

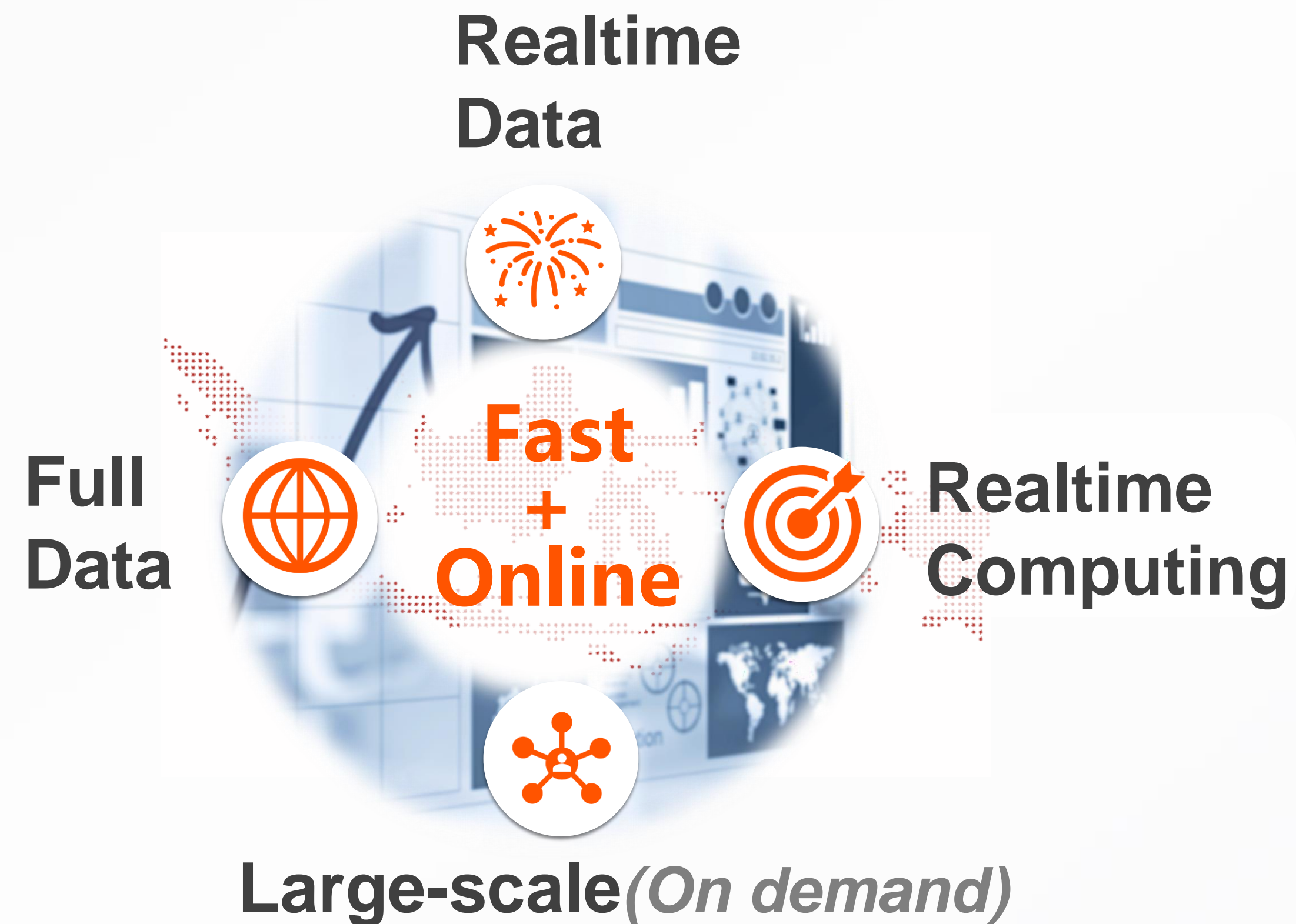
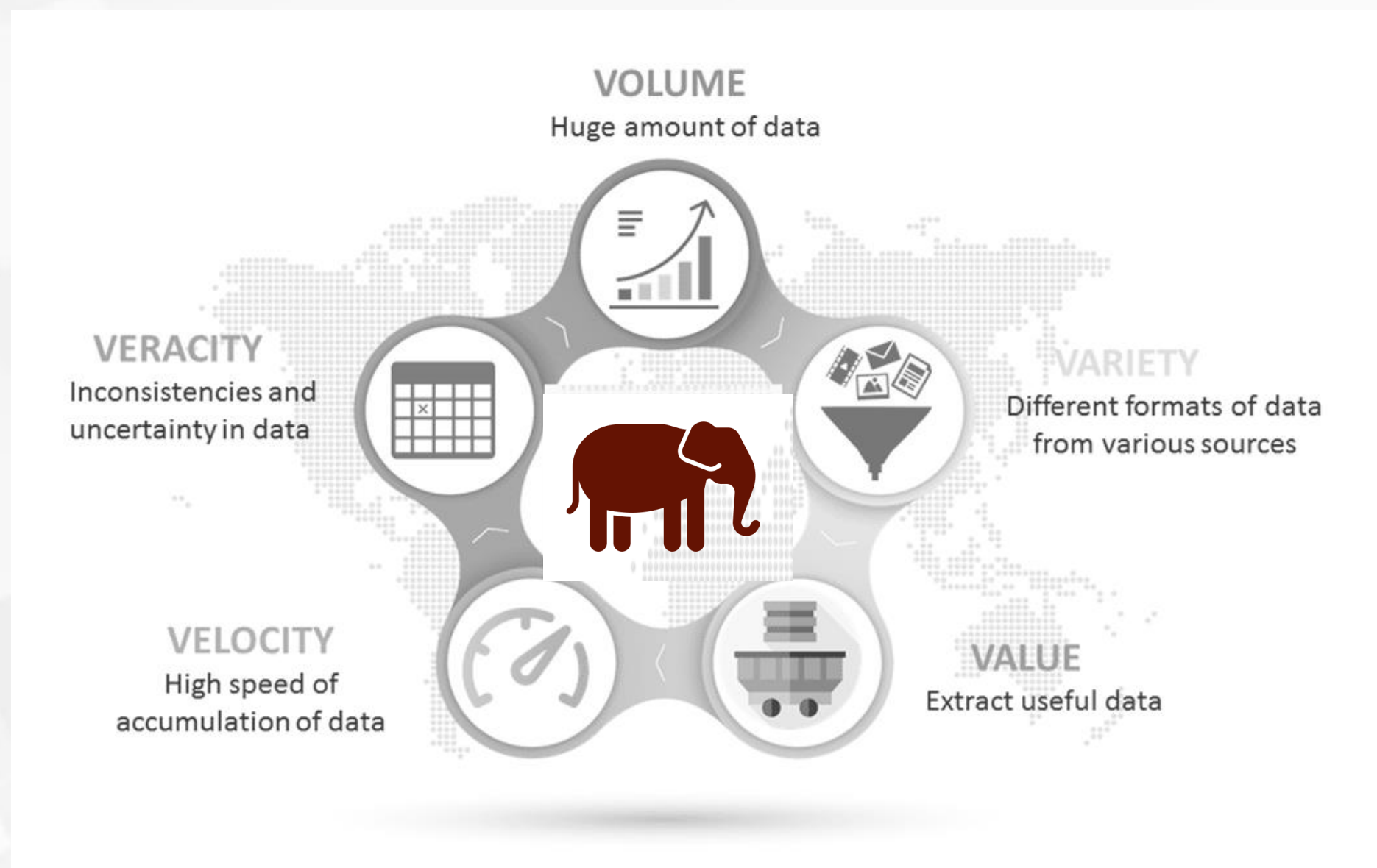
南仙

阿里云智能高级技术专家

AnalyticDB for MySQL: PB级云数仓 核心技术和场景解析

潘岳（南仙）

BigData已是昨日黄花，FastData才是未来



股价41% ↓



寻求买家...



AnalyticDB for MySQL 介绍：PB级实时数据仓库



极致性价比驱动数据价值在线化

- ✓ 千亿 / 万亿 多表关联分析 毫秒级
- ✓ 异构加速的MPP+DAG 融合分布式执行引擎
- ✓ 存储计算分离架构，智能化行列混存实现极速检索

10X

vs Spark

世界级认可

TPC® 1st

Gartner®

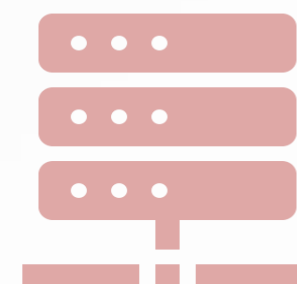
FORRESTER®

Very
Large
Data
Bases



兼容 & 超越 MySQL

- ✓ 全面兼容MySQL生态
 - ✓ ANSI SQL:2013 复杂分析支持
- TPC-DS性能全球第一**



云原生 实时按需极致弹性

- ✓ 存储从GB至100PB
- ✓ 计算节点从3台到2000台
- ✓ 混合负载



非结构化与结构化融合分析

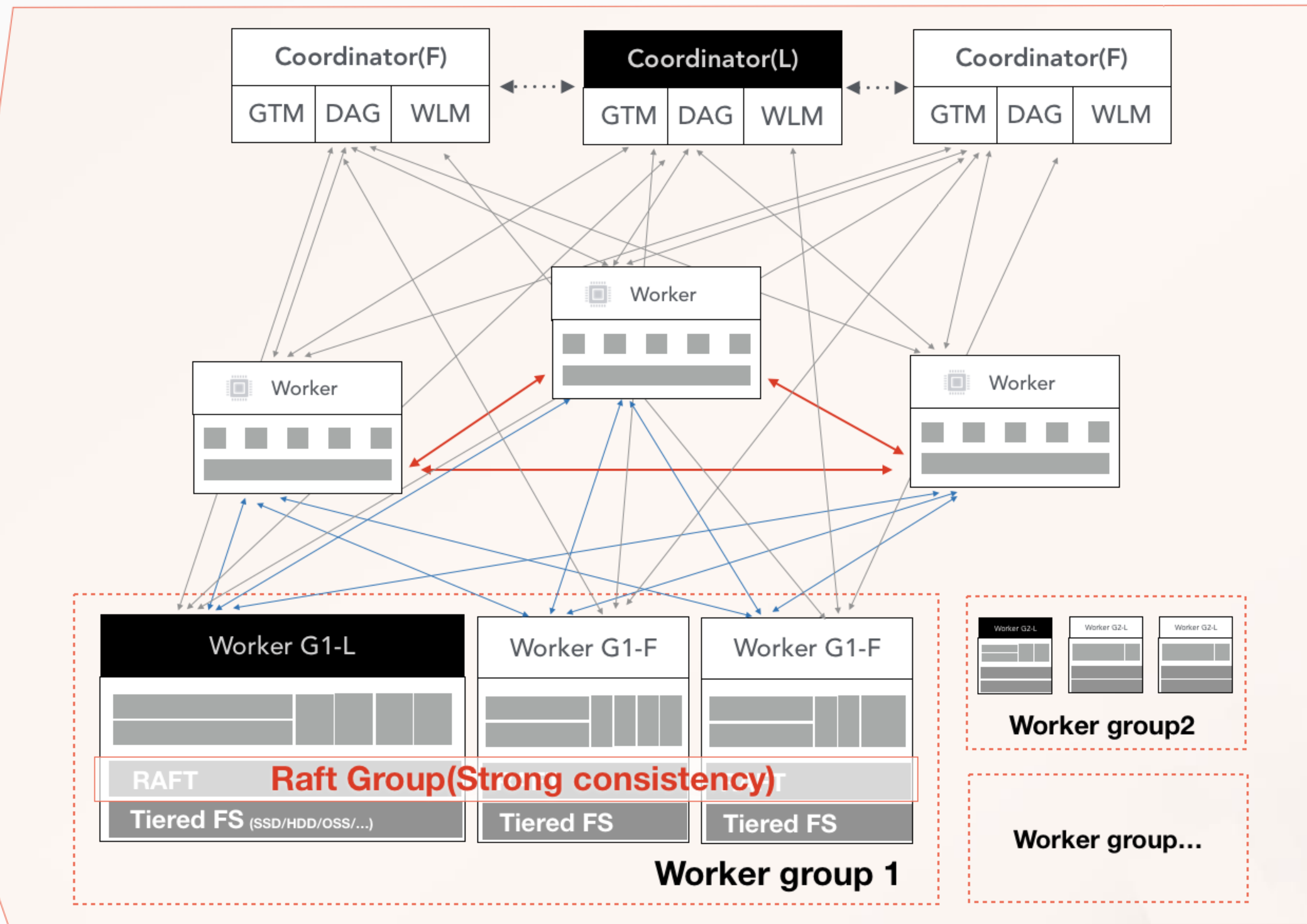
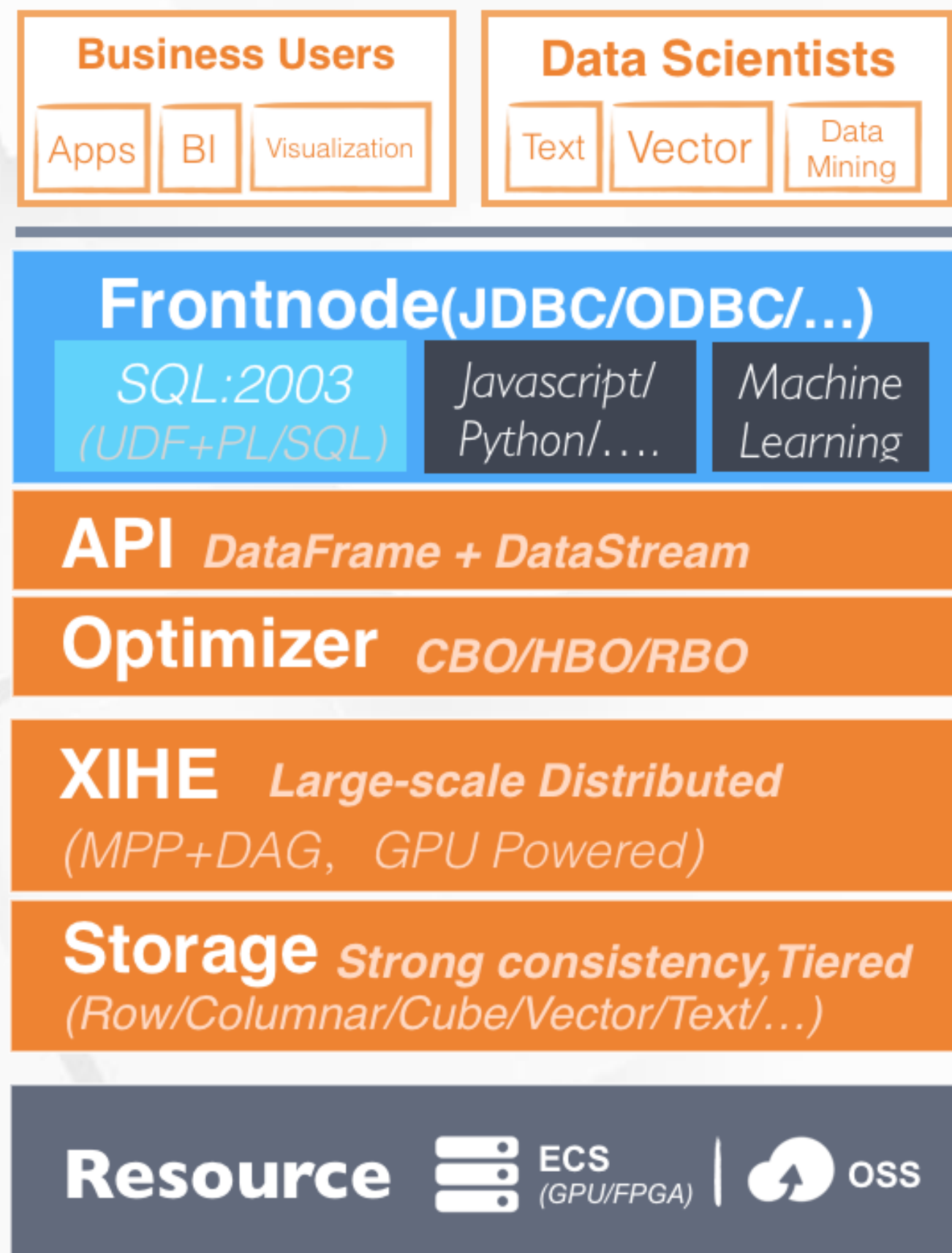
- ✓ 全面支持非结构化数据的存储、检索、碰撞
比对、融合分析
- ✓ 5+ 向量数据类型，5+向量检索算法



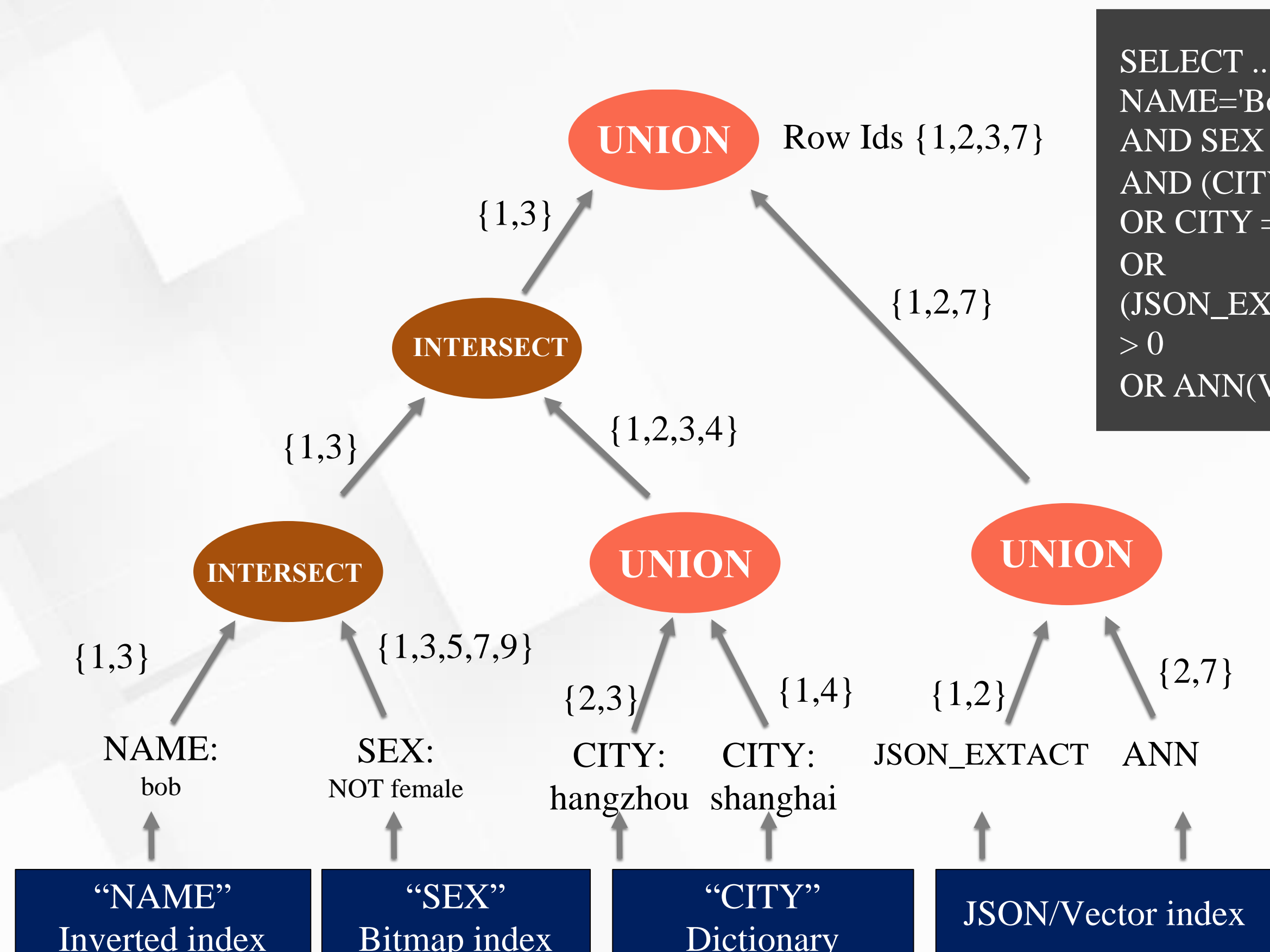
完备的企业级特性

- ✓ 备份/恢复/Flashback/回收站
- ✓ 审计/白名单/自建账号/VPC
- ✓ 跨AZ/跨Region(On-going)

AnalyticDB核心技术：分层存储+存储计算分离架构 带来极致弹性和开放性



AnalyticDB核心技术-智能的“行列混存+全索引”带来极致性能



```
SELECT ... WHERE (  
  NAME='Bob'  
  AND SEX !='female'  
  AND (CITY = 'Hangzhou'  
  OR CITY = 'Shanghai'))  
  OR  
  (JSON_EXTACT(ATTR,'time')  
  > 0  
  OR ANN(VEC, [1,1,1,1], 2))
```

兼备 ✓ 明细查询的所有列
✓ 多维分析的列裁剪和压缩

任意条件组合过滤检索
大幅筛选候选集

Detail Meta File for One Column

Header	
Count	NULL Count
Distinct Count	Sum
Max	Min
Dictionary Offset	Dictionary Length
Block Map Offset	Block Map Length
Dictionary	
Block Entry ₁	
Block Entry ₂	
⋮	

Detail File

Data of One Row			
Column ₁	1	10	6 ...
Column ₂	Bob	Jack	Ann ...
⋮			
Column ₁	23	68	2 ...
Column ₂	Tom	Rio	Lee ...
⋮ Data block			
⋮			

Column File for complex type

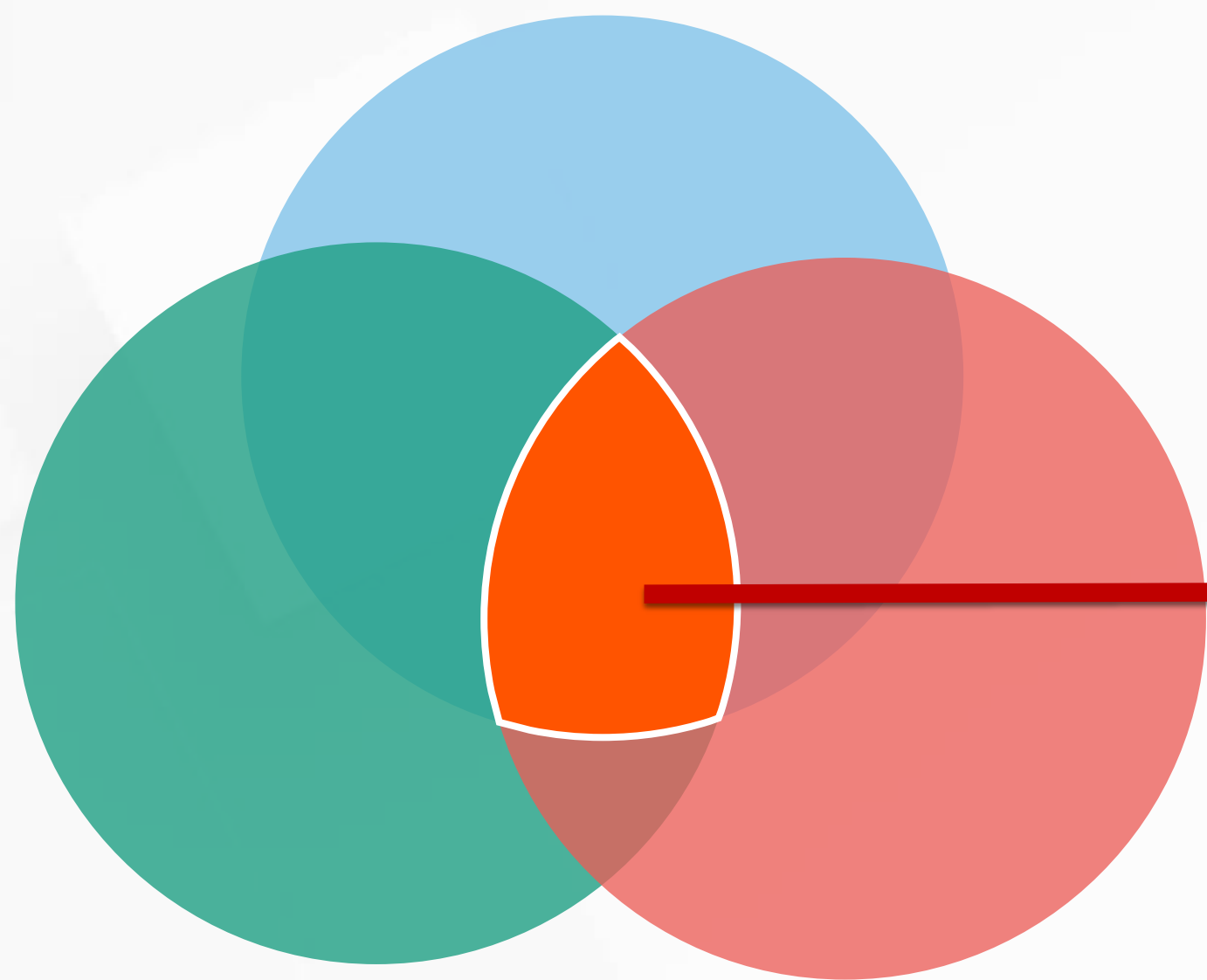
8K fixed storage

8K	8K	8K
8K	8K	8K

AnalyticDB核心技术：一个系统一套存储兼顾多种场景

多维分析

- ✓ 任意列Join
- ✓ 复杂长计算任务、ETL

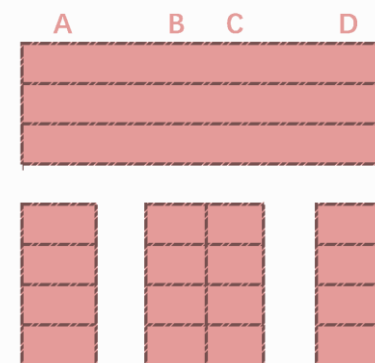


明细查询

- ✓ 1000+ 列 超宽表
- ✓ 半结构化、大字段 (JSON/ARRAY等)

实时查询和写入

- ✓ 实时写入和更新
- ✓ 600+ 万记录/秒 写入
- ✓ 10000+ QPS



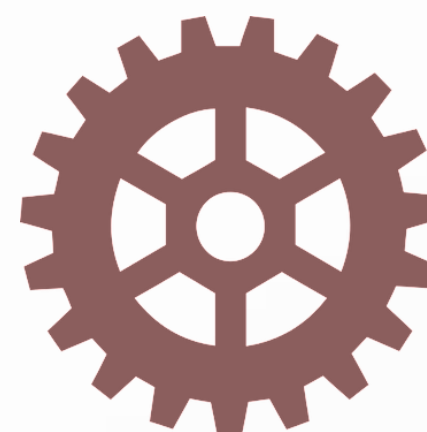
行列混存 (Hybrid Store)

行列混存、Block过滤
智能全索引



混合负载管理 (Hybrid Workload)

CPU/Mem/Net/IO
高并发低延迟 / 复杂ETL



融合计算引擎 (Hybrid MPP+DAG)

大规模分布式执行
Autonomous Optimizer



营销平台

10000+
表数目

100T+
数据量

10+
百亿大表
Join数量

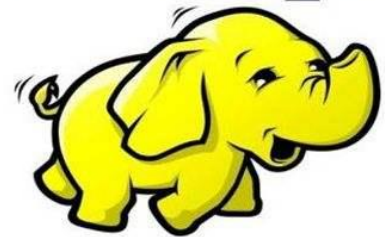
任意纬度
筛选

120万+ 条/秒
峰值写入

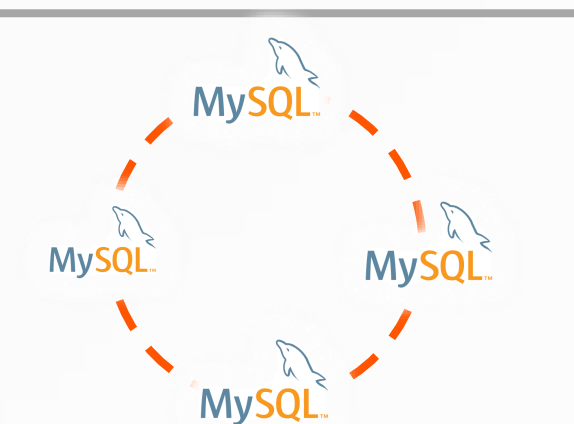
50+ QPS
10张以上百亿大表Join

5000+ QPS
明细查询

hadoop



kafka



MySQL Sharding



Oracle RAC

Datworks

100亿/天

Datworks

600万/秒

DTS

DTS

10亿/天



AnalyticDB

1000+ QPS

APP

DataV

QuickBI

Dataworks
Kettle

....

某云客户

200+ node / 5PB

10+ App



价值关键词

平台统一

无需维护离线在线混合复杂架构

性能大幅提升

提升3-10x

分析实时化

T+1 → 实时

兼容性不错

迁移和接入比较快

Data Lake Analytics: 全域数据、开放分析



无服务器化，扩容能力极强，立即可用的极低成本数仓

- ✓ 智能的存储感知和分层缓存设计 **0**预留存储成本
- ✓ 智能的资源弹性调度和计算引擎感知设计 支持快速扩容至**2000** worker, **0**预留计算成本



丰富的生态兼容

- ✓ 全面兼容MySQL 生态
(BI / ETL / 可视化 / 编程语言)
- ✓ 全面兼容Spark生态
(GraphX/Streaming/ML/SparkR)



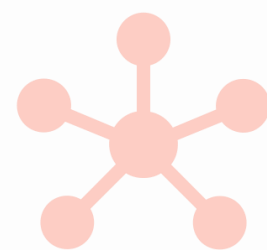
统一元数据，双计算引擎按需切换

- ✓ 极速分析引擎XIHU TPC-H、TPC-DS 领先 **3x**
- ✓ Spark引擎 支持分布式内存池加速 **2x** (On-going)



智能探测,开箱即用

- ✓ Metadata 自动识别
- ✓ SQL查询时动态生成Metadata
- ✓ 智能查询下推

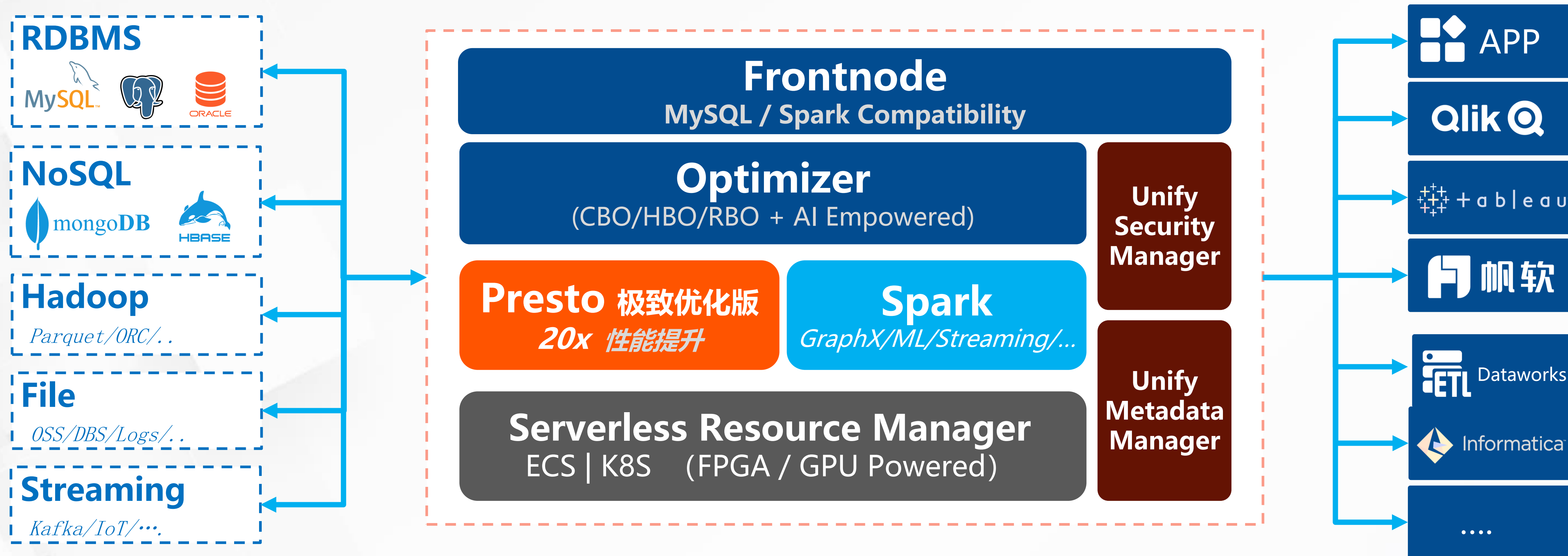


无需迁移即可统一分析全域数据

- ✓ RDMS 5+ (MySQL/Oracle/...)
- ✓ Hadoop 4+ (ORC/Parquet/...)
- ✓ NoSQL 3+ (MongoDB/...)
- ✓ OSS/DBS/SLS/...

15+
数据源

Data Lake Analytics: 全域数据、全局开放分析



价值关键词

统一数据视图

云上所有数据源组合分析查询

实时化

提升10~20x

业务接入极快

兼容度高，一键自动建仓

建仓成本极低

无数据搬迁，按需付费

客户案例1：移动运营APP解决方案

无他相机



业务诉求

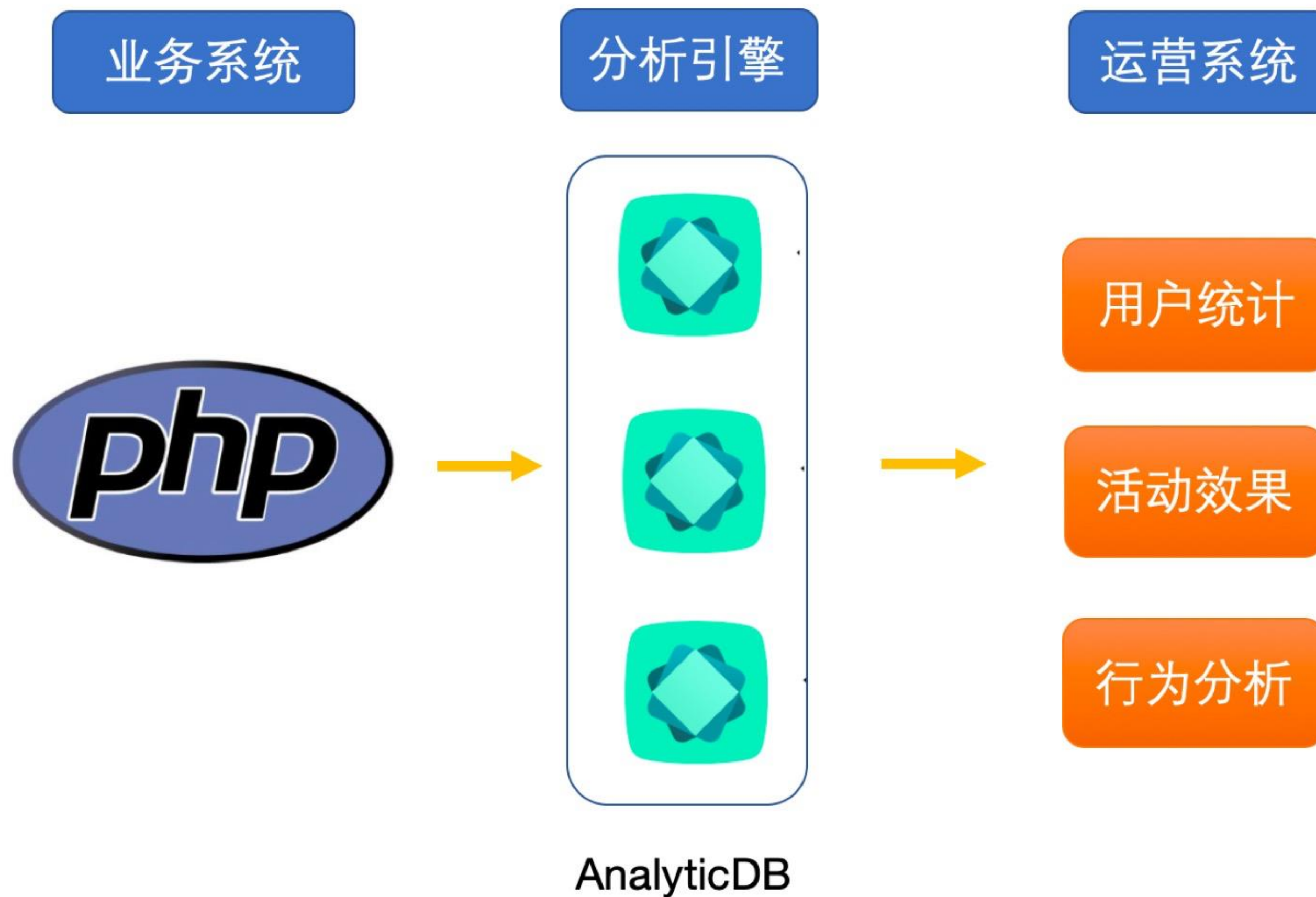
100亿数据，实时入库更新，快速复杂分析

历史问题

MySQL -> MongoDB, 分析性能无法接受

改造收益

实时入库，复杂分析40分钟级到秒级别



客户案例2：物流行业实时数仓



领跑跨境电商物流

➤ 极低成本

投入有限人力，1.5个月完成实时数据平台的基础建设。

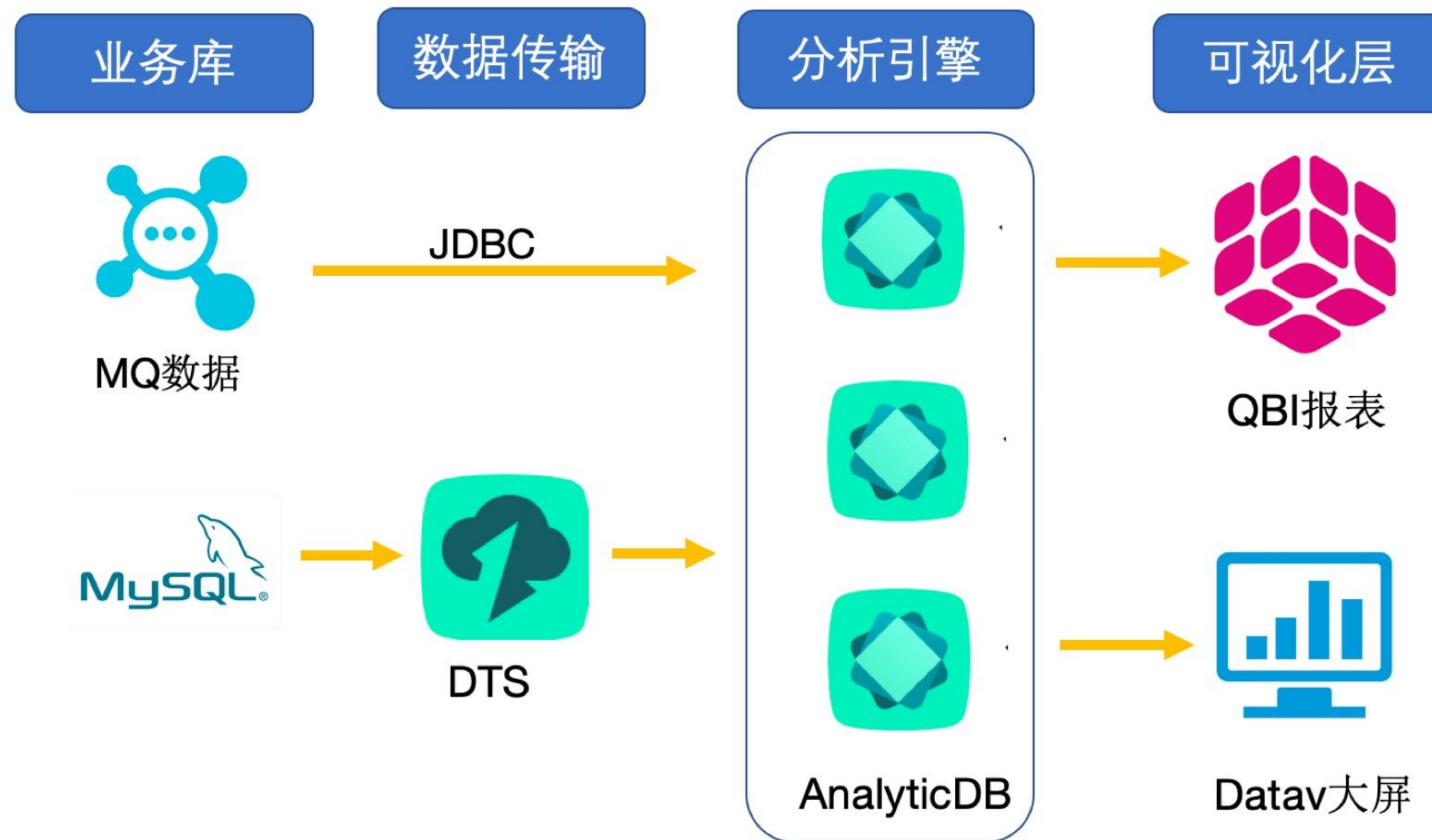
整个方案较其他备选方案，成本降低40%

➤ 极高性能

海量数据多表join，毫秒级返回

➤ 极高效率

从数据接入到数据化上线，仅历时1个半月



AnalyticDB for MySQL: 客户遍布各行各业

泛互联网



新零售



极快 高兼容
多场景

数字政府



金融



AnalyticDB for MySQL 3.0版本：下一代云原生OLAP产品

更易用



- ✓ 支持最多256个DB
- ✓ 兼容性大幅提升 (99.99%)
- ✓ 数据写入立即可见

更高性能



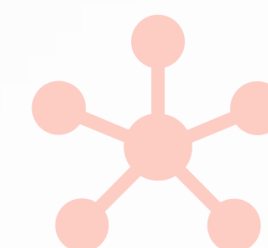
- ✓ 实时写入性能 **1.5x**
- ✓ 查询性能提升 **40%**
- ✓ 批量导入 **TB/小时**

更弹性



- ✓ 磁盘空间弹性伸缩
- ✓ 节点数目任意库扩容
- ✓ 计算节点弹性扩展 (On-Going)

更可靠



- ✓ 完备的权限体系
- ✓ 数据三副本工业安全
- ✓ 完备的备份恢复能力

当前火热公测中，将于7.31商业化，欢迎体验！



阿里云开发者社区

扫码加入社群
与志同道合的码友一起
Code Up



阿里云数据库微信公众号

谢谢！