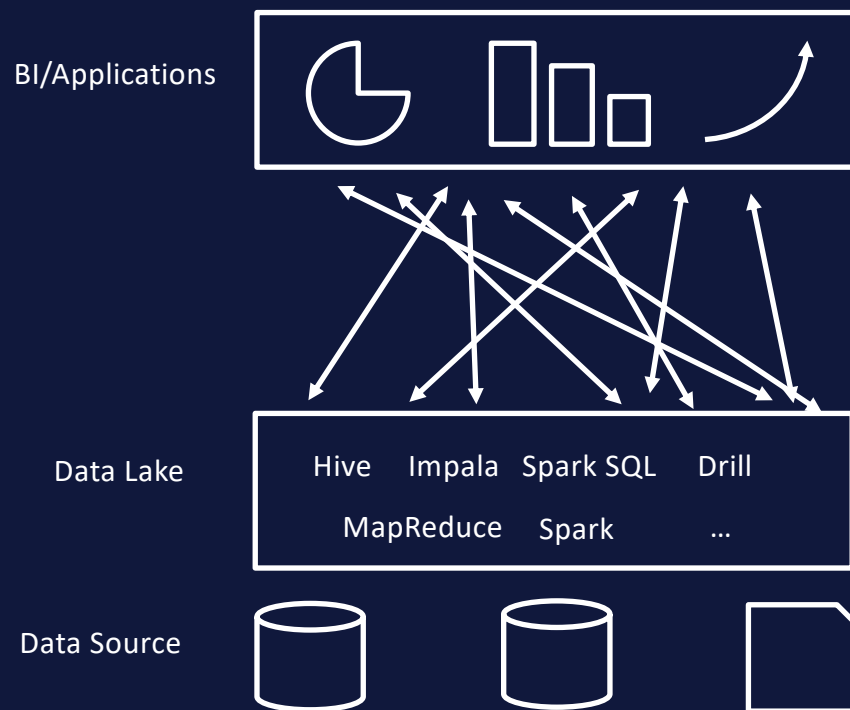

Kyligence: 释放大数据生产力

刘一鸣 yiming.liu@kyligence.io
Kyligence VP
Apache Kylin PMC & Committer

OLAP: The Missing Part of Big Data



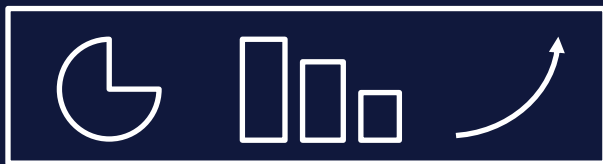
- Too many options
- Low performance
- Long learning curve
- Compatibility issue
- Technology vs Data



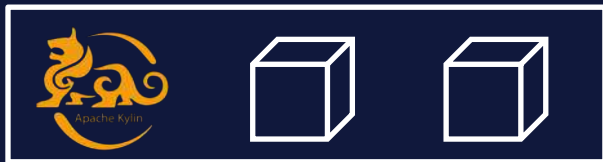
Apache Kylin: Bring OLAP back to Big Data



BI/Applications



OLAP
Data Mart



Data Lake



Data Source



- SQL Acceleration for Big Data
- Semantic Layer
- Speed up Analytics
- ANSI SQL Interface
- High Performance and High Concurrency



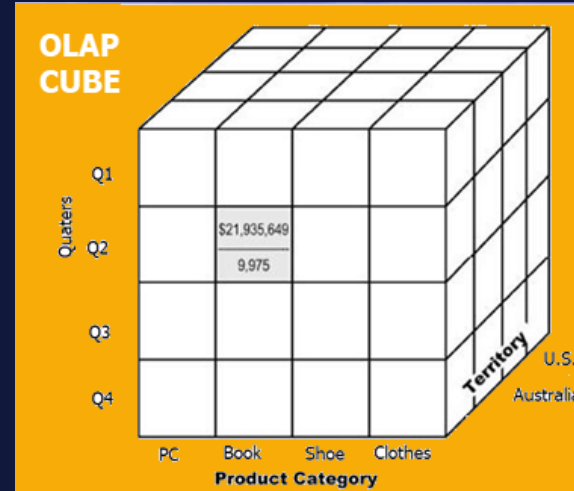
OLAP and OLAP Cube

Online analytical processing, or **OLAP**, is an approach to answering multi-dimensional analytical (MDA) queries swiftly in computing. – Wikipedia

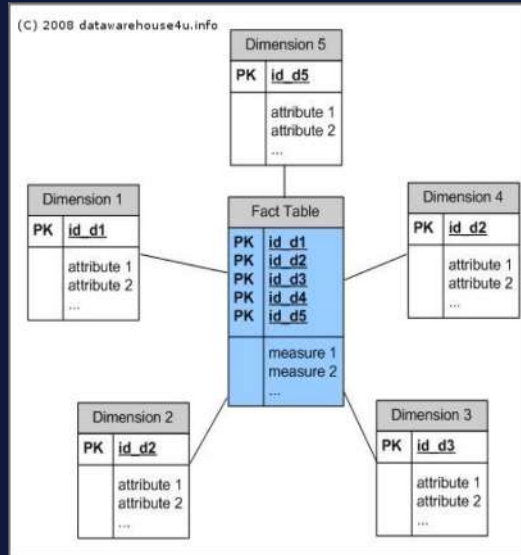
Basic operations

- Roll-up
- Drill-down
- Slice and dice
- Pivot

OLAP cube is a data structure optimized for very quick data analysis.



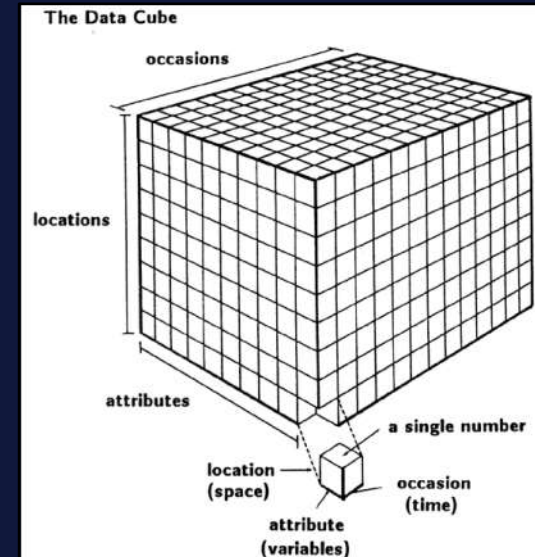
Cube: balance between space and time



Multiple Dimensional Model
--Relational



Classification, aggregation,
and sorting

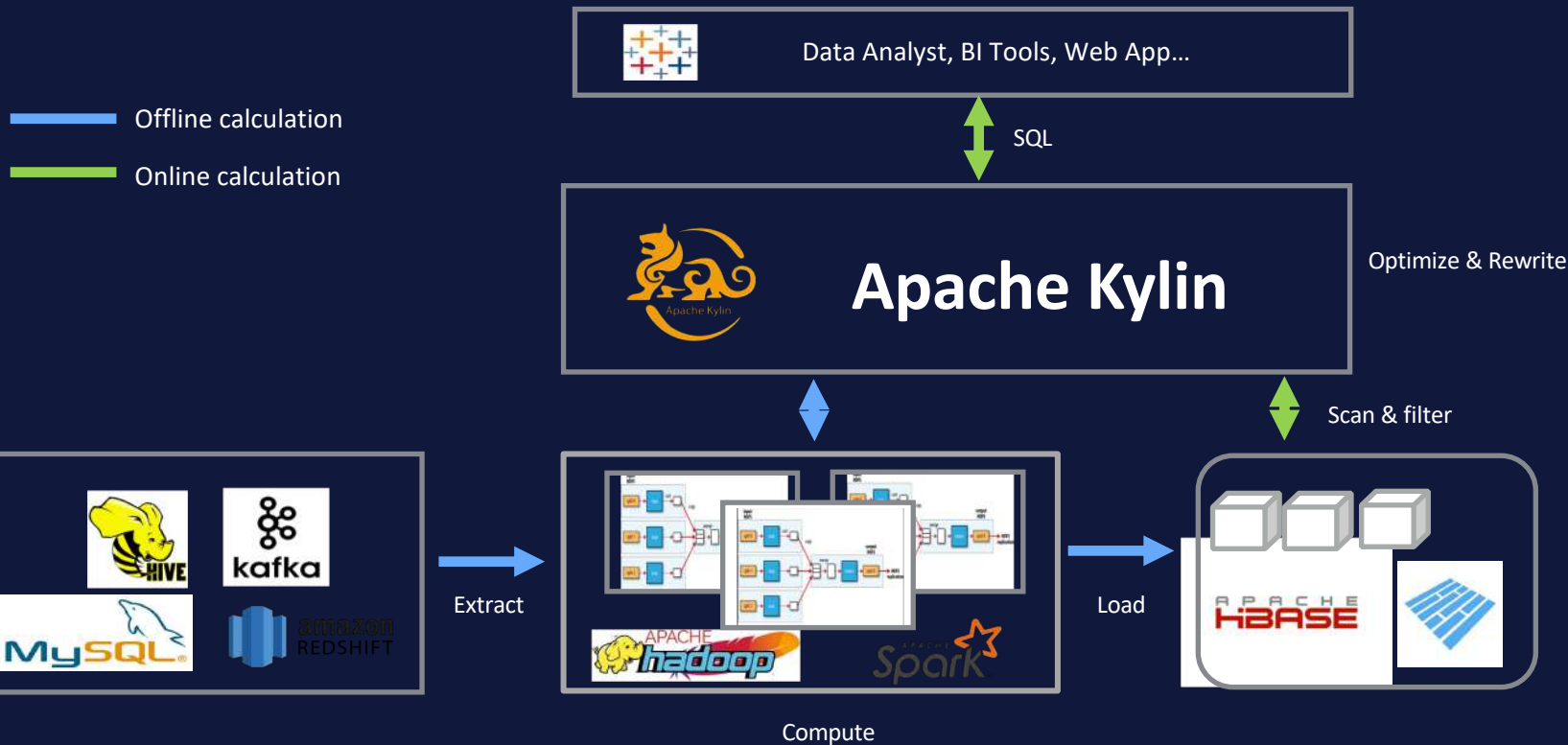


OLAP Cube
--Key-Value



Apache Kylin Architecture Overview

UCLLOUD



SQL execution plan without Cube

Sample: *Check the order return and order status relationship in a time range*

select

```
l_returnflag,  
o_orderstatus,  
sum(l_quantity) as sum_qty,  
sum(l_extendedprice) as sum_base_price
```

from

```
v_lineitem
```

inner join

```
v_orders on l_orderkey = o_orderkey
```

where

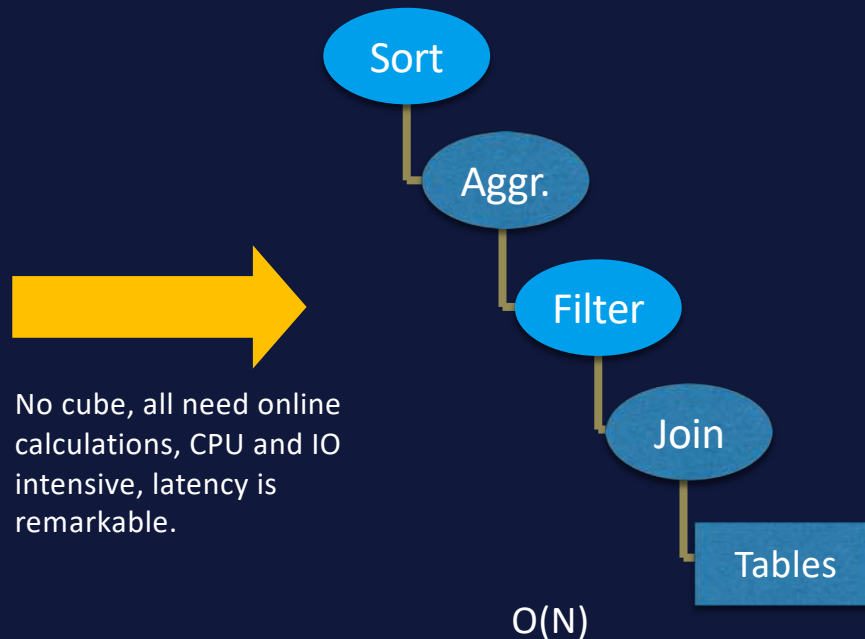
```
l_shipdate <= '1998-09-16'
```

group by

```
l_returnflag,  
o_orderstatus
```

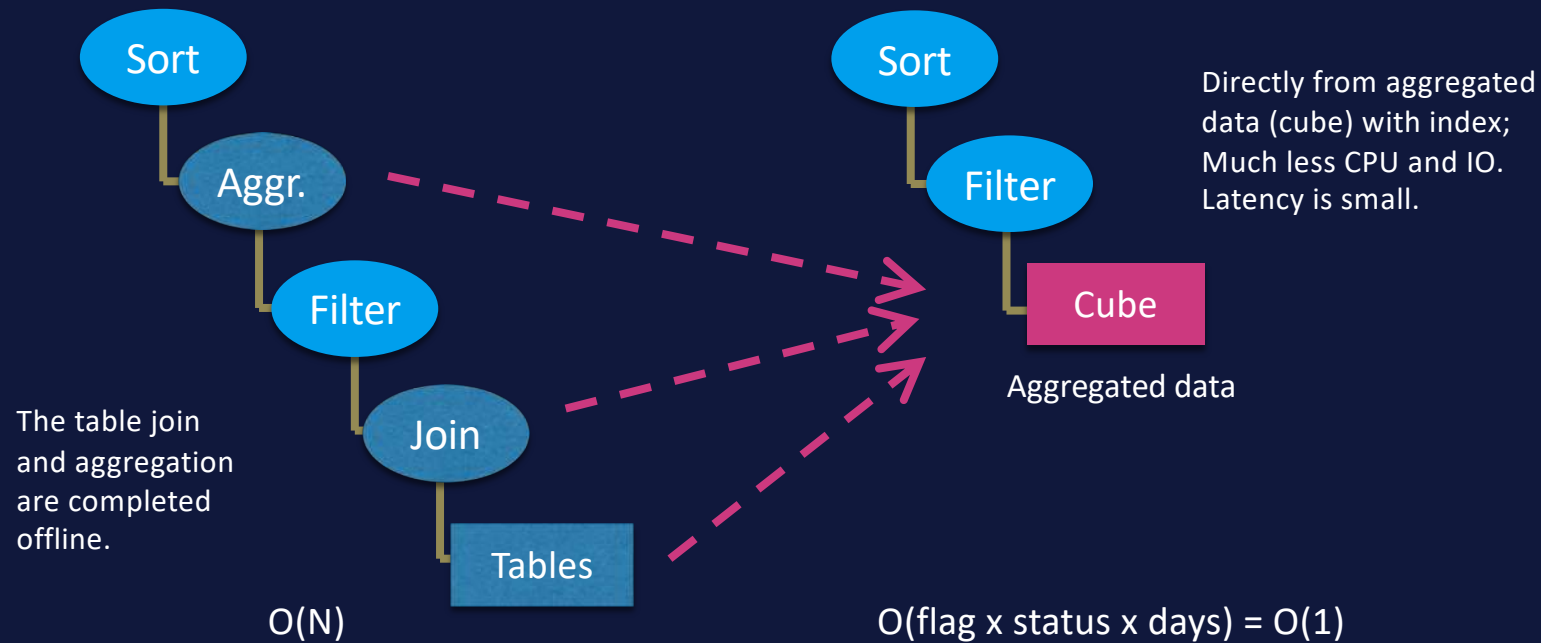
order by

```
l_returnflag,  
o_orderstatus;
```



SQL execution plan with Cube

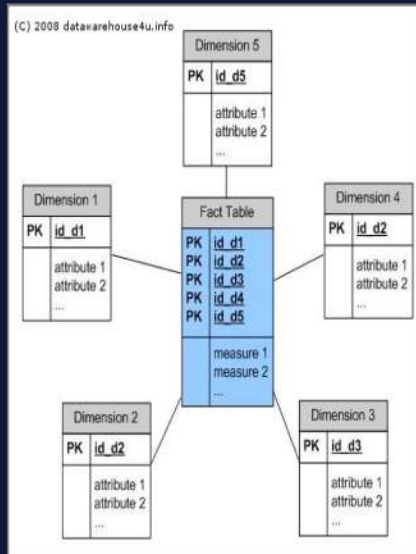
Cube technology speed up query performance with pre-calculation



Persistent the cube in HBase

Relational to Key Value store

UCLLOUD



Dimensions					Metrics		
D1	D2	D3	D4	D5	M1	M2	M2
a1	b1	c1	d1	e1	100	200	300
a2	b2	c2	d2	e2	200	400	600
a1	xxx	c1	yyy	e1	1	1	1

Logical
table for
cuboid
00010101

Dimensions			Metrics		
D1	D3	D5	M1	M2	M2
a1	c1	e1	101	201	301
a2	c2	e2	200	400	600

Row Key	
00010101	a1,c1,e1
00010101	a2,c2,e2

Metric
101,201,301
200,400,600

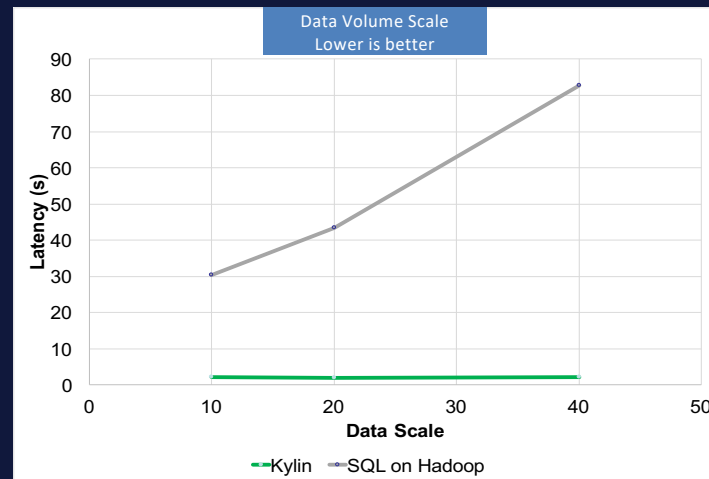
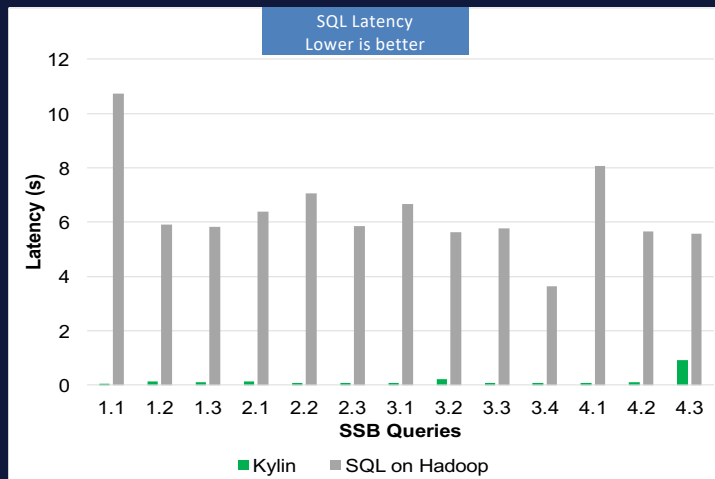
HBase
schema



High performance & High concurrency together



Sub-second latency on PB scale dataset



Star schema benchmark:
<http://www.cs.umb.edu/~poneil/StarSchemaB.PDF>



Traditional Data Warehouse Engineering

Cloud

Low Efficiency, Long Engineering Period

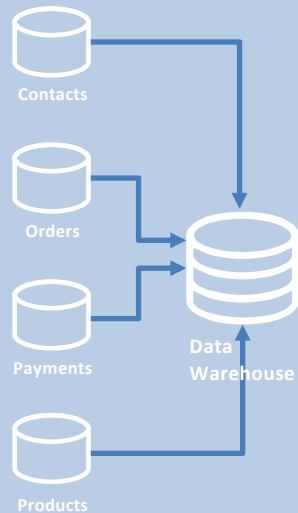


Data Warehouse Evolution

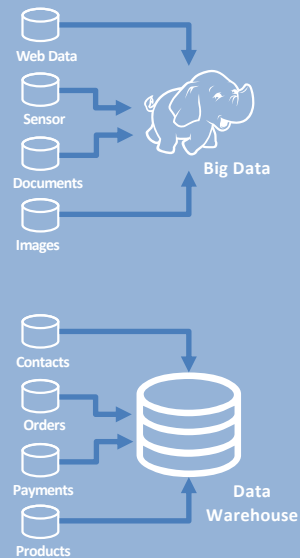
Data Mart



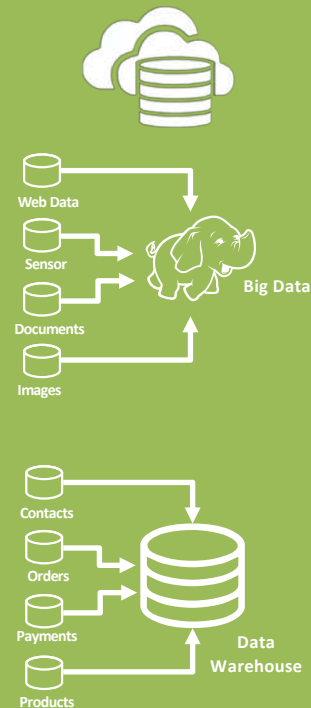
Data Warehouse



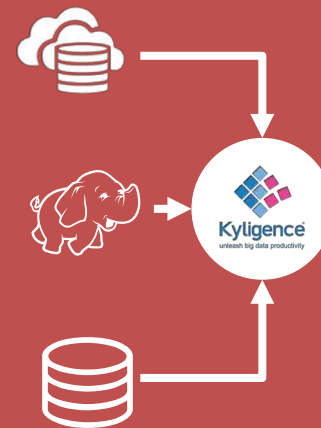
Data Lake



Cloud



Fusion & Intelligent DW



Kyligence = Kylin + Intelligence

Cloud

数据应用

客户分析	用户画像	设备检测
交易流水	资产负债	风险控制

智能 大数据分析

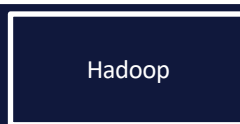


+



Kyligence Enterprise

数据中心



本地部署 + 云端部署 + 混合部署

AI：智能数据准备及分析，释放IT部门的人力资源和成本

Big Data：分布式极速分析引擎，亚秒级数据分析效率，释放业务洞察和数据变现能力

Cloud：混合部署，多源集成，完成新旧数据资产低成本而又高效的融合



AI Augmented Analytics Platform

Cloud



SQL
Query Log



Analytic
Behavior



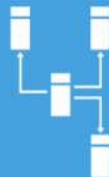
Data
Schema



Data
Profile



ML-based
Pattern
Discovery



Automated
Data
Enrichment &
Modeling



HDD SSD Memory

Optimized
multi-level
Storage Layer



Original

Intelligent
Query Push-
down &
Routing



Self-
Service BI



Real-time
Analysis



Data-as-a-
Service

On-Premises

Cloud

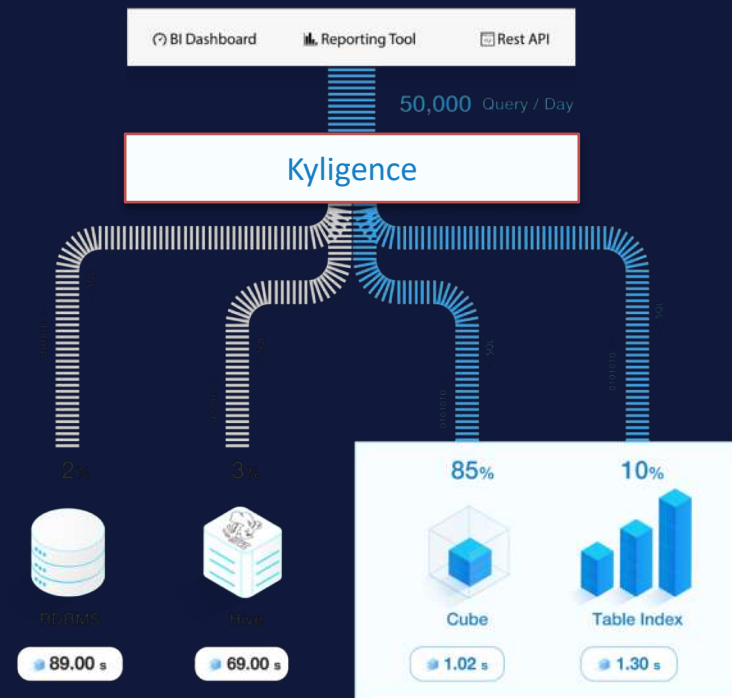
Container

Hybrid





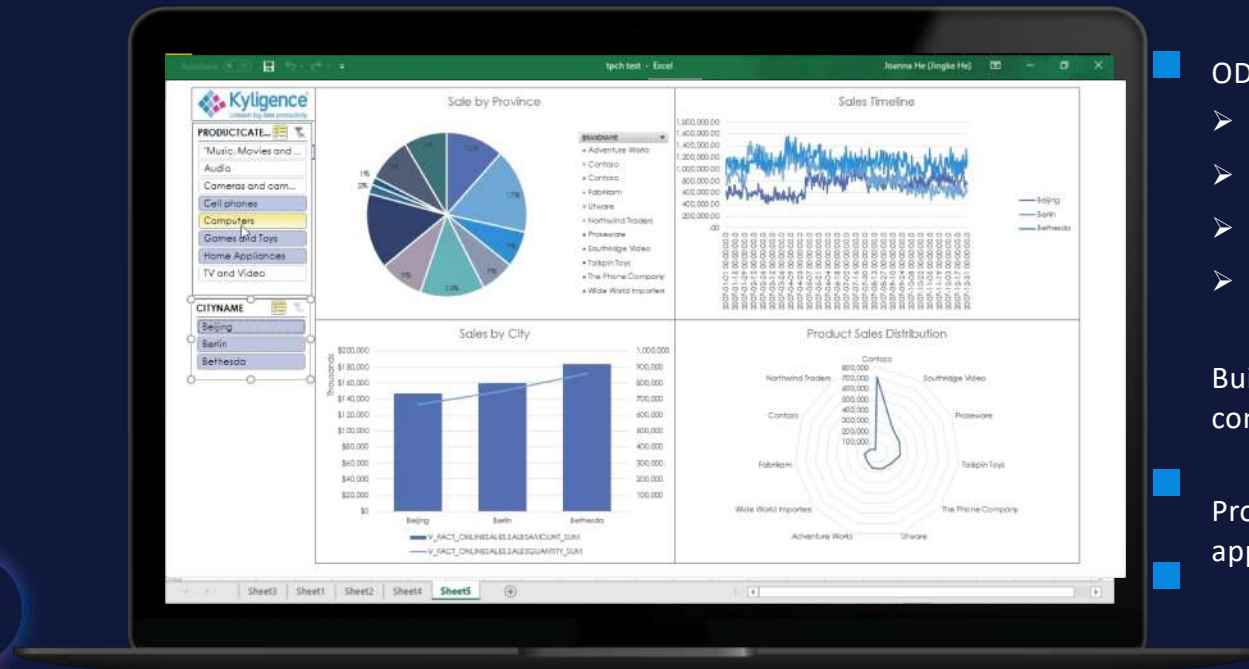
Modernized Architecture **UCLOUD**



Product Screenshot

Smooth Integration with BI and applications

ODBC



ODBC based BI Tool Connection

- Power BI/Excel
- Tableau/MicroStrategy/Qlik
- Cognos/BO/OBIEE
- Zeppelin/Superset

Built-in Data Source by PowerBI, others are coming

Provide REST APIs to integrate with existing applications in the organization



Kyligence Cloud = Apache Kylin + Cloud



Speed Up mission-critical analytics in the cloud



High Performance

Powered by Kyligence
Analytics Platform



One-click
provisioning

Deploy globally in 30
minutes



Auto Scaling

Scale cluster
automatically for
different workloads



Seamless
Integration

Connect to cloud data
sources
Enterprise ODBC
driver for BI



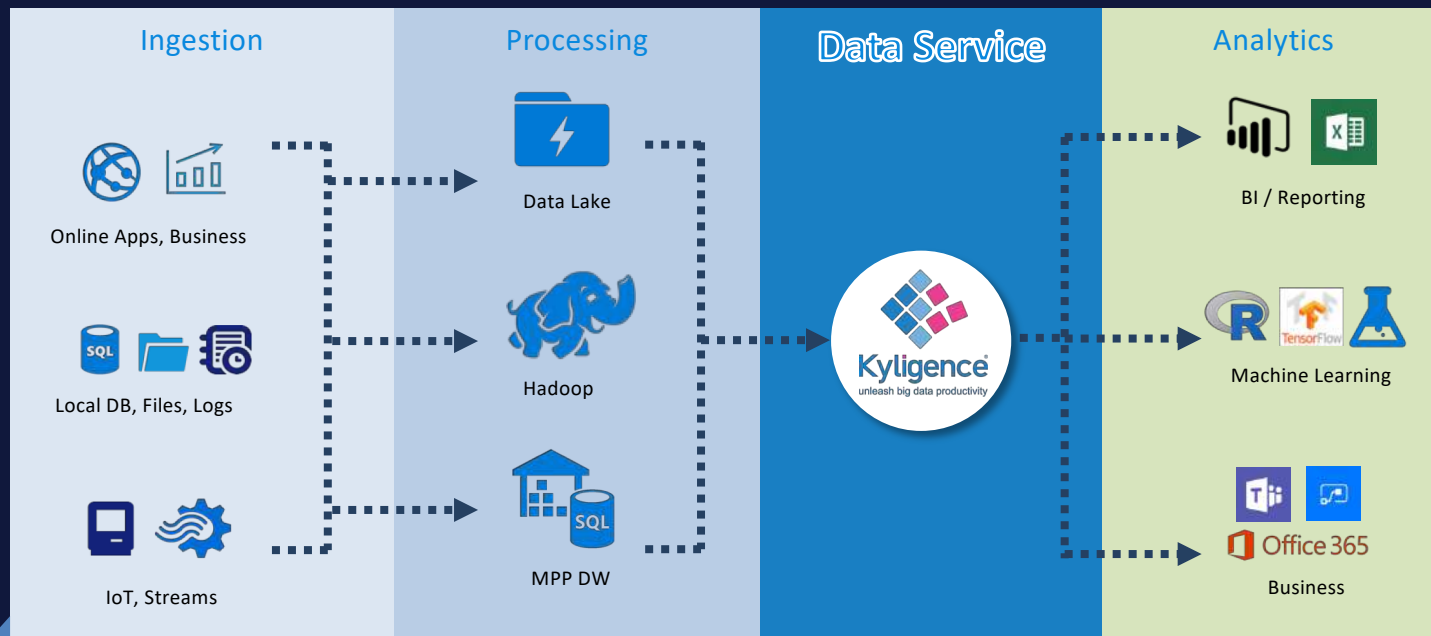
Intelligent Ops

Online diagnosis and
continuous
optimization



Cloud Native Big Data Architecture

Cloud

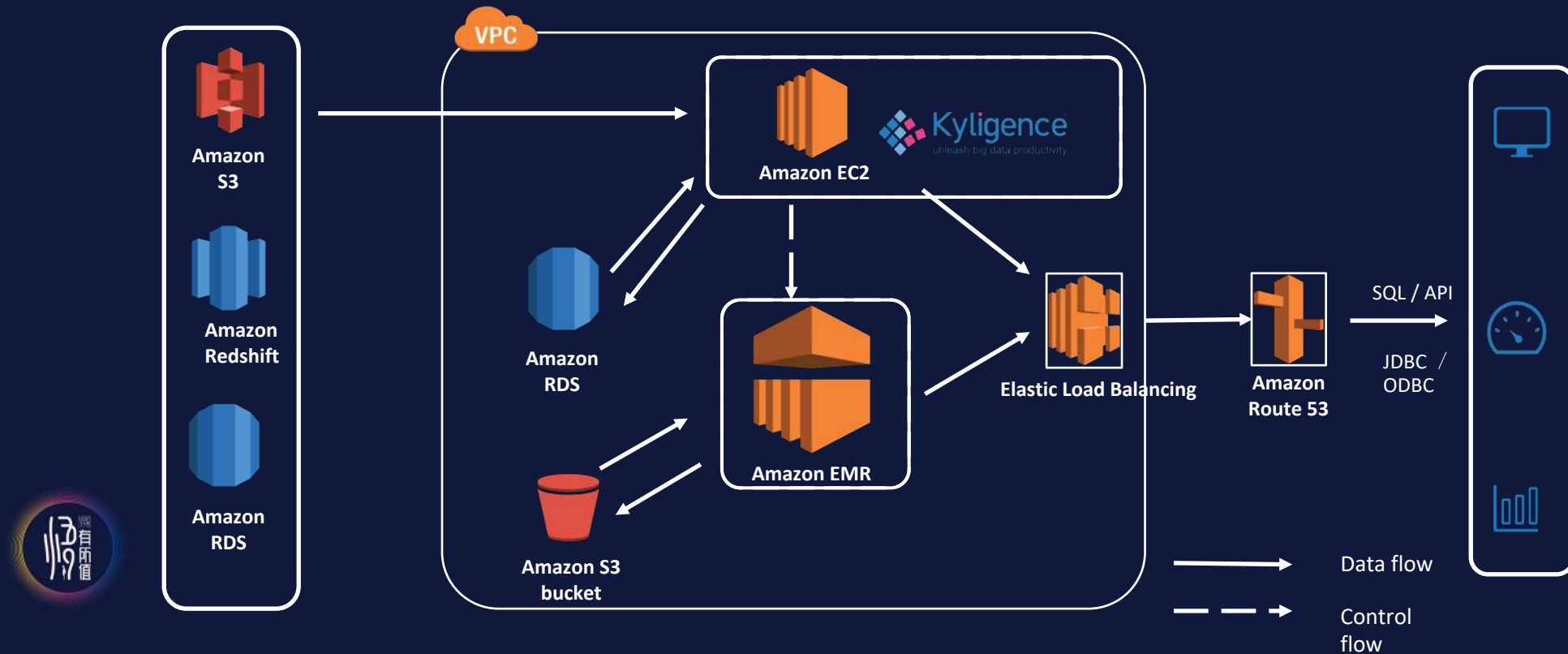


Self-Service & Interactive Analytics for Big Data in the Cloud



Kyligence on public Cloud

UCCLOUD



可靠：企业级数据安全性

Cloud



高可靠云存储

基于超可靠的云端存储，
保障数据的高可用和
永不丢失



账户级数据隔离

轻量级部署架构，
数据的存储、访问、计
算与查询都在客户自己
的云端账户内



企业级数据安全

支持项目、表、单元格
的多级访问控制，严格
控制数据访问安全



效率，30分钟全球一键部署

Cloud

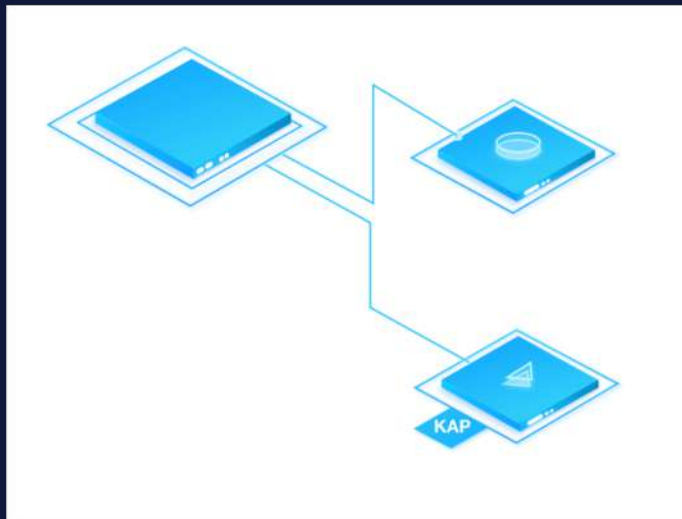
- 依托Cloud平台全球基础设施部署，支撑企业海内外业务发展，降低企业运维成本
- 快速与客户现有云端数据资源对接，让分析更加靠近数据



扩展：按需构建弹性伸缩

UCCLOUD

Head Node: D12 * 1



Worker Node: D12 * 4

- 构建时可根据任务负载动态增加节点
- 查询时可根据性能及并发要求增加节点

Edge Node: D4 * 1

- 用于安装Kylin服务
- 可根据查询并发要求增加节点

注：

- 用户可修改虚拟机型号和节点数，但整体计算能力不应低于以上配置
- 以上为最小配置，当数据量、性能、并发等要求过高，应增加节点数或提高配置以提升集群计算能力



按需使用的自动成本管理

Cloud

以成本和性能为目标的自动伸缩，节省成本多至 70% ！

— 查询集群

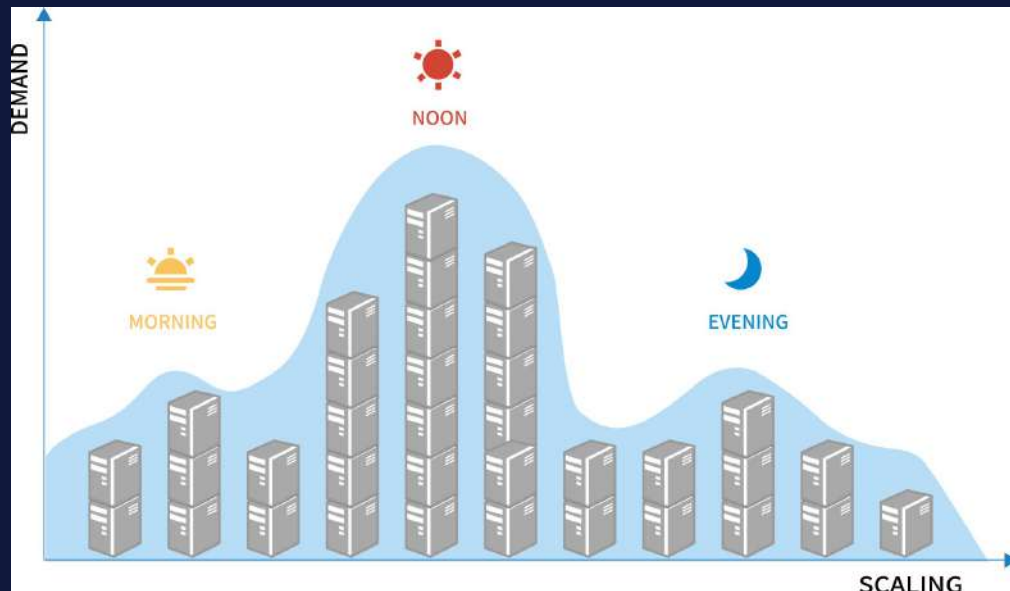
— 追踪性能指标 CPU / Mem / Network

— 按需自动升缩集群

— 构建集群

— 按需创建集群

— 用完销毁



Kyligence = Kylin + Intelligence

U CLOUD

- Founded by the team who created **Apache Kylin**, in 2016
- From eBay, Microsoft, IBM, Cisco, Teradata
- **Apache Kylin: Leading open source OLAP on Big Data**
- Funding by leading VCs: Redpoint Ventures, Cisco, CBC Capital, Shunwei Capital and Fidelity International
- Dual-Headquarters: **Shanghai China**, **Silicon Valley US**
- **CRN Top 10 Big Data Startups 2018**



Kyligence Open Ecosystem



No.1

No.1 open source
OLAP on Hadoop



Community

Leading the Open Source
community and culture

更多技术干货分享



更多深度技术交流

