Strata DATA CONFERENCE

Kyligence

Augmented OLAP for Big Data

Luke Han | luke.han@Kyligence.io

Co-founder & CEO of Kyligence

Apache Kylin PMC Chair

Microsoft Reginal Director & MVP

Strata Global Sponsor

BOOTH #410

About Luke Han

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

About Apache Kylin

- Leading Open Source OLAP for Big Data
- Rank 1 from googling “big data OLAP”
- Rank 1 from googling “hadoop OLAP”
- Open sourced by eBay in 2014
- Graduated to Apache Top Project in 2015
- 1000+ Adoptions world wild
- 2015 InfoWorld Bossie Awards
- 2016 InfoWorld Bossie Awards

Agenda

- About Kyligence
- Pains in Big Data Analysis
- Kyligence's solution: Augmented OLAP
  - Video Demo
  - Benchmark
- Use Cases

Kyligence = Kylin + Intelligence

- Founded in 2016 by the original creators of Apache Kylin
- CRN Top 10 Big Data Startups 2018
- Backing by leading VCs:
  - Redpoint Ventures
  - Cisco
  - CBC Capital
  - Shunwei Capital
  - Eight Roads Ventures (Fidelity International Arm)
  - Coatue
- Global Offices:
  - Shanghai
  - Beijing
  - Shenzhen
  - San Jose
  - New York
  - Seattle
  - …

Trusted by Global Leaders

Most of them are Global Fortune 500

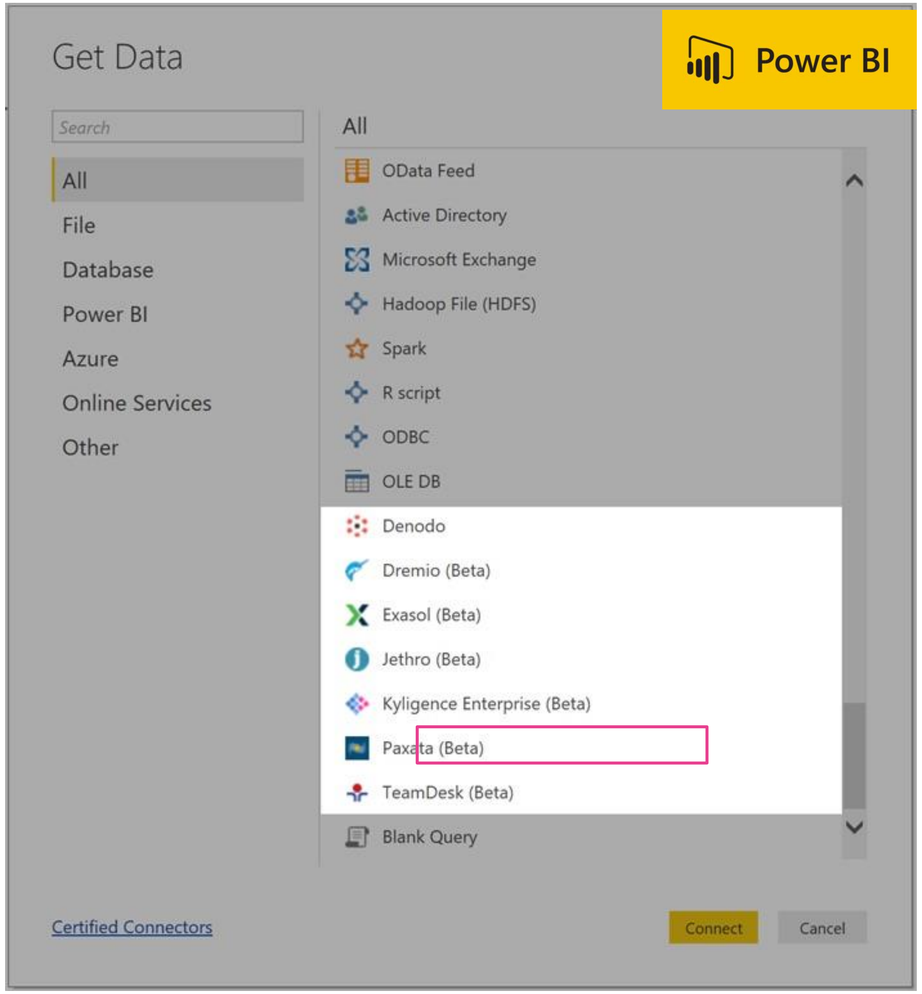
Finance

Telecom

Manufacturing

Retail &
Others

Global Partners



Agenda

- About Kyligence
- Pains in Big Data Analysis
- Kyligence’s solution: Augmented OLAP
- Use Cases

Let’s talk about photography story…

https://technave.com/data/files/mall/article/201812271418327393.jpg

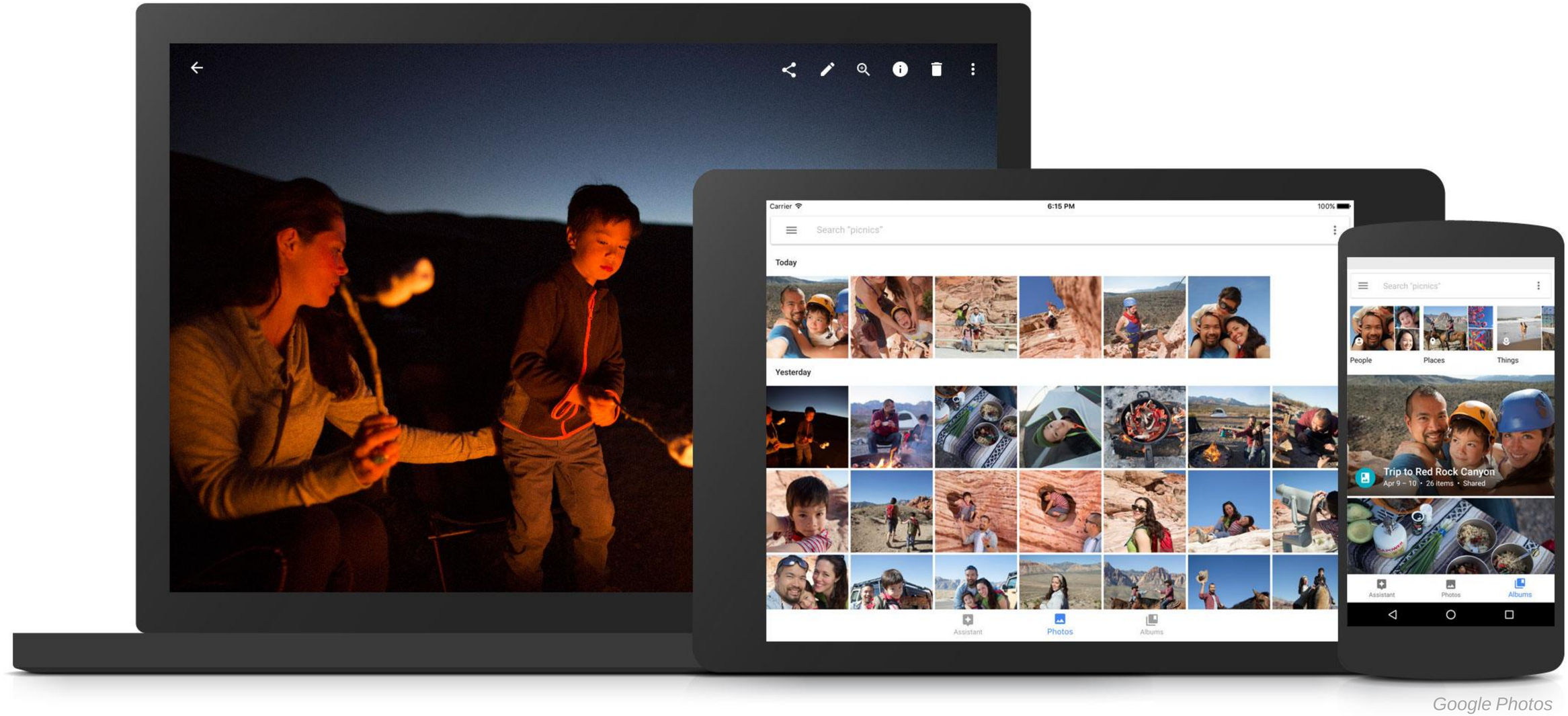
Let’s talk about photography story…

How many people really know how to setup those?

Let’s talk about photography story…

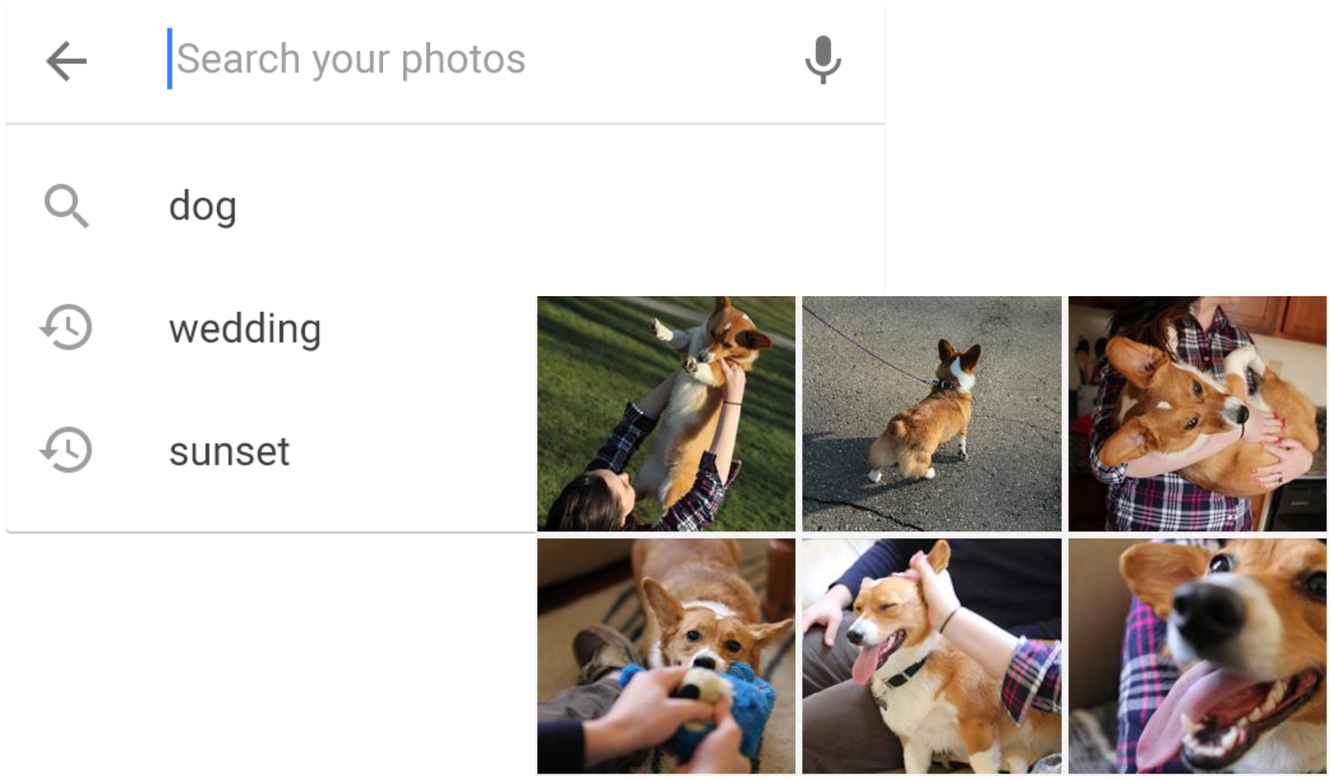
https://technave.com/data/files/mall/article/201812271437395703.jpg

Let’s talk about photography story…



Google Photos

Let’s talk about photography story…



Google Photos

How do you manage your 100,000+ photos?

Then…

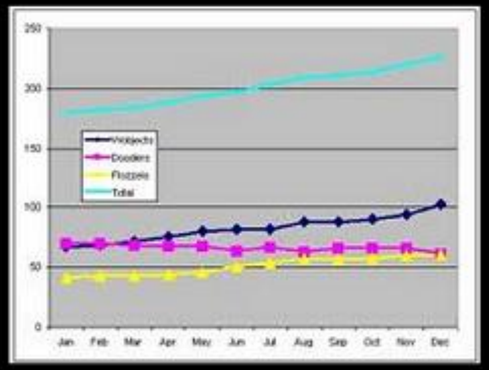
how about your enterprise data?

Data Scientist

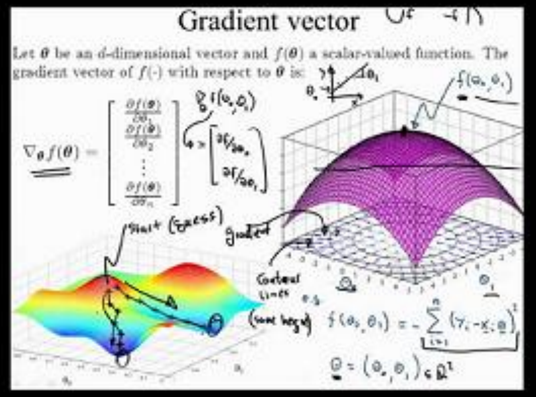
What my friends think I do

What my mom thinks I do

What society thinks I do



What my boss thinks I do



What I think I do

What I actually do

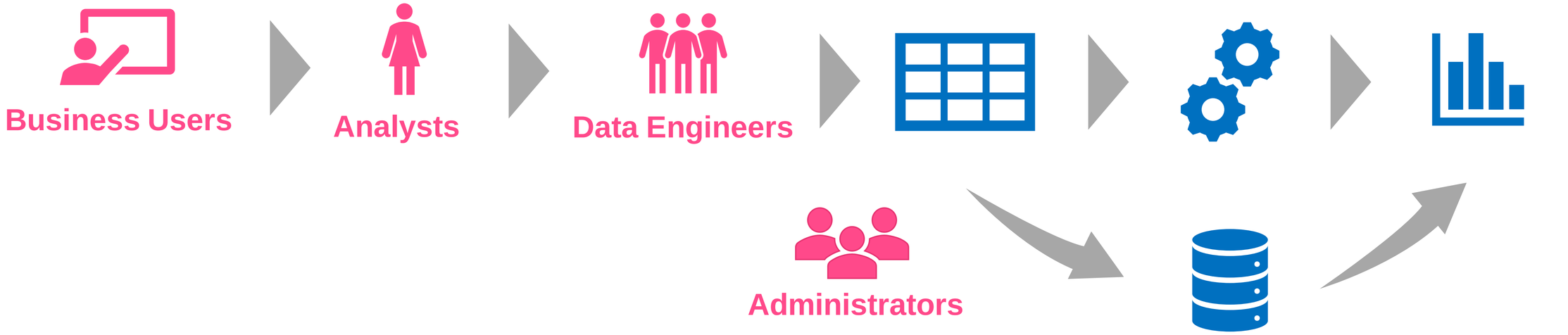
https://www.sintetia.com/wp-content/uploads/2014/05/Data-Scientist-What-I-really-do.png

Fast and Changing
Analysis Demand

vs

Slow and Heavy
Big Data Operations

The Typical “Throw in some People” Approach

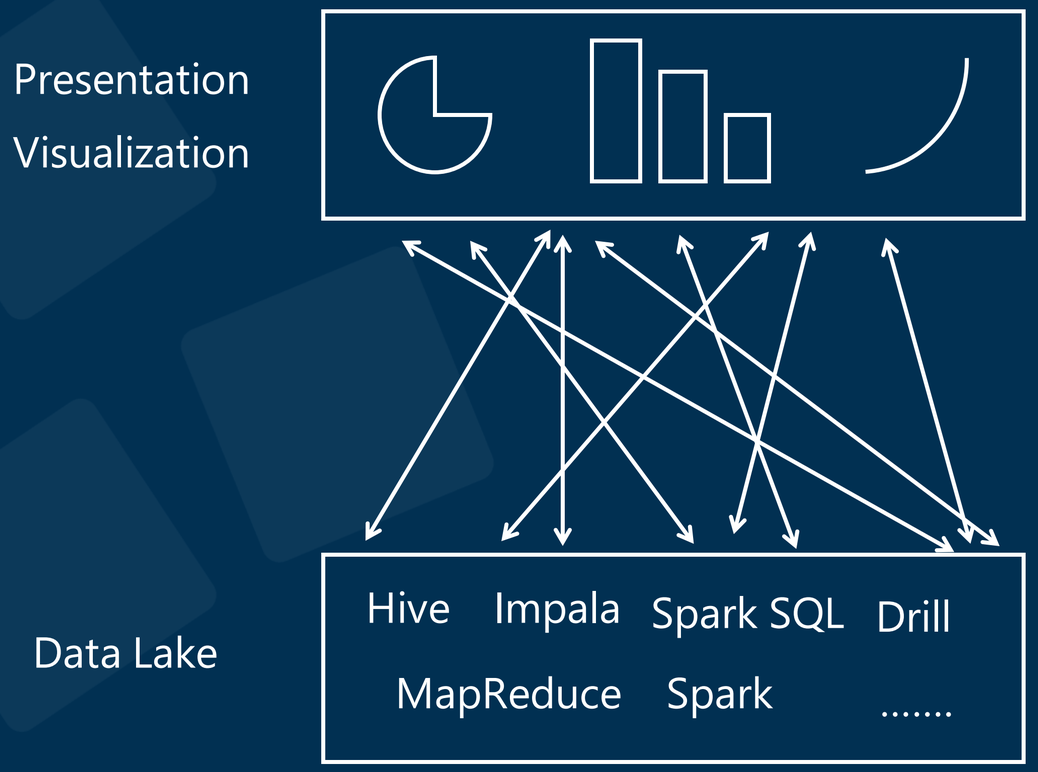


Business Analysis → Data Modeling → Lake → Warehouse → Mart → Reporting

Slow Time-to-Insight:

High Cost : $$$

Pains in the “Throw in some People” Approach



Time-to-value Pain

Weeks of waiting breaks the “online” promise.

Collaboration Pain

Hard to reuse asset across teams. Each team fights their own path.

Resource Pain

Hard to scale. Where to find so many skilled big data engineers?

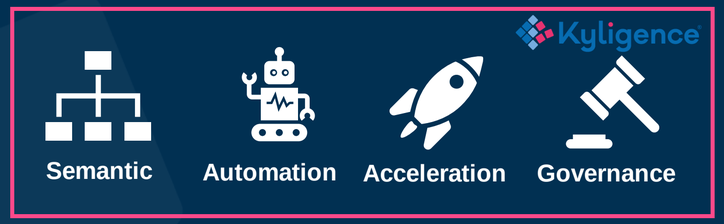
Agenda

- About Kyligence
- Pains in Big Data Analysis
- Kyligence’s solution: Augmented OLAP
- Use Cases

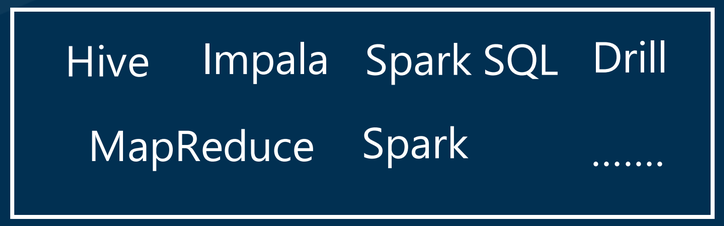
Throw in some Intelligence!

Presentation
Visualization

Augmented OLAP
Data Mart



Data Lake



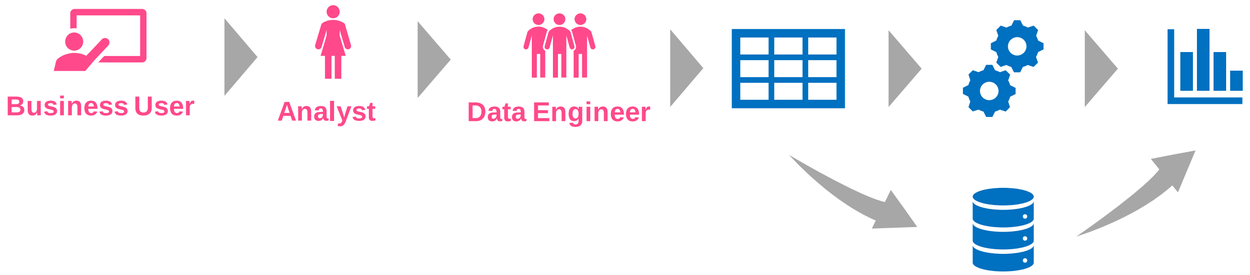
Let a system replace the people.

- Transparent SQL Acceleration
- On-demand Data Preparation
- Interactive Query Performance
- High Concurrency
- Centralized Semantic Layer

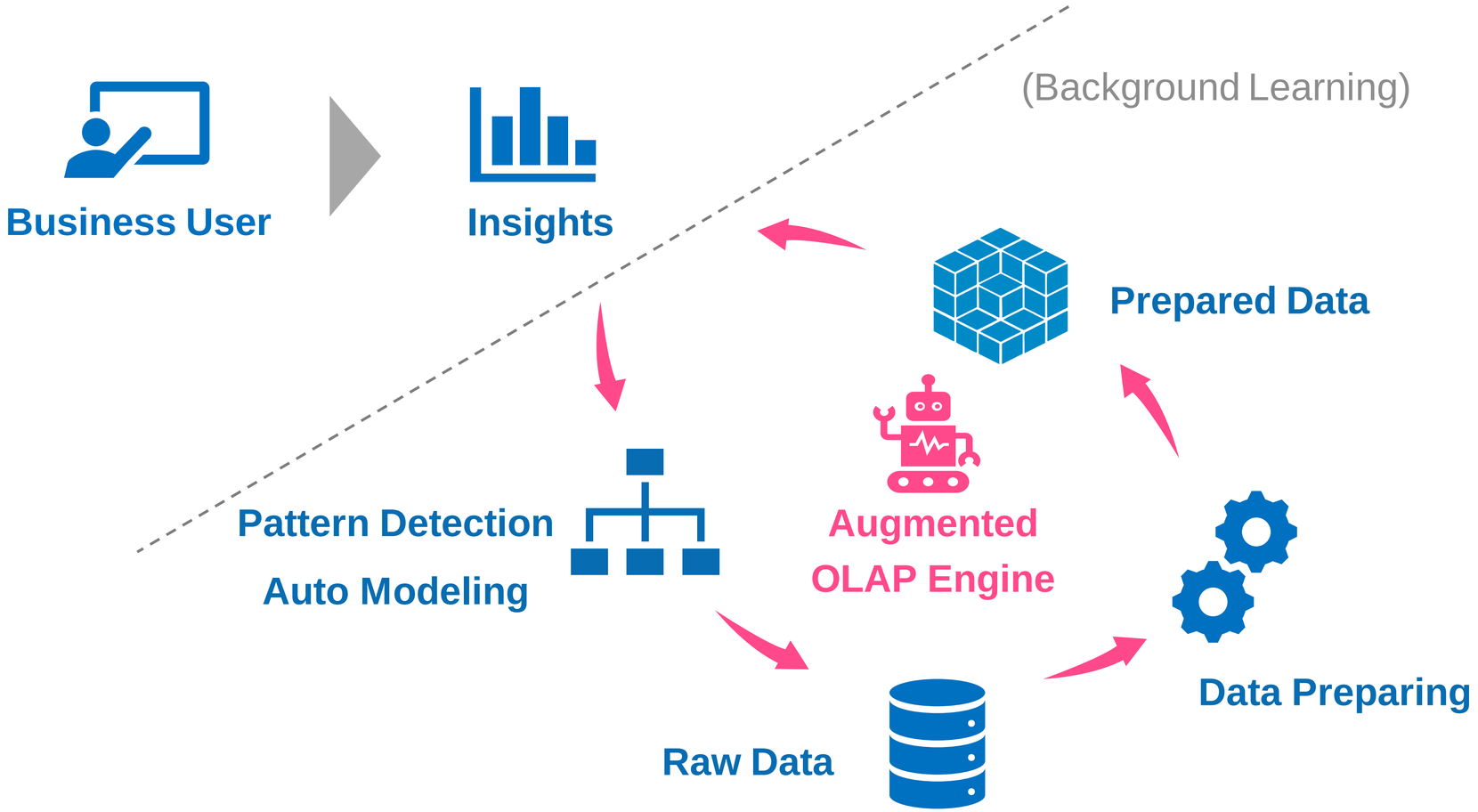
Faster time to market. Stay “online”.

A Learning OLAP System

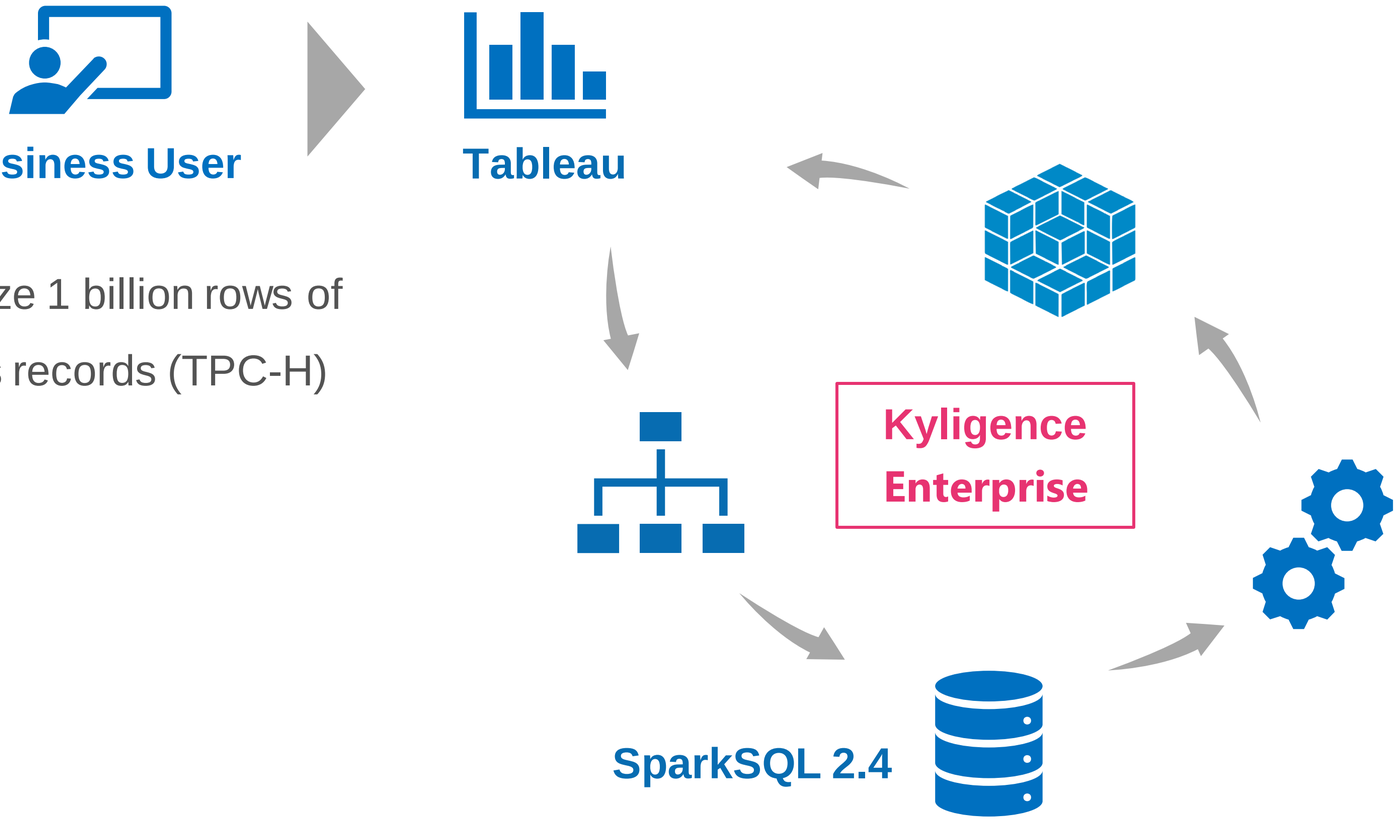
VS



A Learning OLAP System



Demo Setup



Analyze 1 billion rows of sales records (TPC-H)

(Embed the Demo Video)

Demo FAQ

How to improve the first slow exploration?

What if the analyst operates differently the second time?

More comprehensive performance benchmark?

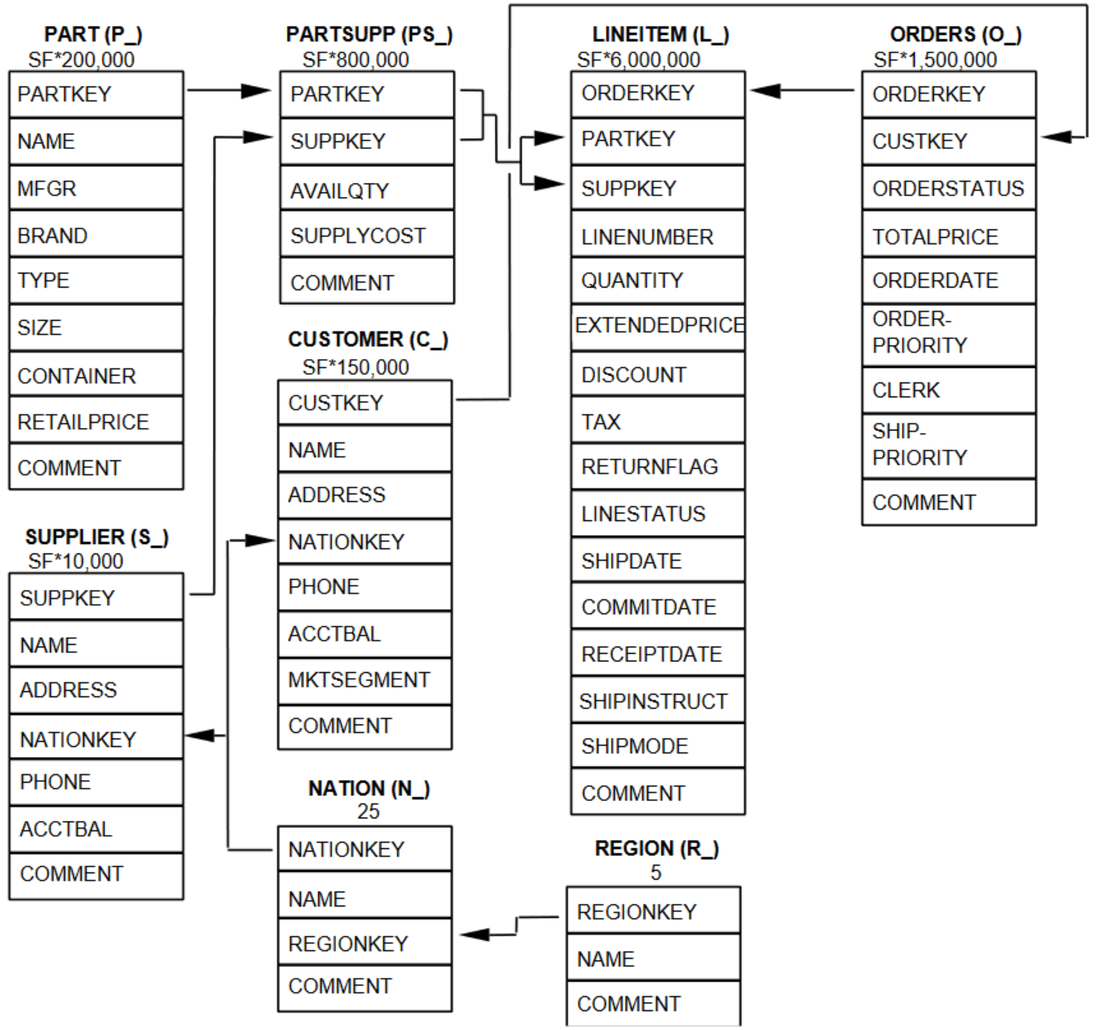
TPC-H Decision Support Benchmark

TPC-H Benchmark

- Examine large volumes of data
- High complexity queries
- Answers critical business questions
- 22 decision making queries

E.g. The Shipping Priority Query

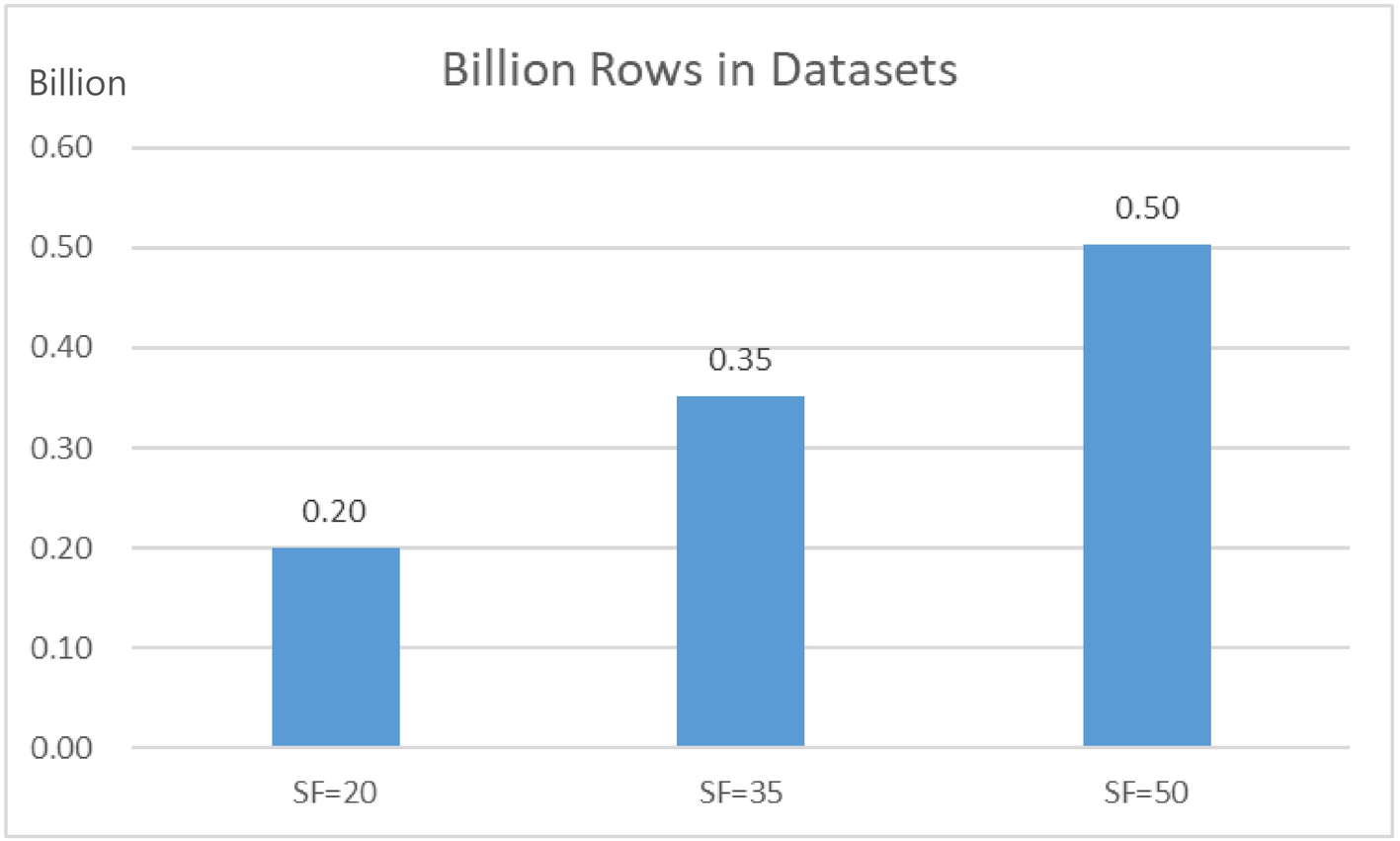
retrieves the shipping priority and potential revenue of the orders having the largest revenue among those that had not been shipped as of a given date. Top 10 orders are listed in decreasing order of revenue.



Kyligence Enterprise 4 Beta vs SparkSQL 2.4

To see the trend as data grows

- 3 datasets
- Scale Factor = 20, 35, 50



- TPCH_SF1: Consists of the base row size (several million elements).
- TPCH_SF20: Consists of the base row size x 20.
- TPCH_SF35: Consists of the base row size x 35.
- TPCH_SF50: Consists of the base row size x 50 (several hundred million elements).

Hardware Configurations

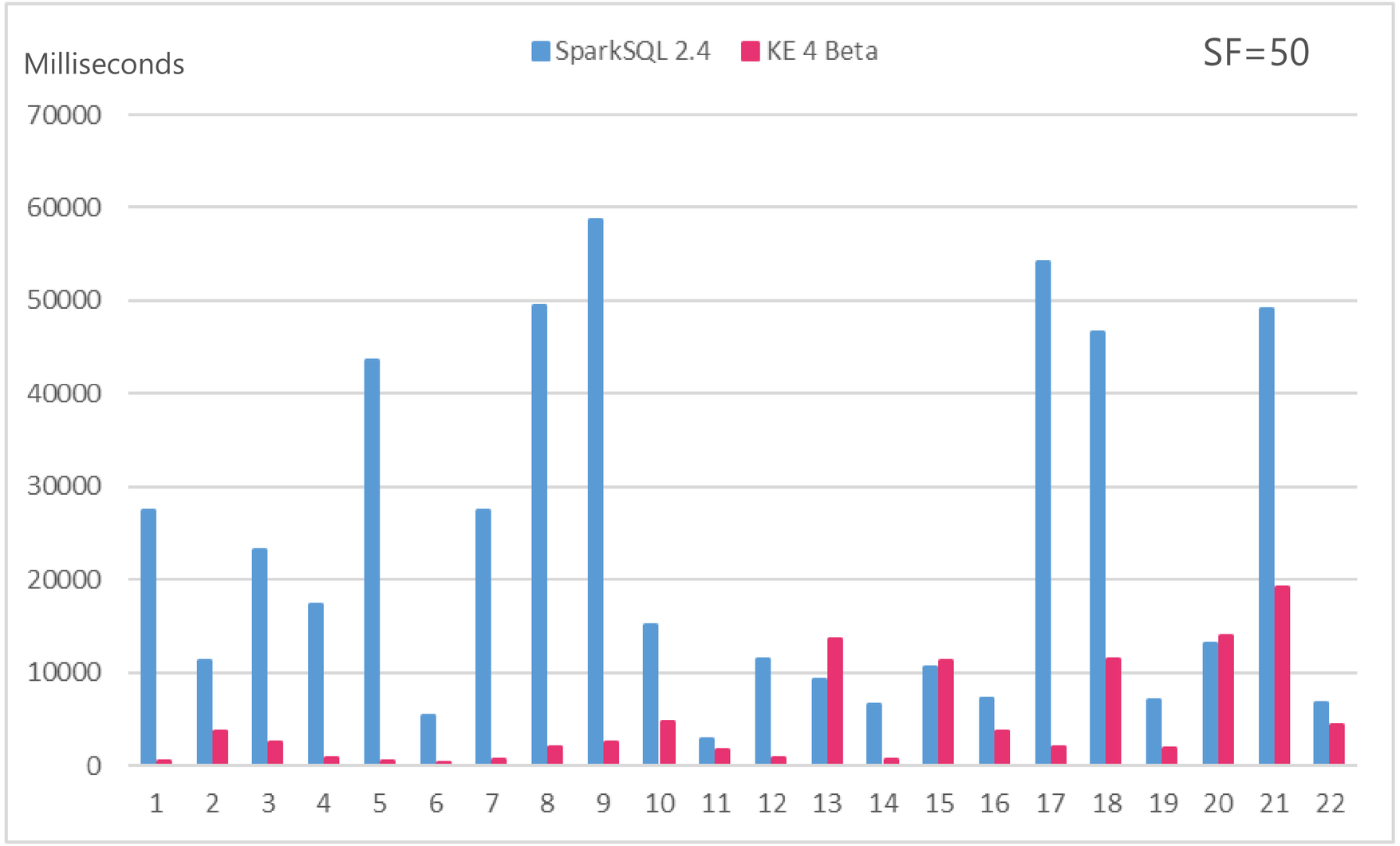
Same 4 physical nodes

- Intel(R) Xeon(R) CPU E5-2630 v4 @ 2.20GHz * 2
- Totally 86 vCores, 188 GB mem

Same Spark configuration for both KE 4 Beta and SparkSQL 2.4

- spark.driver.memory=16g
- spark.executor.memory=8g
- spark.yarn.executor.memoryOverhead=2g
- spark.yarn.am.memory=1024m
- spark.executor.cores=5
- spark.executor.instances=17

Query Response Time | KE 4 Beta vs. SparkSQL 2.4



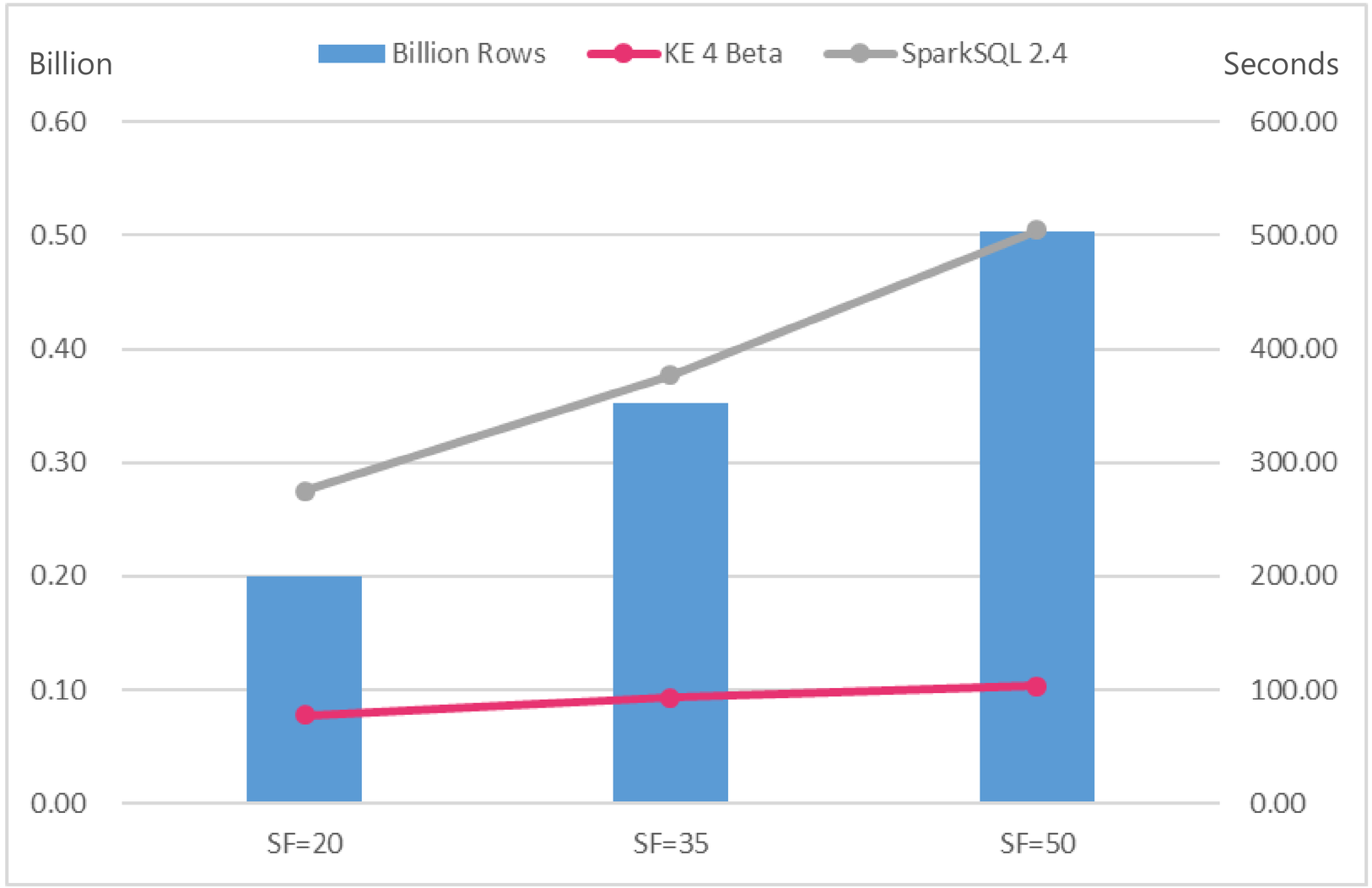
TPC-H 22 queries

For each dataset

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

Lower is better.

Total Response Time | KE 4 Beta vs. SparkSQL 2.4

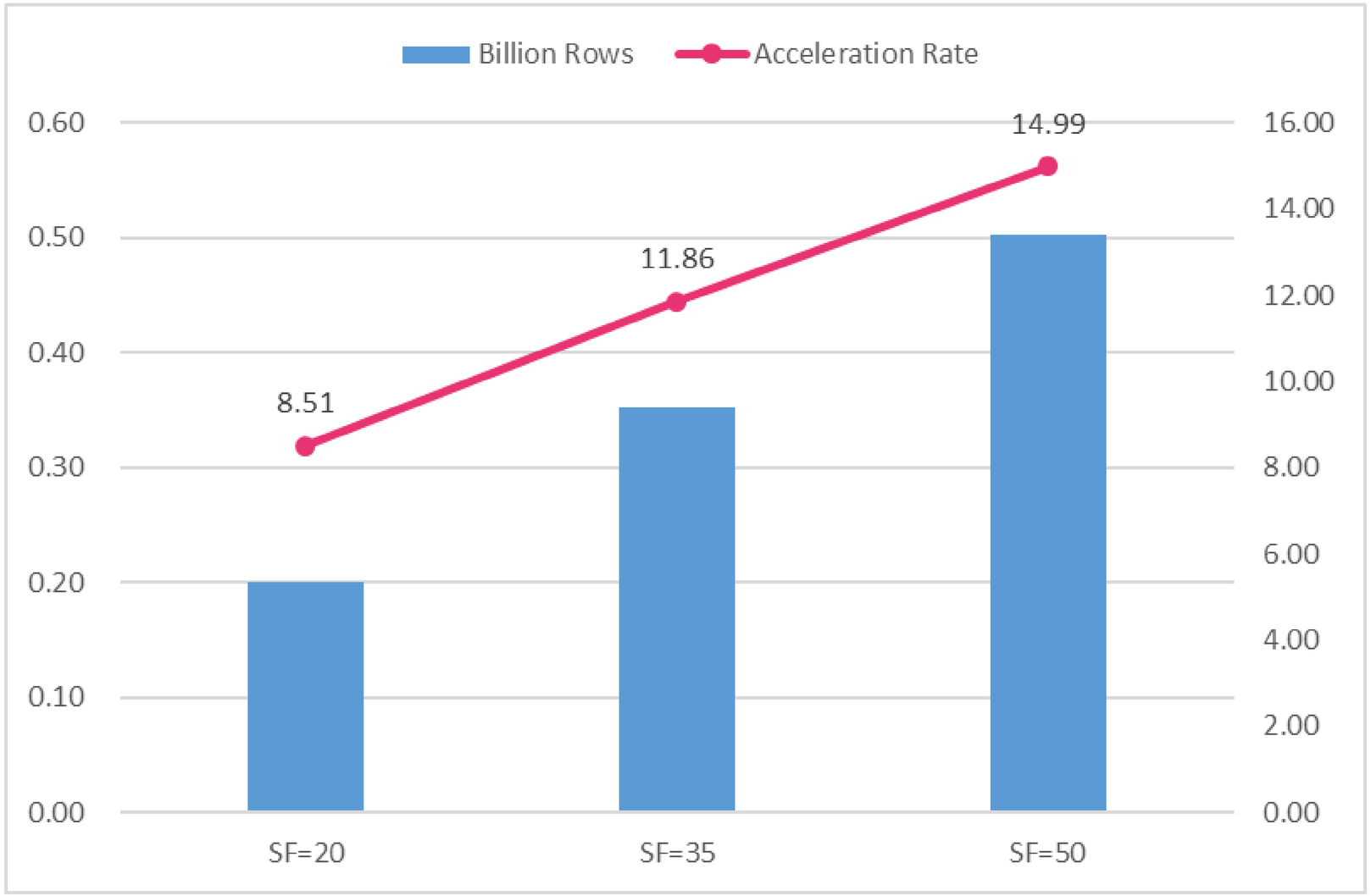


Total response time is the sum of 22 queries' response time.

Compare over the size of datasets and feel the trend.

Scale out for the future.

Avg. Acceleration Rate | KE 4 Beta vs. SparkSQL 2.4

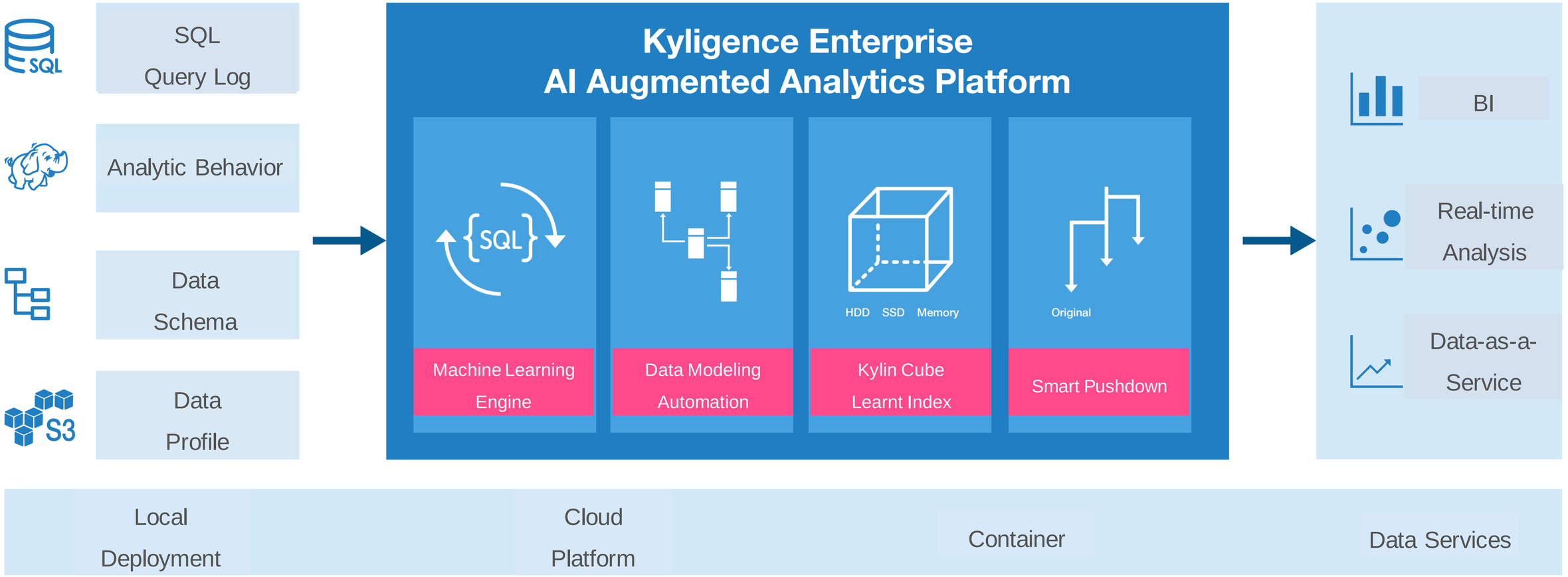


Acceleration Rate

= SparkSQL time / KE time

Take average of the 22 and compare over size of datasets.

AI-Augmented Analytics Platform



Agenda

- About Kyligence
- Pains in Big Data Analysis
- Kyligence’s solution: Augmented OLAP
- Use Cases

Use Case: IBM Cognos Replacement

One Kyligence Cube for 800+ Cognos Cubes

Kyligence

Data: 300+ B Records

Merchants: 10+m

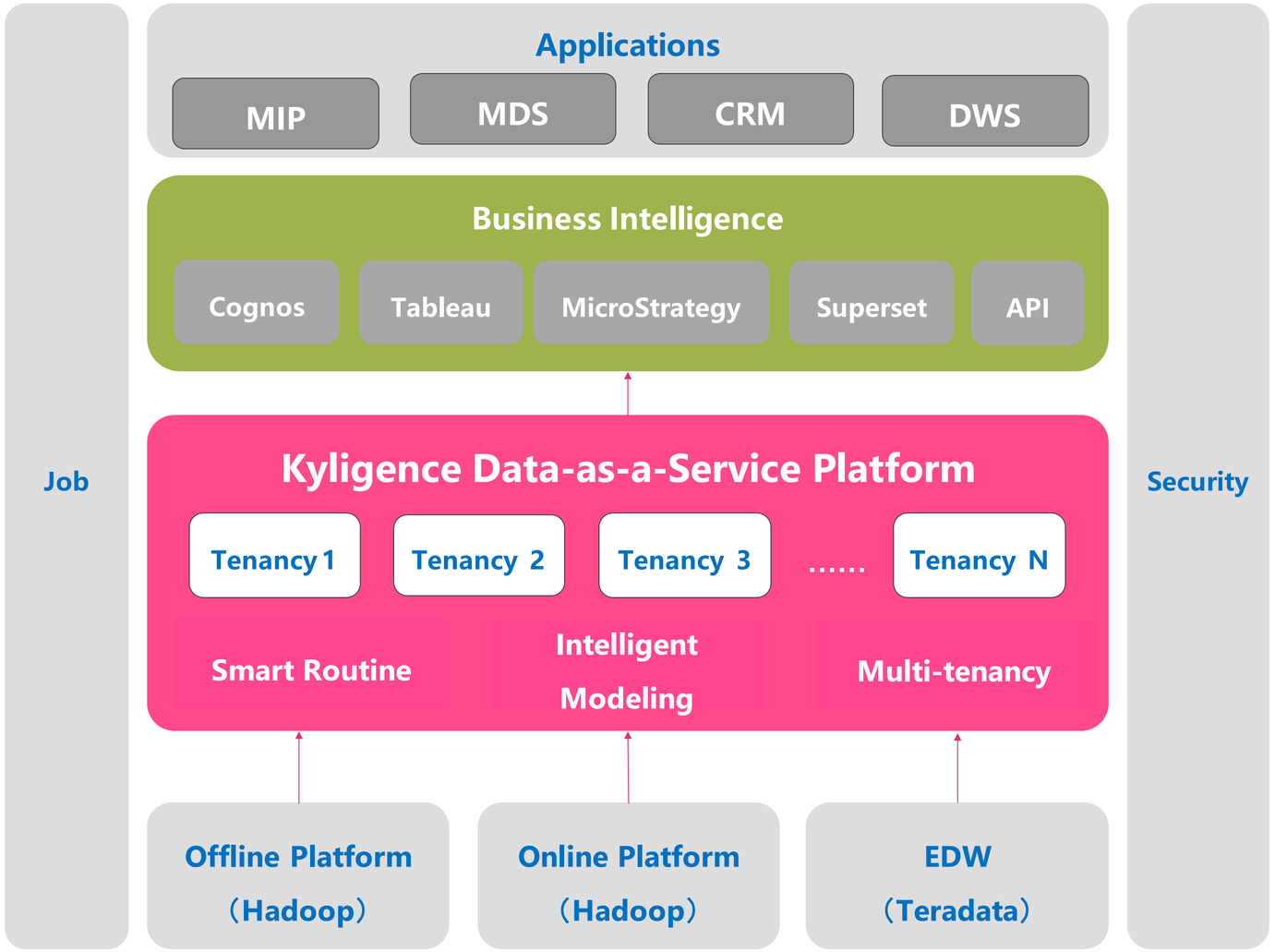
Cards: 10+B

Card Transaction

Dimensions: 167

Measures: 20

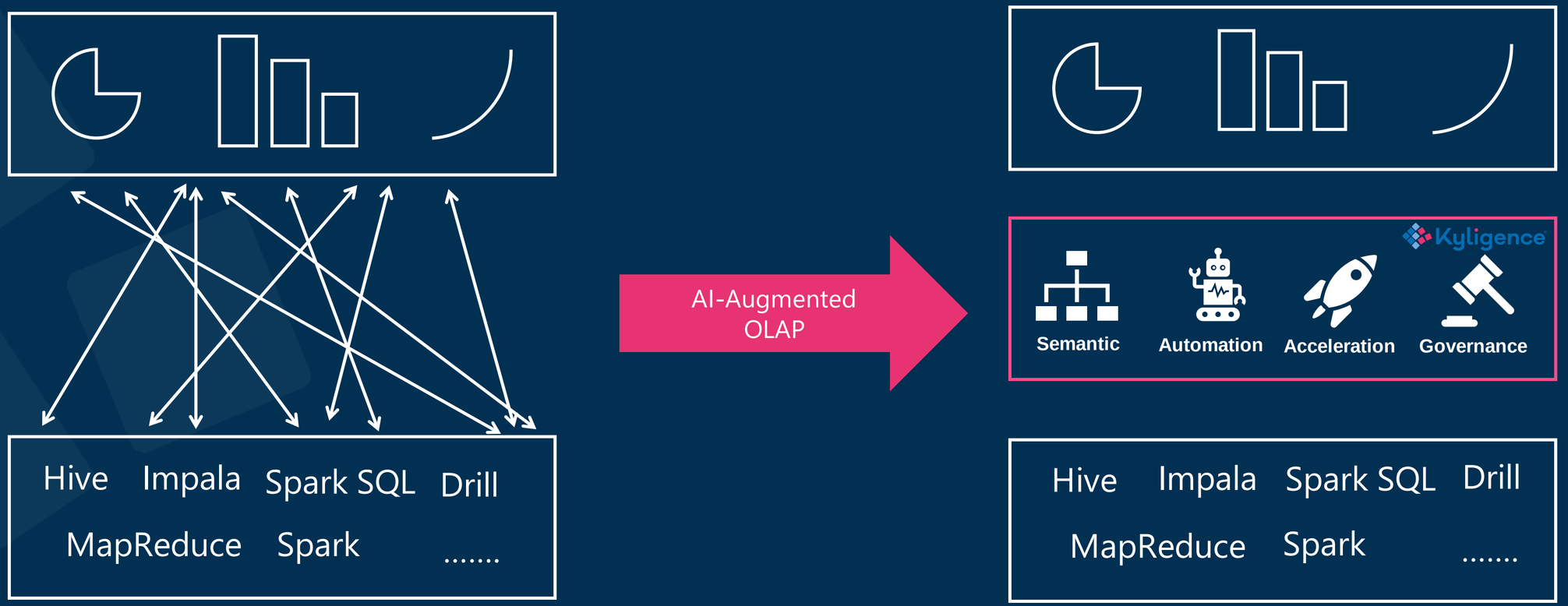
Use Case: Data as a Services Platform



In the past, due to the limitations of our previous multi-dimensional analytic tool, we faced challenges of constrained time range in queries……We are considering leveraging multi-dimensional data cubes to replace a number of fragmented legacy tabular reports in more business units, so that we can provide better analytic services to our business users.”

-- Laments Wu Ying, VP of CMBs Development Center,

Take away: Augmented OLAP, the future for analytics



Kyligence

Thanks

luke.han@Kyligence.io | @lukehq

Homepage: http://kyligence.io

Twitter: @kyligence

Booth: #410