengupta

◆ Apache Kylin



Top Level Apache Project

- The only open-source OLAP on big data platform



Sub-Second Interactive Query

- Large scale, high concurrency, sub second query latency, multi-dimension



Best Open-source Big Data Tool

- InfoWorld's Bossies (Best of Open Source Software Awards) in 2015 & 2016



1000+ Organizations

- Adopted by thousands of organizations globally

◆ Kyligence = Kylin + Intelligence

Accelerate Critical Business Decisions with AI-Augmented Data Management and Analytics

- Founded in 2016 by the creators of Apache Kylin
- Built around Kylin, with augmented AI and enhanced to deliver unprecedented enterprise analytic performance
- CRN Top-10 big data startups in 2018
- Global Presence: San Jose, Seattle, New York, Shanghai, Beijing
- VCs: Fidelity International, Shunwei Capital, Broadband Capital, Redpoint, Cisco, Coatue

Founded Pre-A



Redpoint
Cisco

Series A



CBC
SHUNWEI

Series B

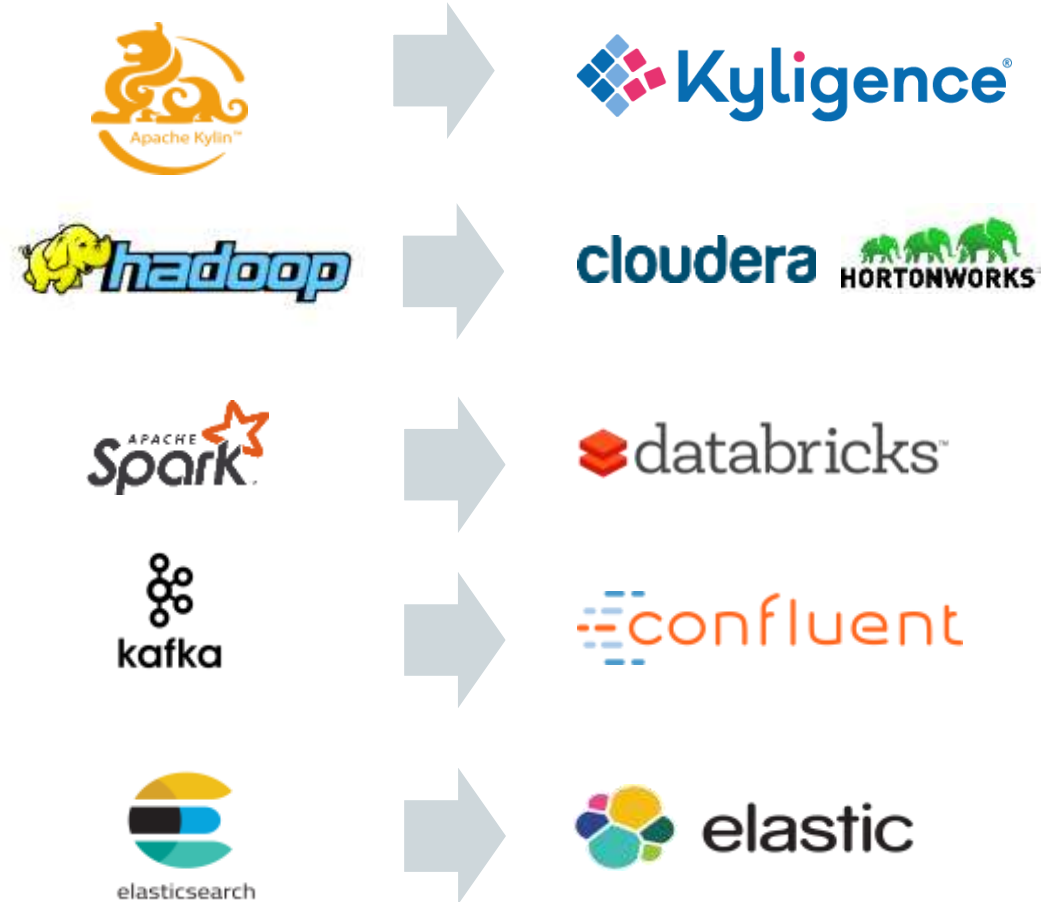


8Roads

Series C



Coatue



◆ Trusted by Global Fortune 500

BFSI



J.P.Morgan

WeBank



招商银行
CHINA MERCHANTS BANK



太平洋保险
CPIC



国泰君安证券
GUOTAI JUNAN SECURITIES



财通证券
CAITONG SECURITIES
更懂你的综合金融服务商

Telecom



China
unicom中国联通



中国移动
China Mobile



中国电信
CHINA TELECOM



telecoming

Technology

SAMSUNG

Xactly
Inspire Performance

Lenovo

oppo

vivo

Manufacturing,
Retail, Etc.



L'ORÉAL
PARIS



Shinhotokai
欣和



上海通用汽车
SHANGHAI GM



EXPRESS

◆ Pains in Collaboration

Low efficiency in development to fulfill business requirements



Data Engineer

- Manage data source
- Design data model to keep one source of truth
- ETL and load data

Limited dimension and measures in a model to serve complex calculations

Time to insight is slow

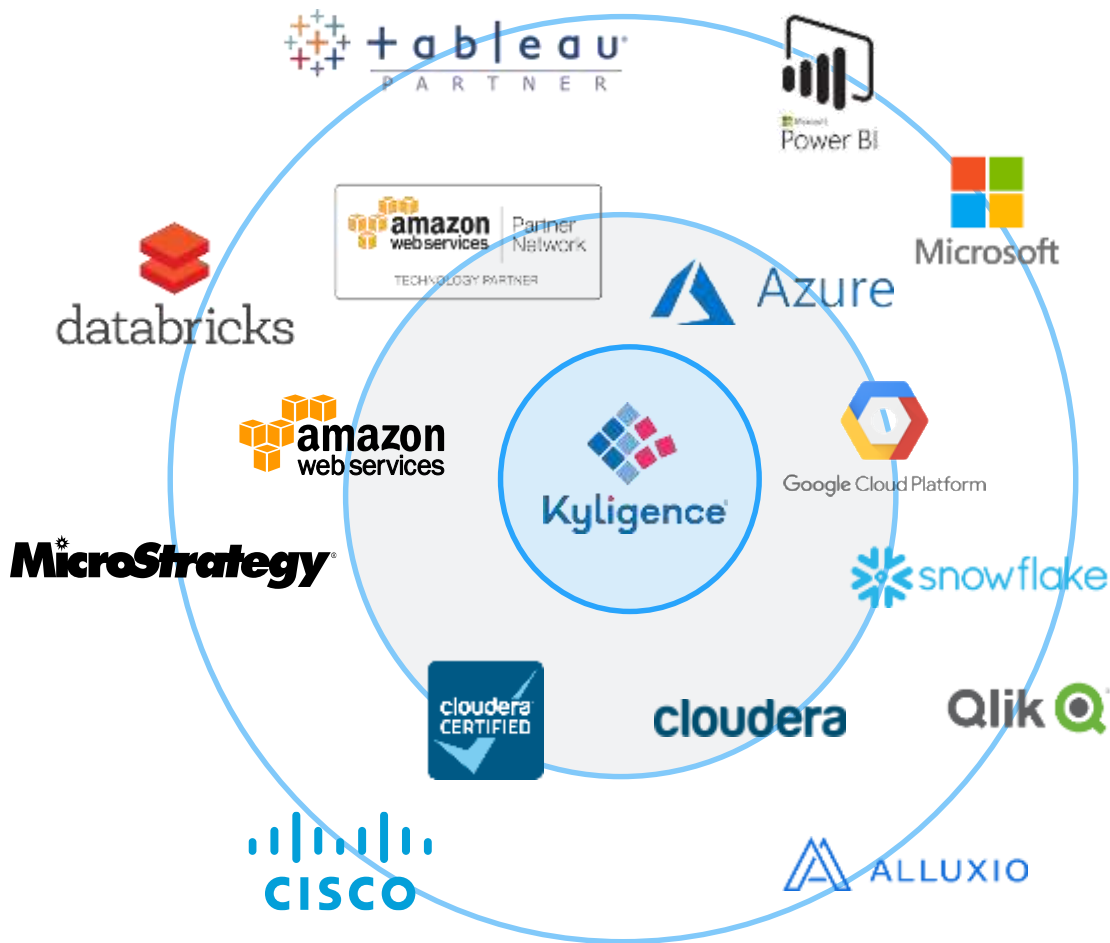
Difficulty if analytics requirements or source data change



Data Analyst

- Develop dashboard/reporting
- Self-service analysis to answer business questions

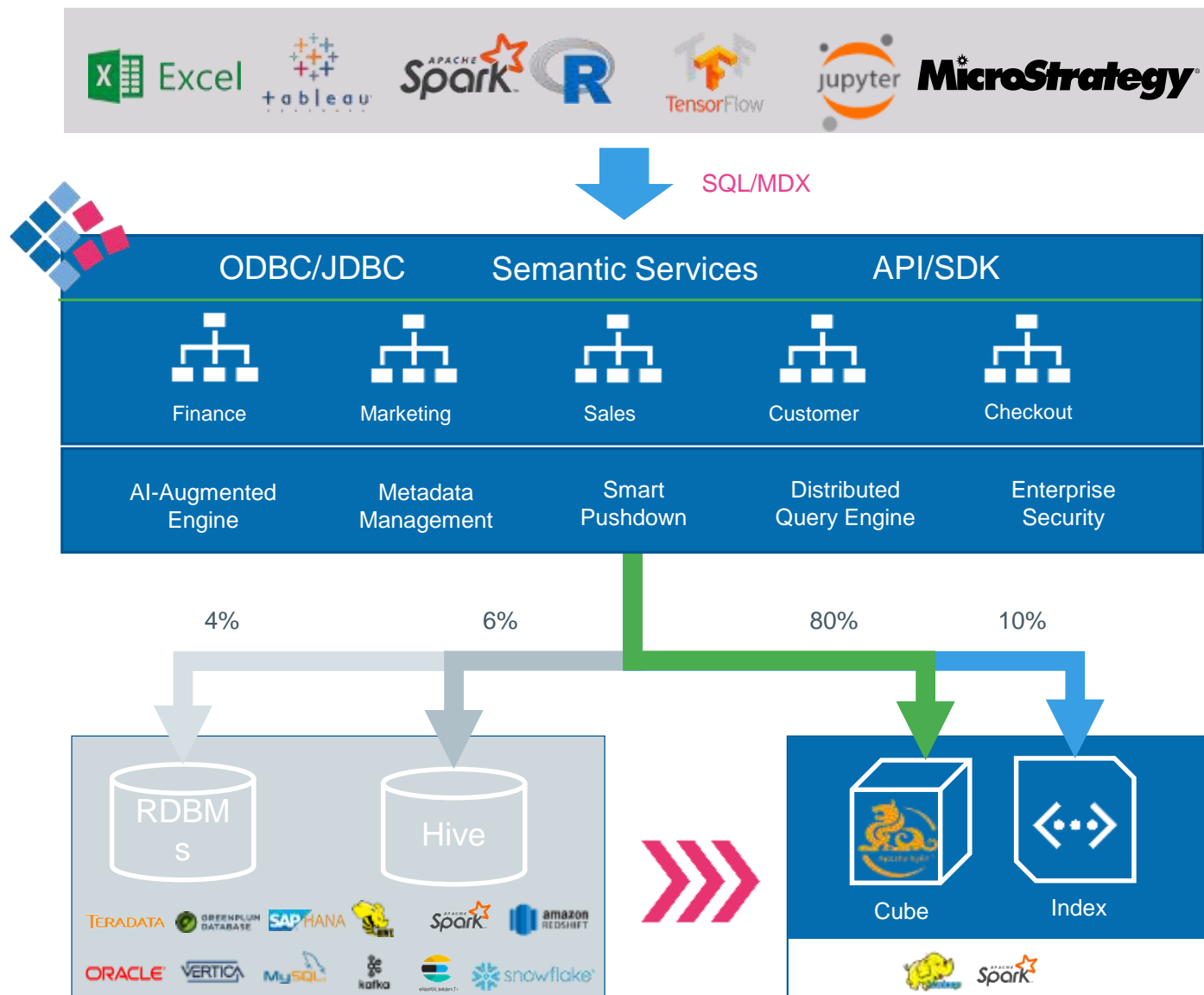
◆ Kyligence Ecosystem



Global Partners

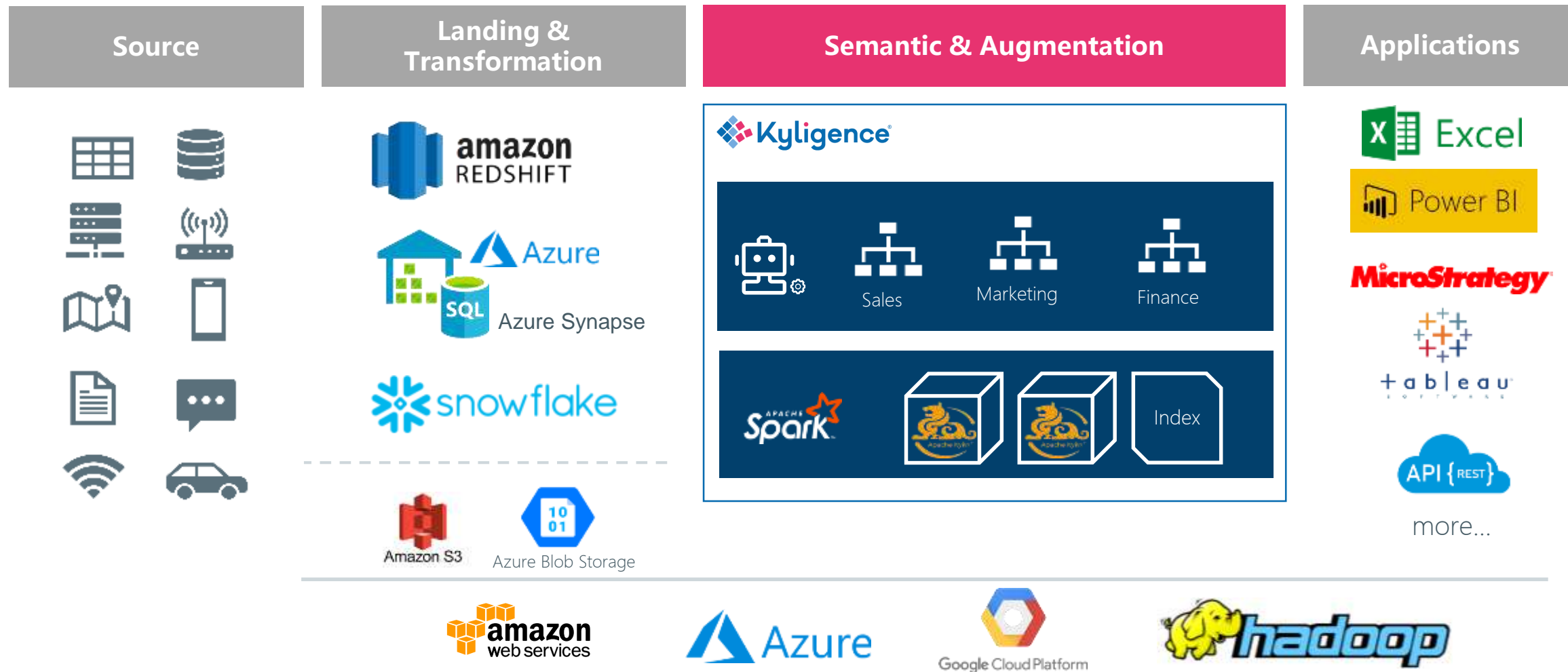
- Fully enabled on leading cloud and data platforms (Azure, AWS, Google Cloud, Cloudera)
- Integrated with popular BI and virtualization (Tableau, Power BI, Qlik, MicroStrategy)
- Certified on main Hadoop distributions (CDP)

◆ Kyligence Enterprise Accelerate Mission-critical Analytics Intelligently

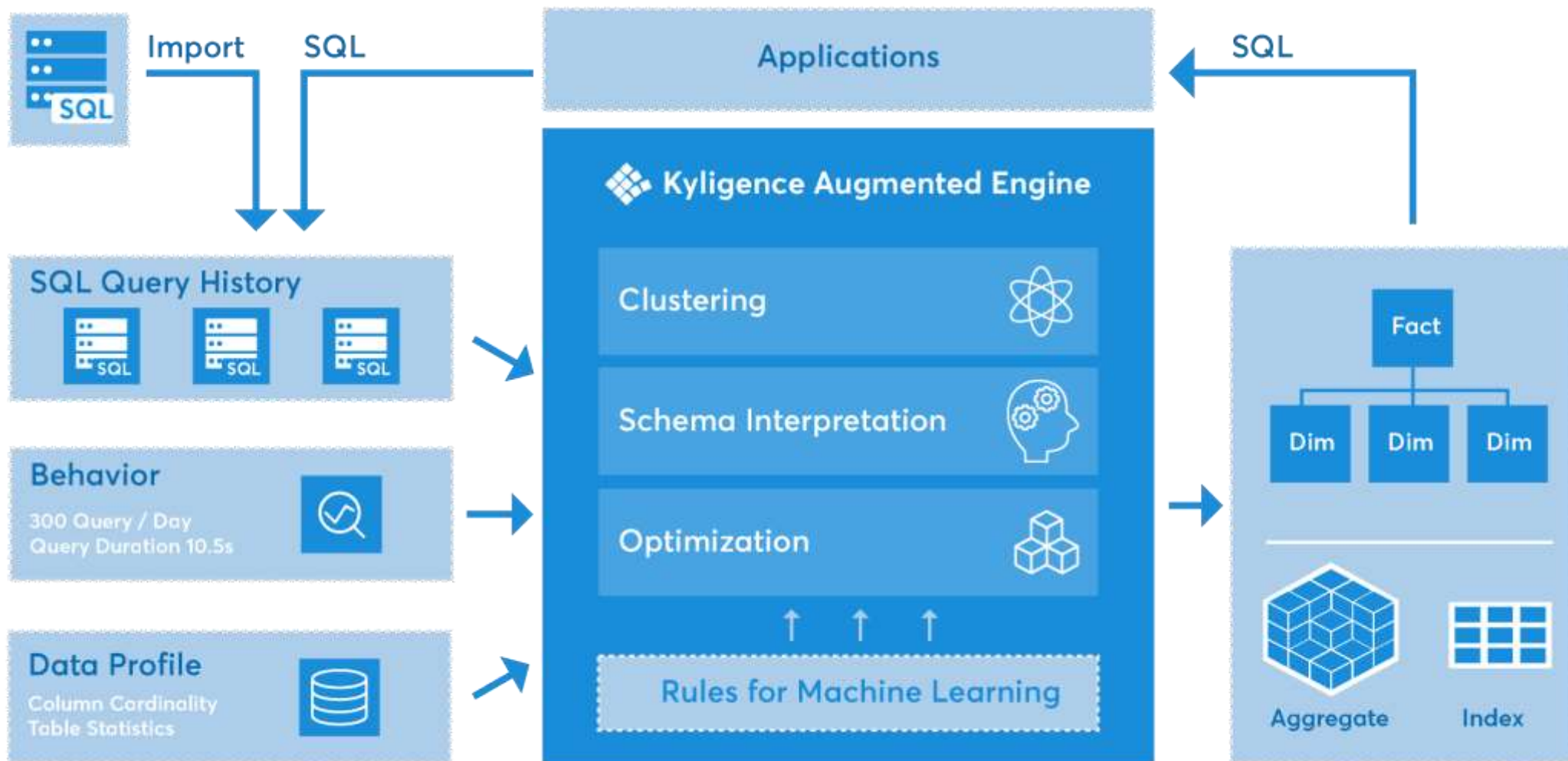


- Unified Query Entrance
- Business Semantic Layer
- Query Pattern for all data
- High Performance Engine

◆ Kyligence Cloud

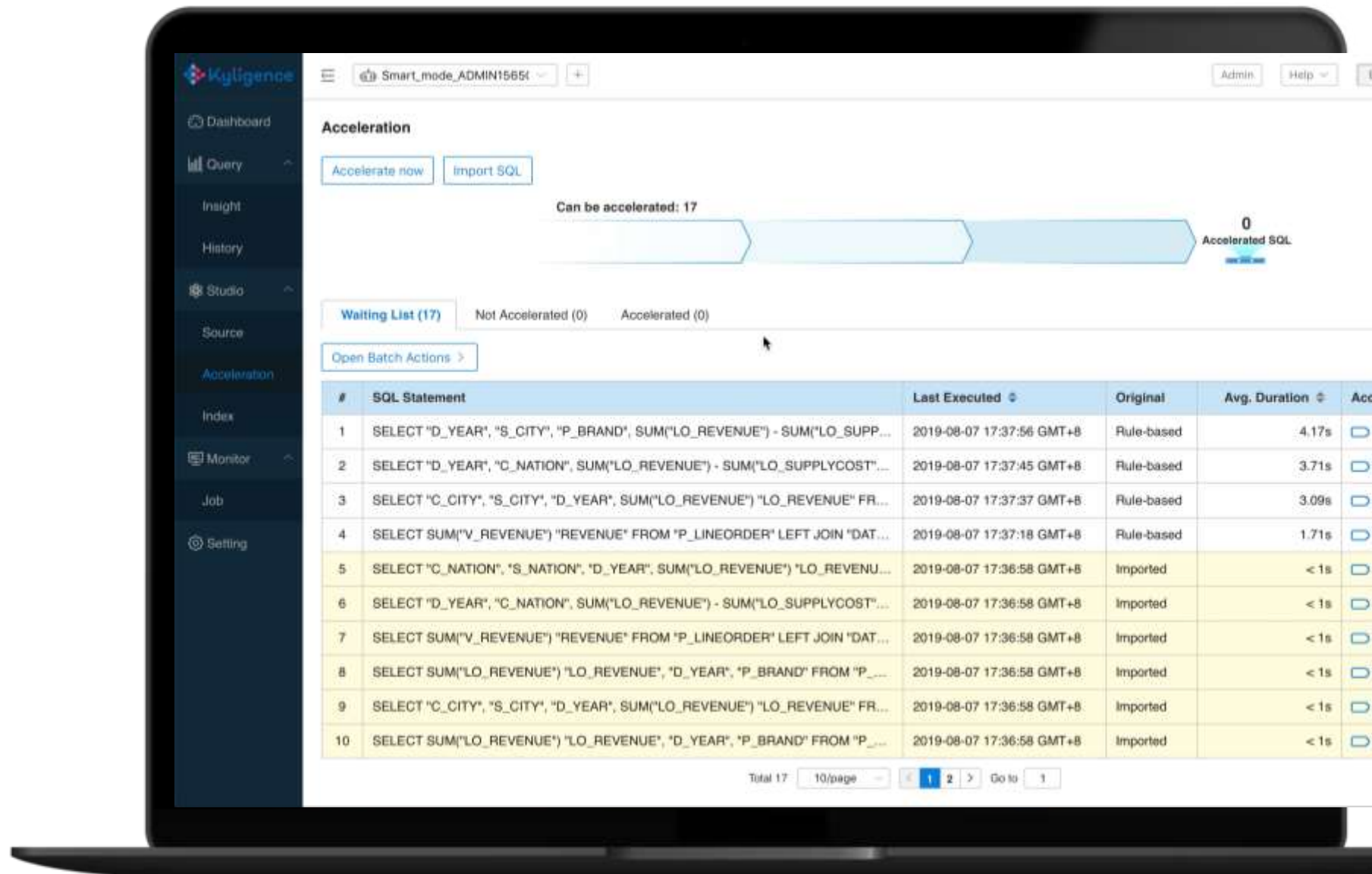


◆ AI Augmented Engine: Intelligent Data Development

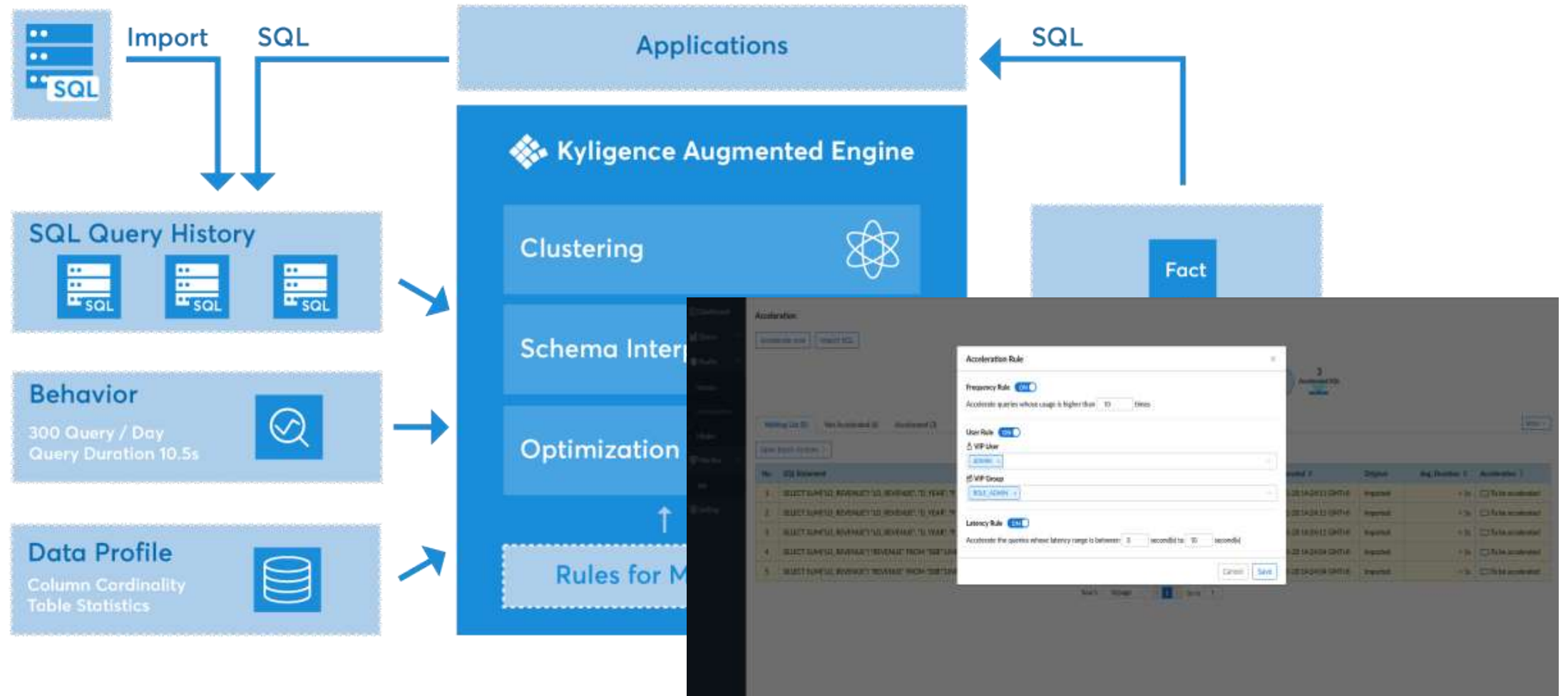


◆ AI Augmented Engine: One-click Acceleration

- Self-maintaining
- Dynamic auto-modeling
- Self-learning engine
- One-click acceleration
- Adaptive model



◆ AI-Augmented Engine — Learn From Your Analytics History



◆ Advanced Tuning Features – Push Down and Aggregate Index

Segment

Aggregate Index

Table Index

JSON

SQL

Full Screen

Edit Aggregate Group

Build Index

Aggregate Index Tree

Aggregate Amount:53

Empty Aggregate: 0

Broken Aggregate: 0

Aggregate Detail

Search Aggregate ID

| | |
|------------------|---------------------------|
| Last Modify Time | 2020-01-08 12:26:20 GMT-5 |
| Data Size | 9.08 MB |
| Index ID | 650000 |
| Data Range | Null |
| Usage | 0 |

| Order | Content |
|-------|--------------------|
| 1 | STORE_AND_FWD_FLAG |

It is recommended to select the grouping dimensions and filtering dimensions frequently used in the query into the aggregation group according to the cardinality from high to low.

+ Add Aggregate Index

Aggregate-Group-1 (31)

Select all columns Clear all

Include

GREEN_TRIP_DATA.STORE_AND_FWD_FLAG

GREEN_TRIP_DATA.PICKUP_DATETIME

PICKUP_LOCATION.BOROUGH

PICKUP_LOCATION.LONGITUDE

PICKUP_LOCATION.LATITUDE

DROPOFF_LOCATION.BOROUGH

DROPOFF_LOCATION.LONGITUDE

DROPOFF_LOCATION.LATITUDE

Mandatory

Please search or select

Hierarchy

Please select or search

Joint

PICKUP_LOCATION.BOROUGH

PICKUP_LOCATION.LONGITUDE

DROPOFF_LOCATION.BOROUGH

DROPOFF_LOCATION.LONGITUDE

Segment

Aggregate Index

Table Index

JSON

SQL

Table Index

Build Index

Search

Long Ago(7)

Table Index ID:20000000001

2020-01-07 10:25:00 GMT-8

19 Columns

Table Index ID:20000000001

2020-01-07 11:25:00 GMT-8

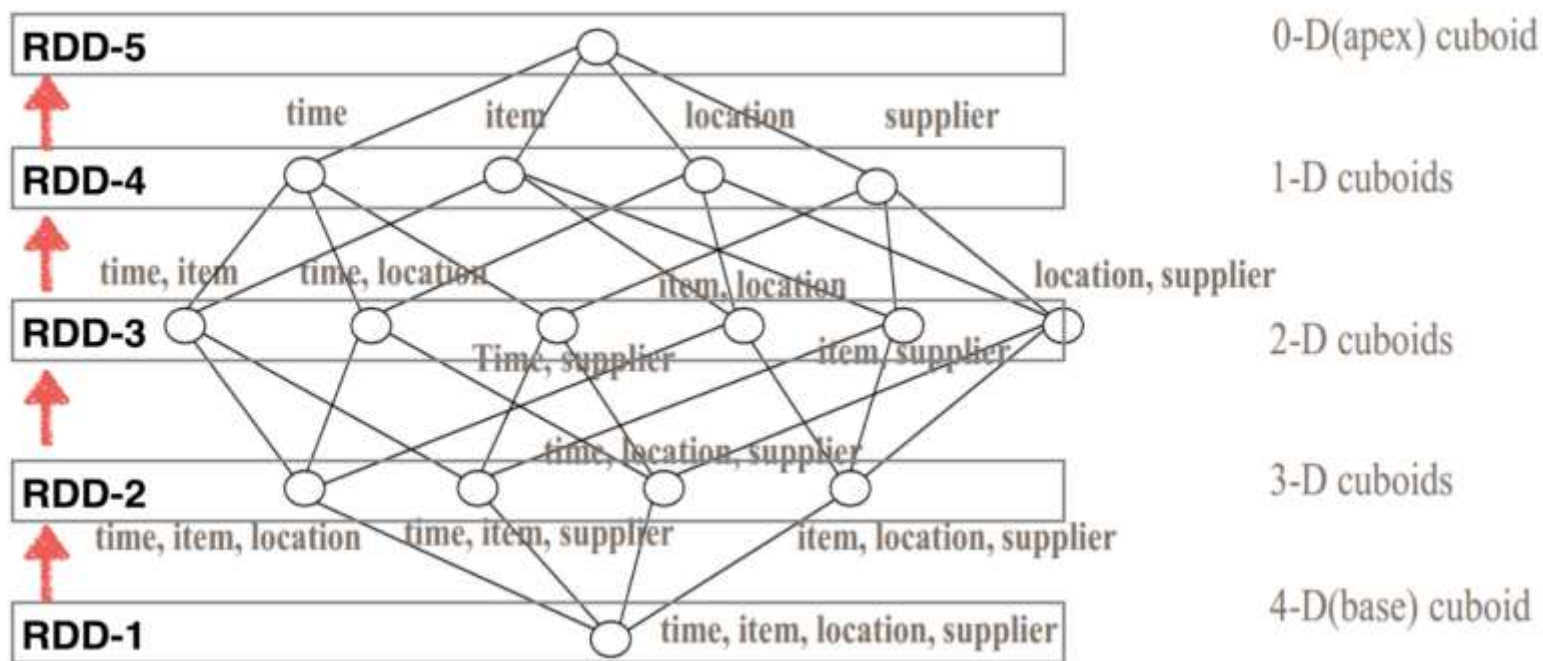
2 Columns

Table Index ID:20000000001

2020-01-02 18:00:01 GMT-8

8 Columns

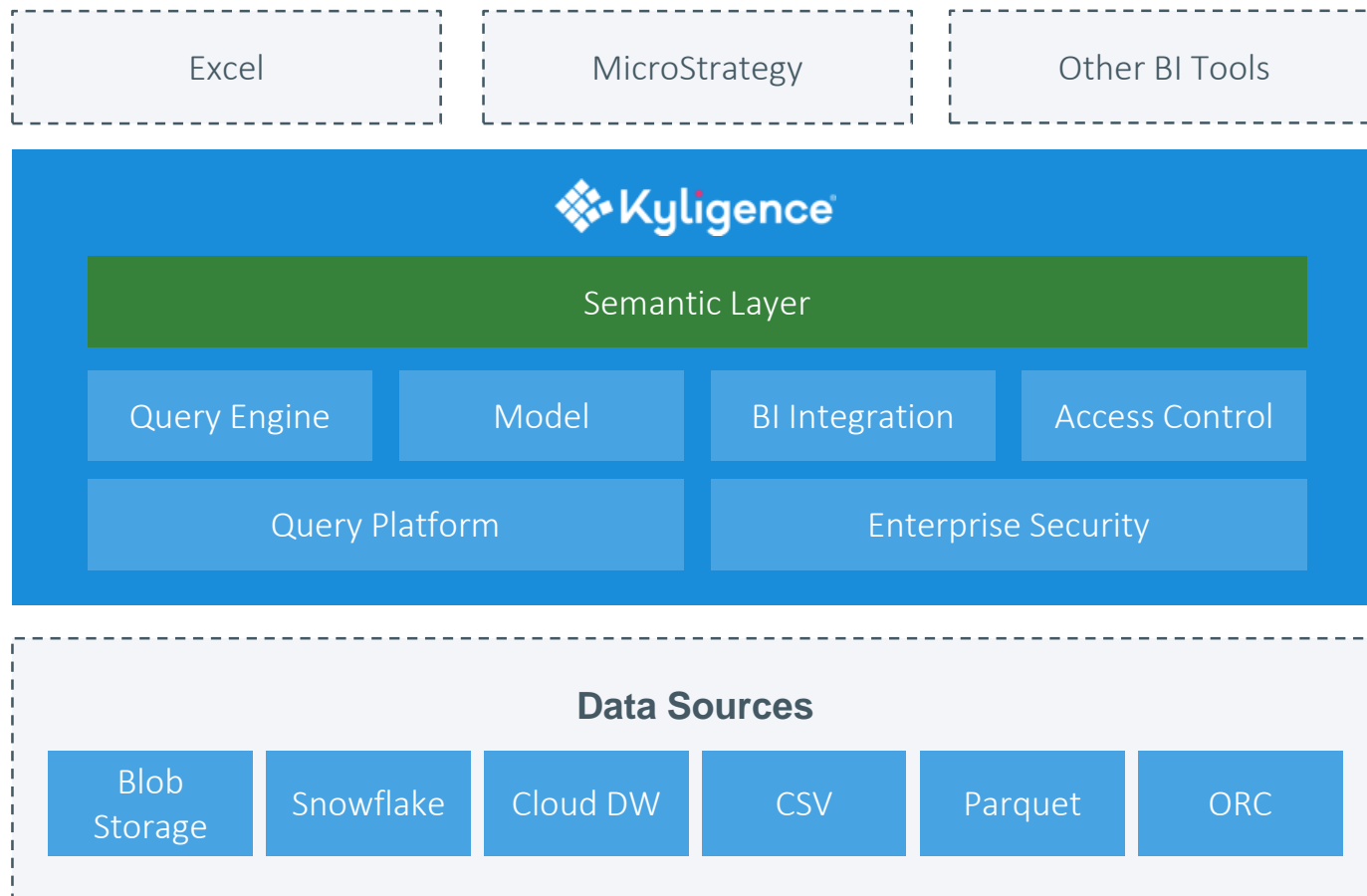
◆ Under the hood : Smart Cuboids



- Each Model consists of N-Dimension Cuboids which is a combination of several dimension in different permutations and combinations.
- Apache Spark is used to build the cuboids making query results extremely fast.
- When the user sends a query the model intelligently looks for the Cuboids/segment returns the results extremely fast.

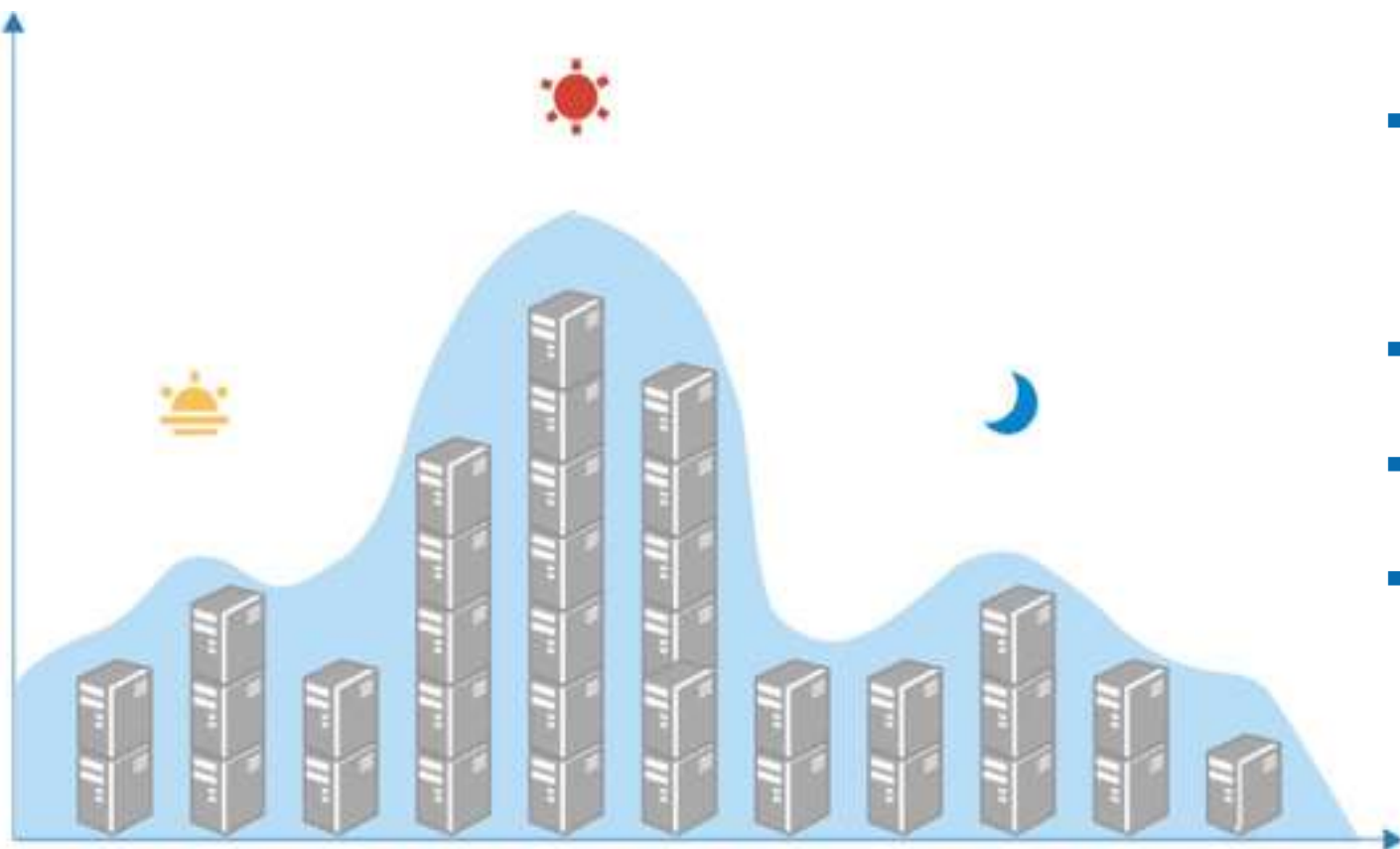


Unified Semantic Layer



- Translate technical details into business terminology
- Synchronize semantics across major BI tools
- Unified business definitions
- Flexible business calculations

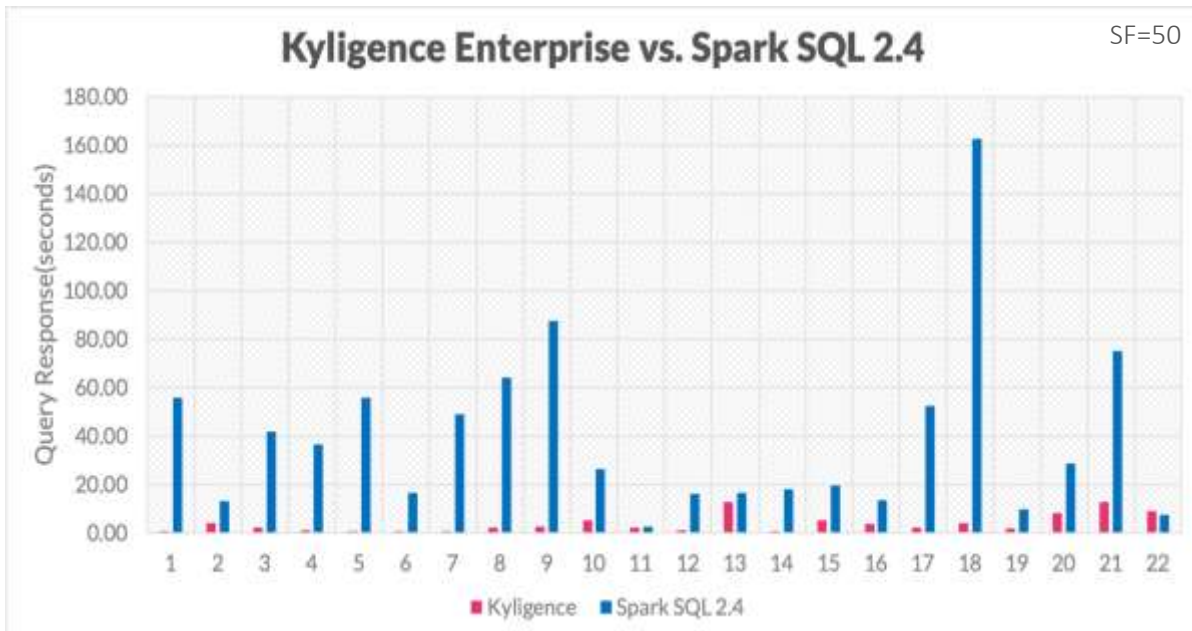
◆ Elastic Scaling — Handle Peak Time Automatically



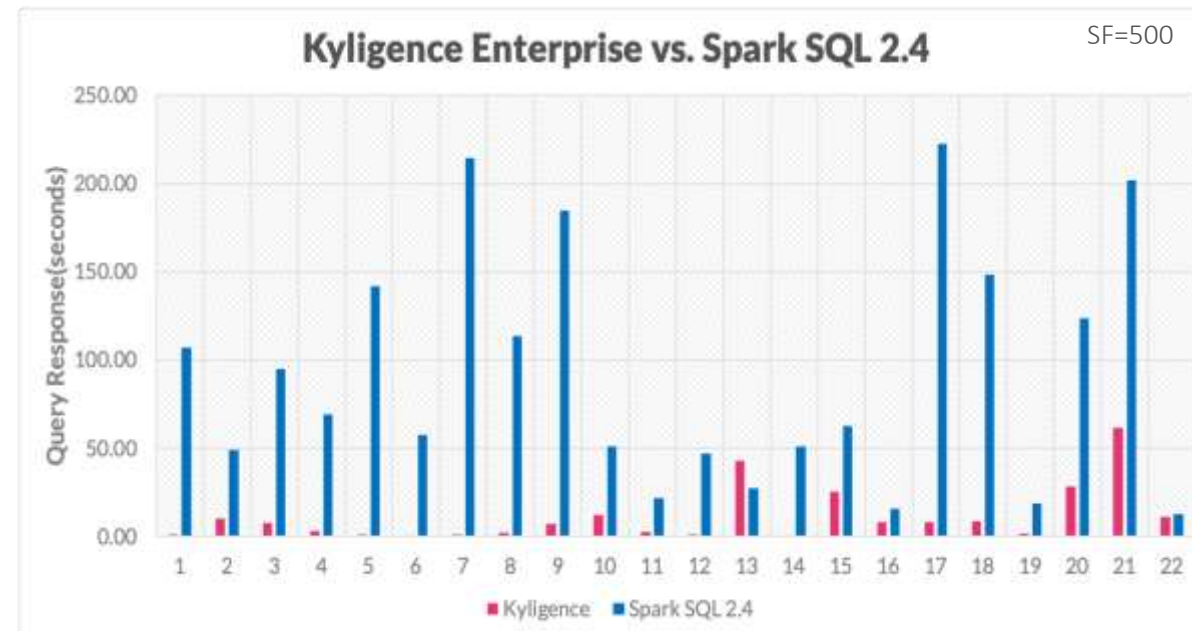
- Fewer compute and storage resources utilized
- Dynamic on-demand cluster resizing
- Uses spot instances
- Efficient planning for data growth

◆ TPC-H 22 Queries

Query Response Time | 0.5 Billion



Query Response Time | 5 Billion



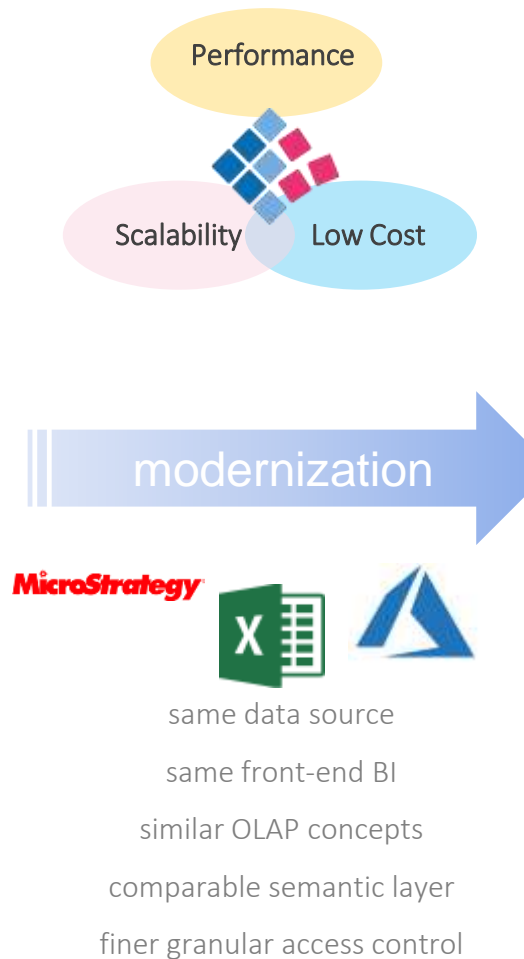
For each Dataset:

- Run each query 3 times
- Record the average time
- No warm up
- Lower is better

◆ Financial Risk Management - replacing the large SSAS cube

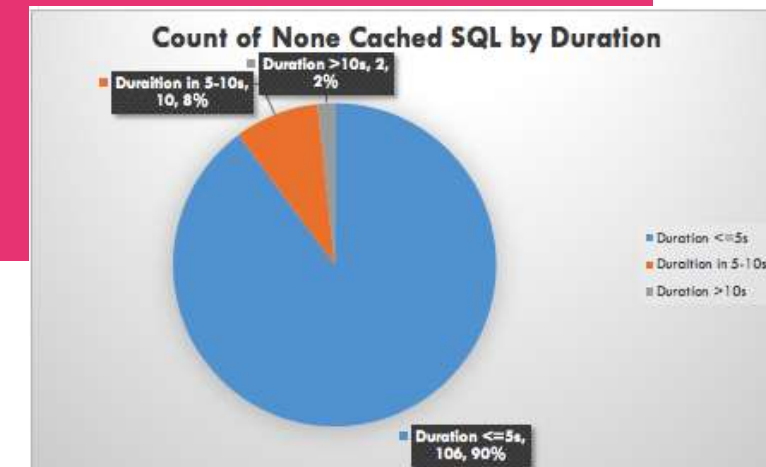
Challenges

- 5TB SSAS cube with 5 Billion rows daily incremental data
- 14 Lookup tables, half over 20M cardinalities (largest 200M)
- 600+ dimensions
- 30+ analysis users
- Analysts' work locked by incremental loading workload, system crashes happen frequently
- Poor performance on data loading and queries (especially on UHC, Count Distinct, Correlation)
- Limited concurrent users



Kyligence's Solution

- Single cube easy management
- Analysts' work no longer interrupted
- Transparent to business users, same analysis tool Excel
- Improved query and loading performance
- Support 1000+ concurrent users
- Meet future requirement - prediction of 40% data volume growth, migration to cloud, Realtime





THANK YOU