



# 李飞飞

阿里巴巴集团副总裁，高级研究员，  
阿里云智能数据库总负责人



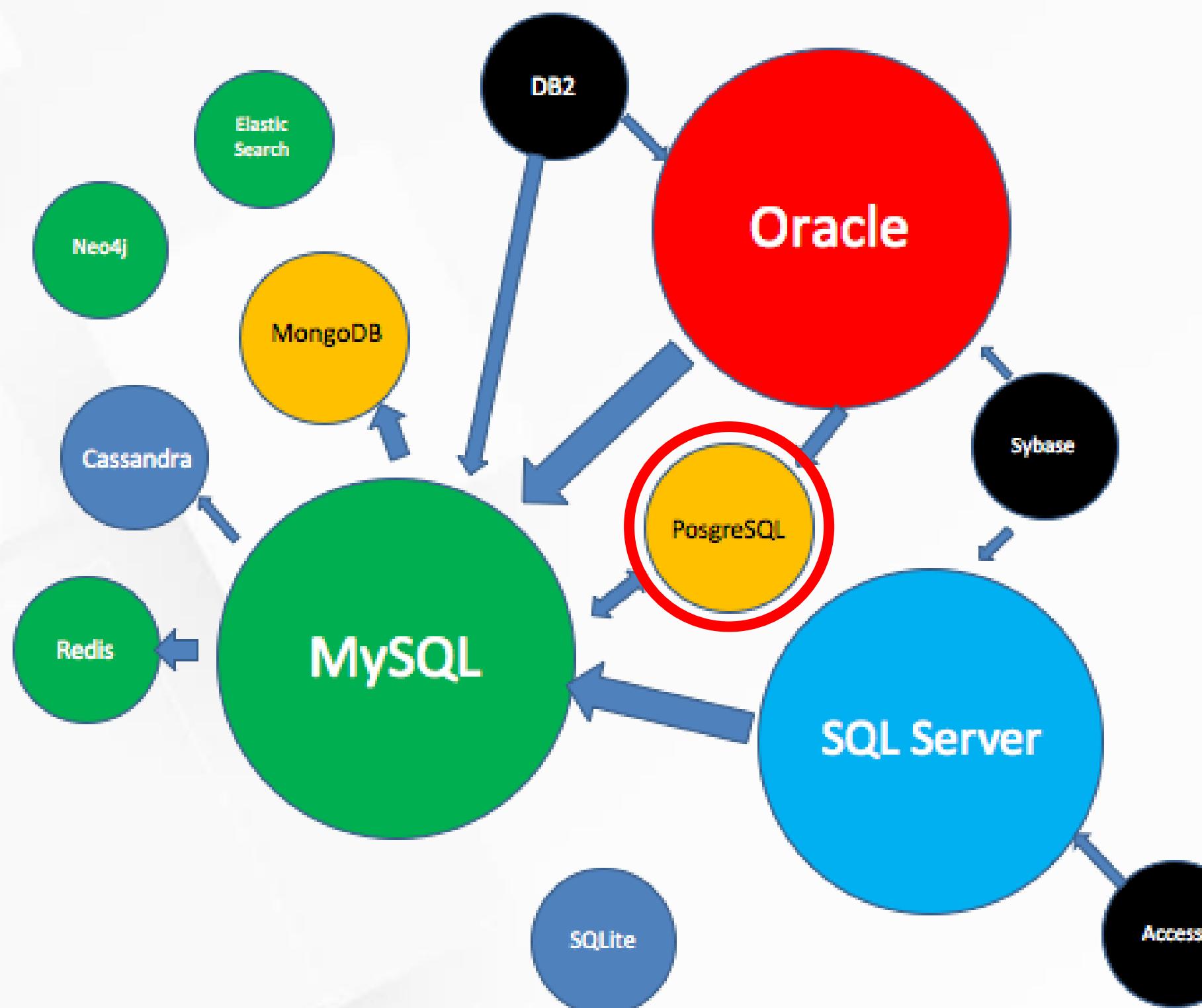
2019阿里云峰会·上海  
开发者大会  
DEVELOPER CONFERENCE

# 下一代云原生数据库技术趋势

李飞飞

阿里巴巴集团副总裁  
数据库产品事业部总裁  
达摩院数据库首席科学家, ACM杰出科学家

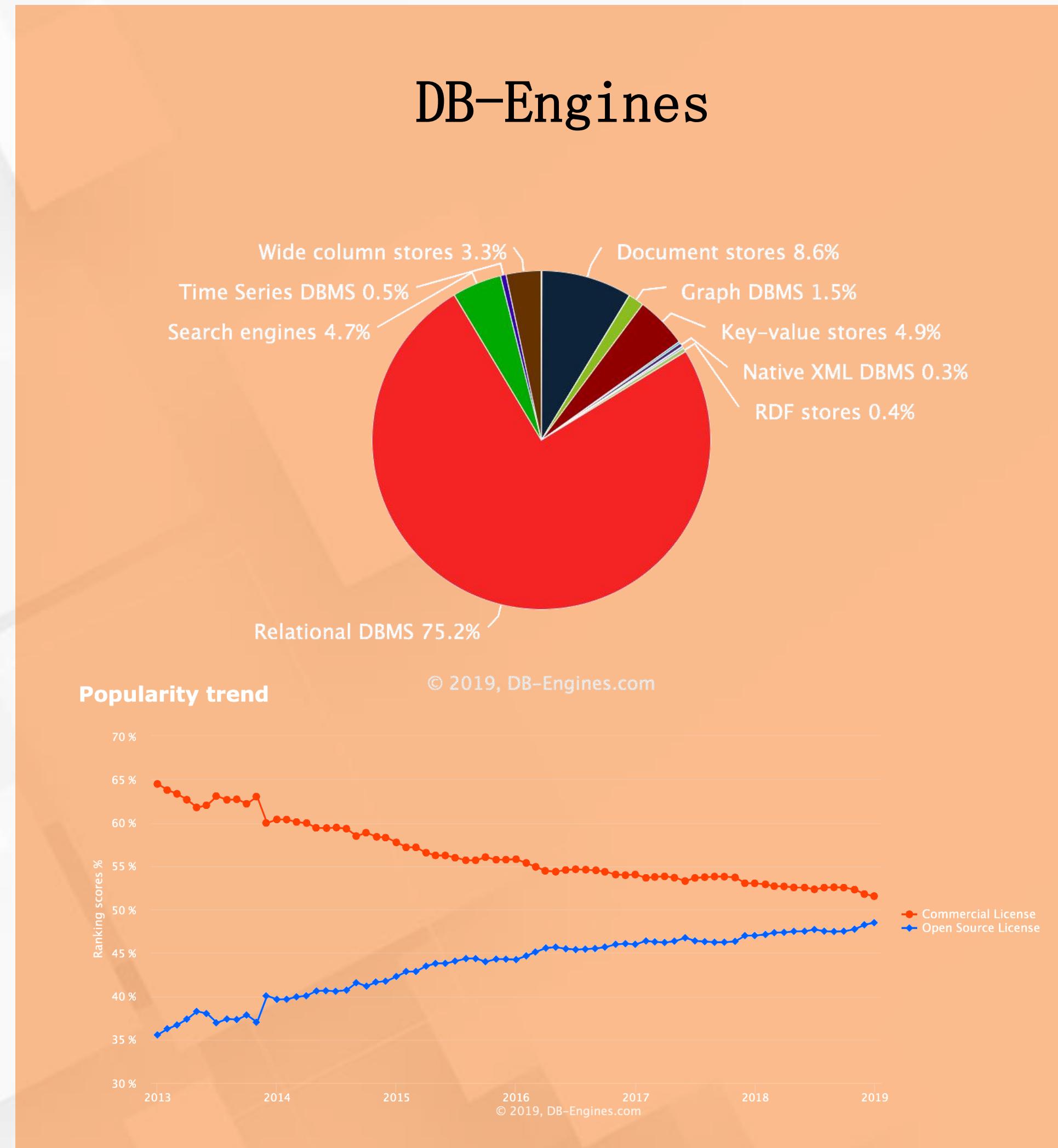
# 不同种类的数据库系统



TOP 30 Database Products

Rank	DBMS	Database Model	Score		
			Jan 2019	Dec 2018	Jan 2018
1.	1. Oracle	Relational DBMS	1268.84	-14.39	-73.11
2.	2. MySQL	Relational DBMS	1154.27	-6.98	-145.44
3.	3. Microsoft SQL Server	Relational DBMS	1040.26	-0.08	-107.81
4.	4. PostgreSQL	Relational DBMS	466.11	+5.48	+79.93
5.	5. MongoDB	Document store	387.18	+8.57	+56.24
6.	6. IBM Db2	Relational DBMS	179.85	-0.90	-10.43
7.	7. Redis	Key-value store	149.01	+2.19	+25.88
8.	8. Elasticsearch	Search engine	143.44	-1.26	+20.89
9.	9. Microsoft Access	Relational DBMS	141.62	+2.10	+14.92
10.	10. SQLite	Relational DBMS	126.80	+3.78	+12.54
11.	11. Cassandra	Wide column store	122.98	+1.17	-0.89
12.	12. Splunk	Search engine	81.43	-0.76	+17.42
13.	14. MariaDB	Relational DBMS	78.82	+1.56	+20.52
14.	13. Teradata	Relational DBMS	76.19	-2.98	+3.56
15.	15. Hive	Relational DBMS	69.91	+2.52	+14.42
16.	16. Solr	Search engine	61.48	+0.13	-2.89
17.	17. HBase	Wide column store	60.39	+0.38	-1.24
18.	18. FileMaker	Relational DBMS	57.15	+0.50	+1.95
19.	19. SAP HANA	Relational DBMS	56.64	+0.33	+10.47
20.	21. Amazon DynamoDB	Multi-model	55.09	+0.80	+17.17
21.	20. SAP Adaptive Server	Relational DBMS	55.04	-0.78	-10.42
22.	22. Neo4j	Graph DBMS	46.80	+1.24	+6.83
23.	23. Couchbase	Document store	34.59	-1.00	+3.16
24.	24. Memcached	Key-value store	29.54	-0.07	+1.39
25.	25. Microsoft Azure SQL Database	Relational DBMS	27.20	+0.17	+4.17
26.	24. Informix	Relational DBMS	26.76	-0.17	-1.92
27.	27. Microsoft Azure Cosmos DB	Multi-model	24.38	+1.01	+9.53
28.	28. Vertica	Relational DBMS	21.78	+0.40	+1.47
29.	29. Firebird	Relational DBMS	20.18	+0.13	+2.79
30.	33. Amazon Redshift	Relational DBMS	20.05	+0.05	+6.99

# DB-Engines (343 in total)



**The top 5 commercial systems, January 2019**

Rank	System	Score	Overall Rank
1.	Oracle	1269	1.
2.	Microsoft SQL Server	1040	3.
3.	IBM Db2	180	6.
4.	Microsoft Access	142	9.
5.	Splunk	81	12.

**The top 5 open source systems, January 2019**

Rank	System	Score	Overall Rank
1.	MySQL	1154	2.
2.	PostgreSQL	466	4.
3.	MongoDB	387	5.
4.	Redis	149	7.
5.	Elasticsearch	143	8.

# Top 云厂商数据库产品



# 数据库: 云上应用关键一环

“The real battle will be in databases”

Source:

- “How Amazon Web Services aims to win cloud computing's next big battle” SiliconANGLE
- “AWS to Oracle: Now it's our turn and we got next”ZDNet

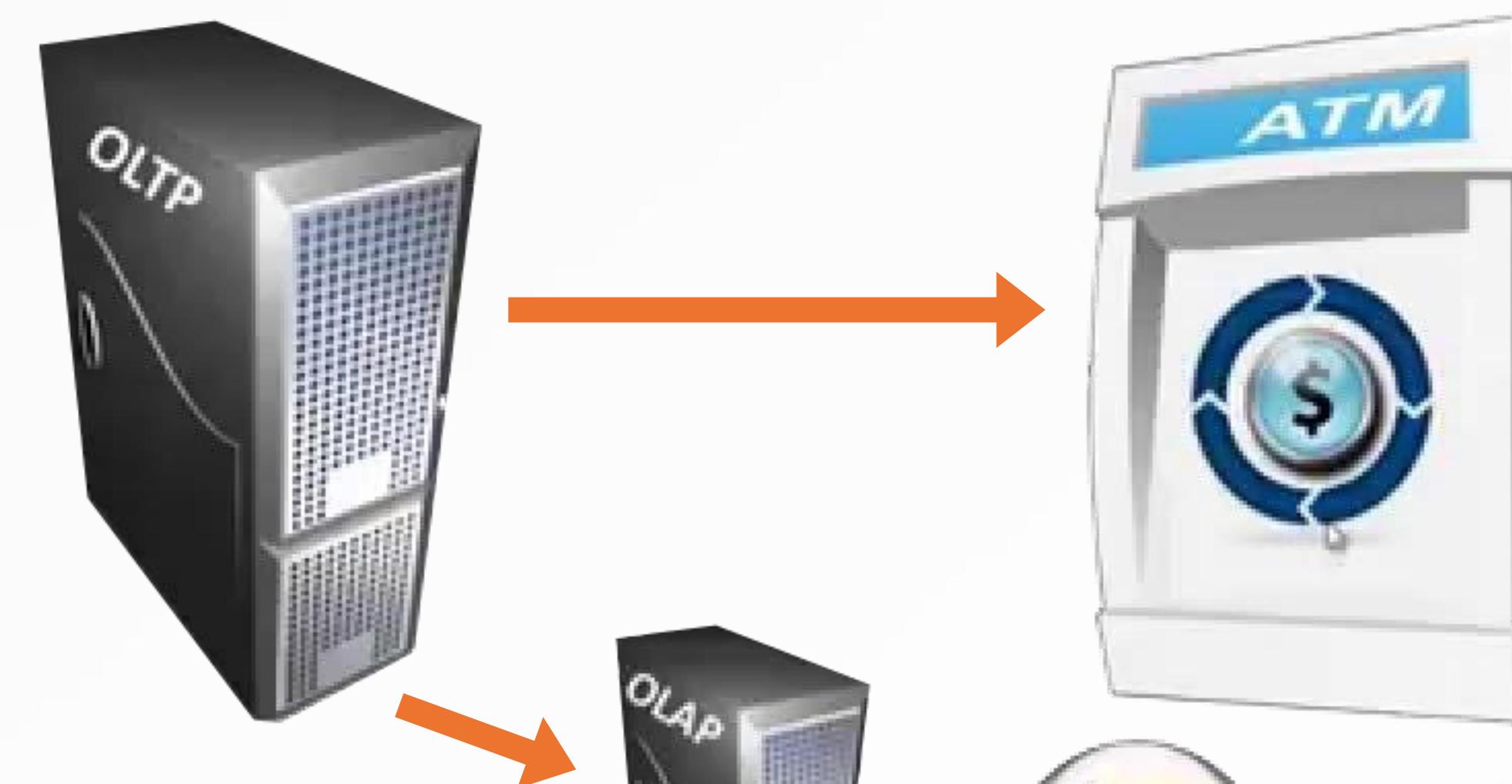


数据的产生, 存储, 和消费

# 数据库发展-业务视角

## ◎ 事务性数据库 (OLTP)

- › 结构化数据库
- › 银行转账记账, 淘宝下单订单商品  
库存管理
- › 挑战: 高并发, 高可用, 高性能下的  
正确性和数据一致性



## ◎ 分析型数据库 (OLAP)

- › 海量数据, 数据类型复杂
- › 分析条件复杂
- › 深度智能化分析
- › 挑战: 高性能, 分析深度, 与TP数据库的联动联动

## ◎ 数据库服务+管理类工具

- › 数据传输, 数据备份, 数据库管理等

# 数据库系统演进

结构化数据



RDBMS  
[ SQL+OLTP ]

结构化数据



Data warehouse  
Data Cube  
[ ETL+OLAP ]

Heterogeneous Data

Structure Data

Graph

Time Series

Vector

Spatial Data

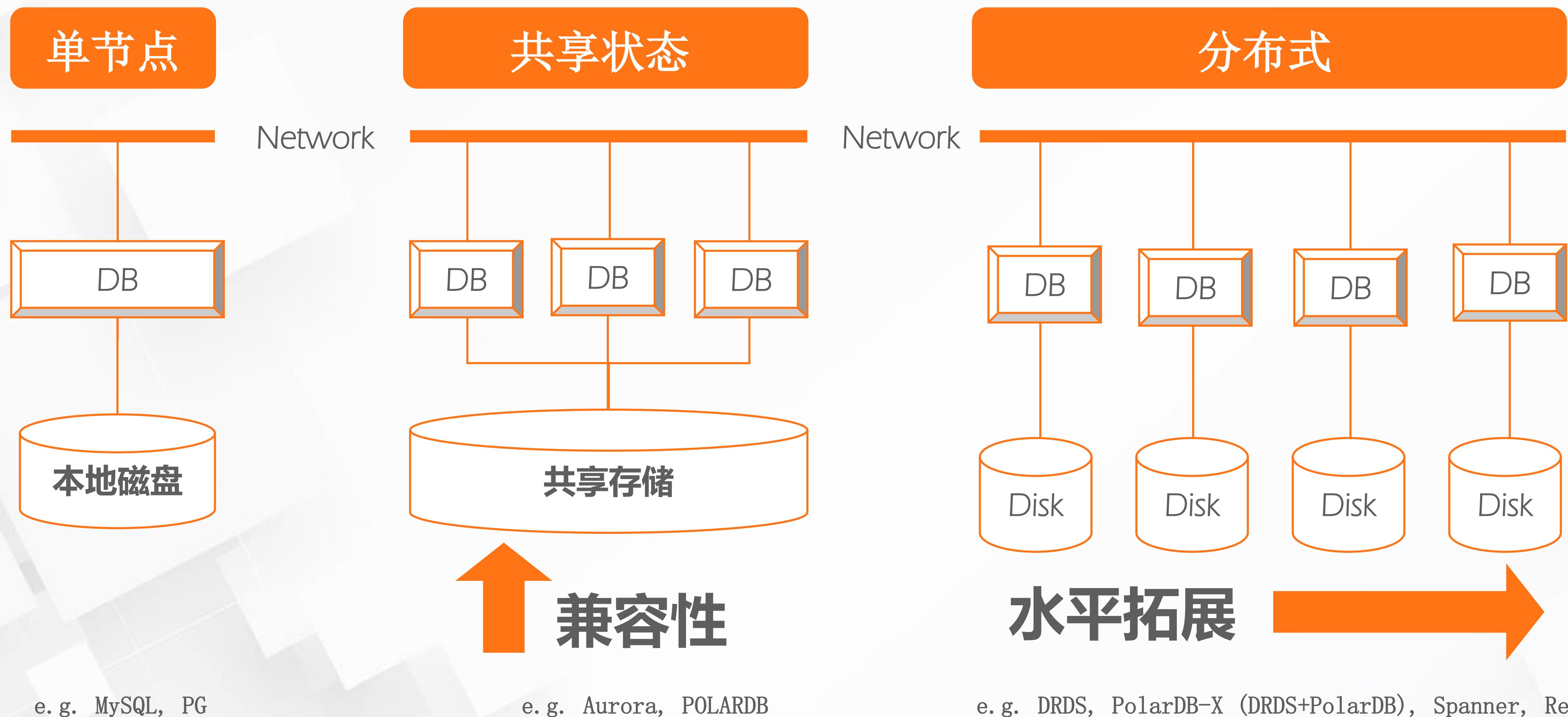
Text

RDBMS

NoSQL/NewSQL DB

[ Multi-Model + HTAP ]

# 云原生架构+分布式: 弹性计算、弹性存储、水平拓展



# Multi-Model多模数据库系统

北向

查询接口与标准

SQL

Put/Get

SPARQL

DocQL

GQL

...

数据库引擎

数据结构与类型

结构化数据

图

时序

向量

文档

时空数据

非结构化数据

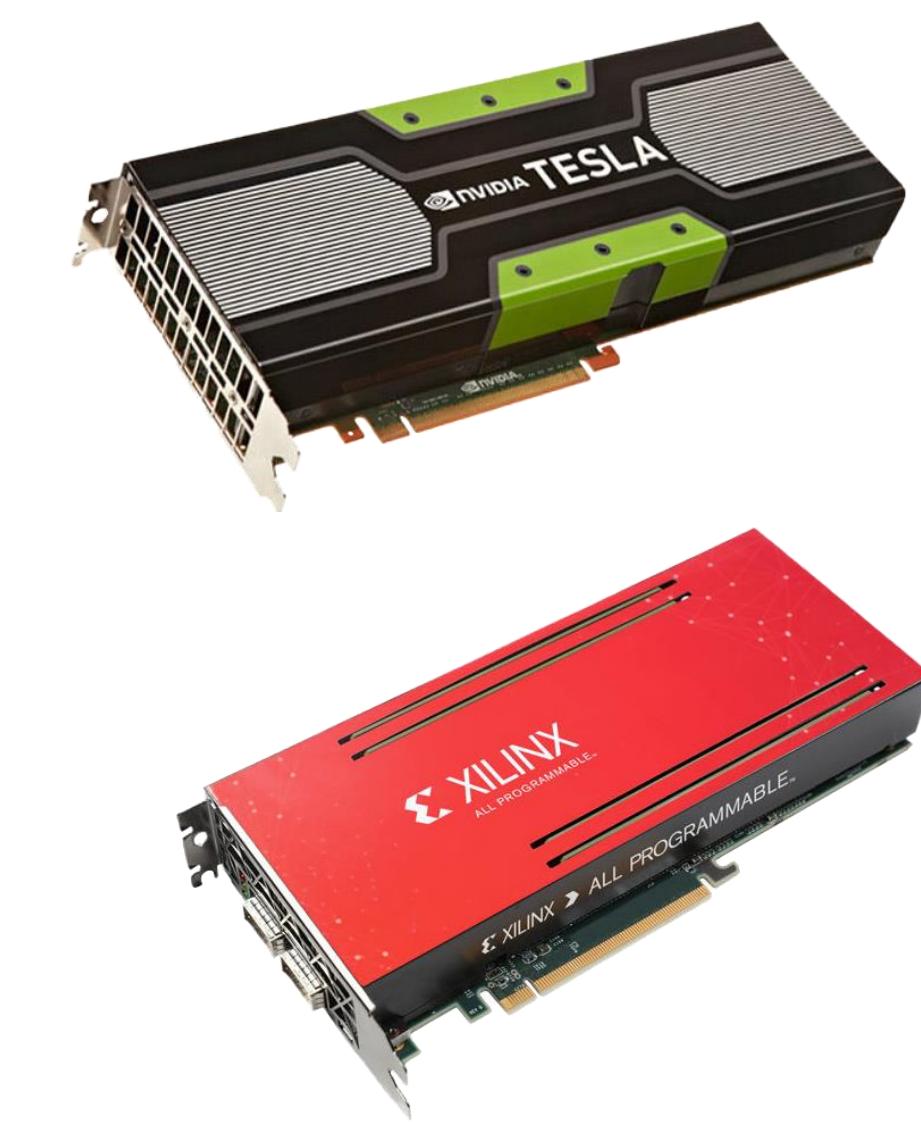
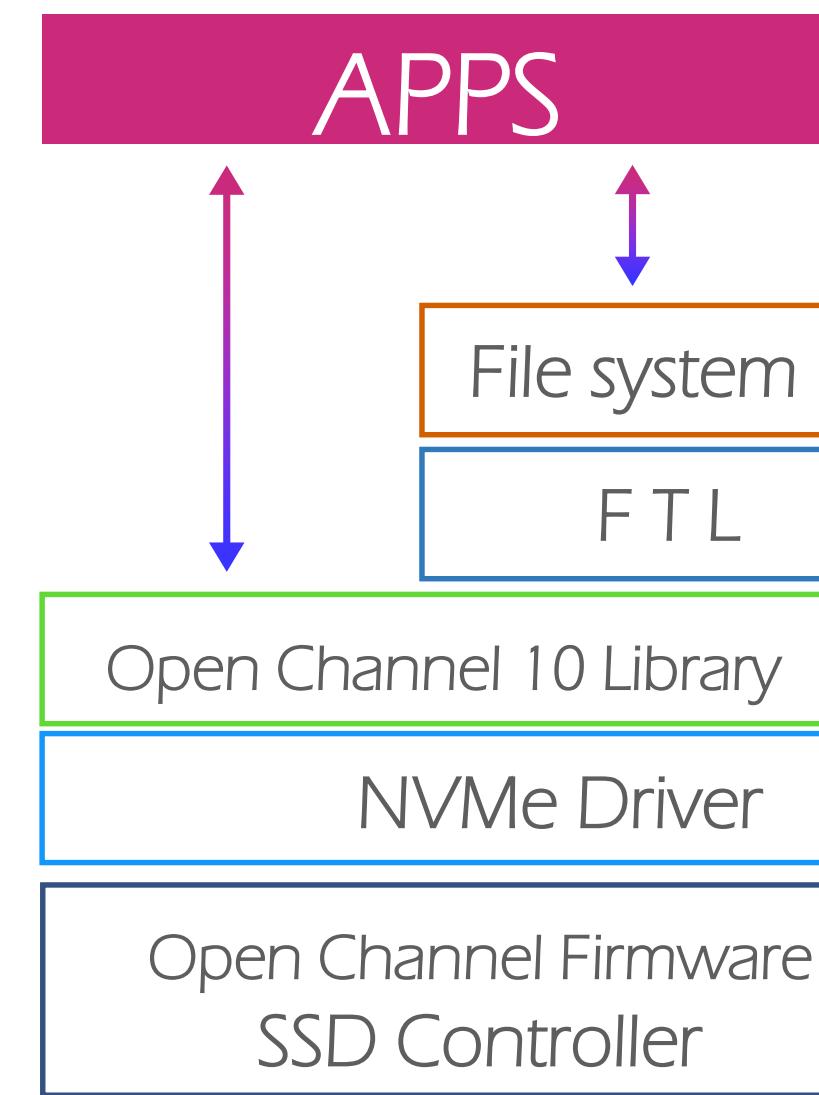
南向

# 数据库智能化+自动化管控平台

提升数据库内核与数据库平台运行效率、节省运行成本

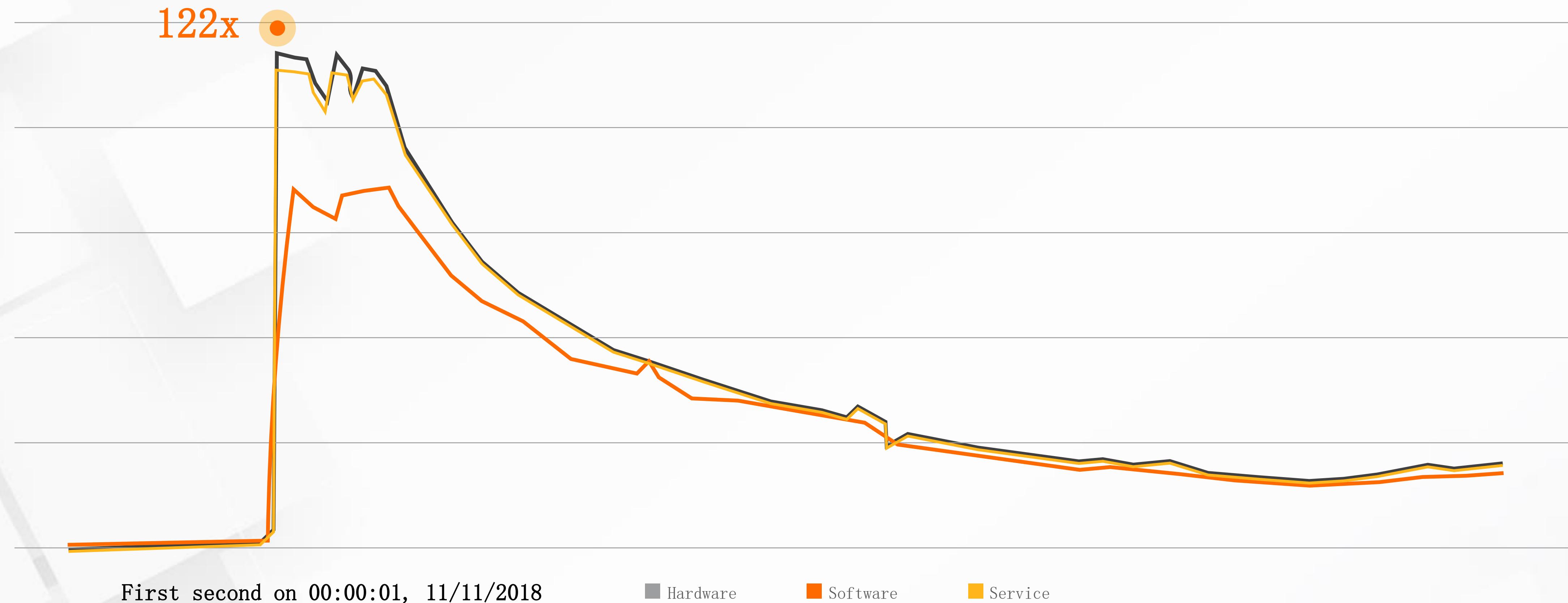


# 新硬件: 软硬件一体化设计



# Singles Day (11/11) 2018

Database workloads



# 阿里云数据库产品是一个完整的生态体系



数据库和应用迁移 ADAM

应用评估、迁移与实施



数据传输 DTS

一键无缝迁移上云



POLARDB



阿里云市场

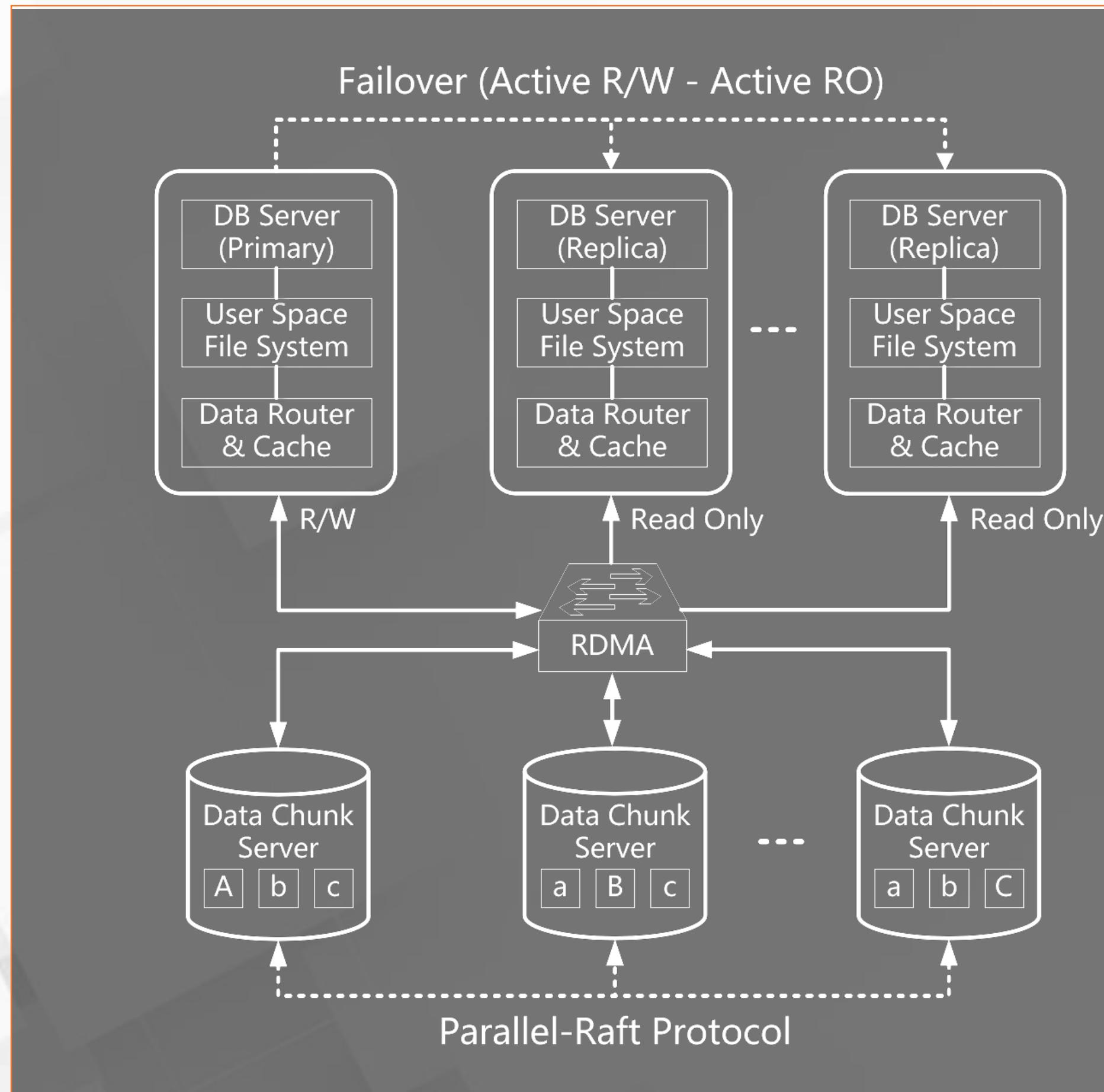
丰富的第三方合作生态



AnalyticDB

支持PB级数据库实时分析

# 云原生数据库: POLARDB (VLDB2018, SIGMOD2019)



## ◎ 共享存储 + 一写多读: 弹性+按需+按量

- ▶ 存储计算分离
- ▶ 100TB 存储容量
- ▶ 100万 QPS / 节点

## ◎ 唯快不破

- ▶ 分钟级 备份恢复
- ▶ 秒级 弹性扩容

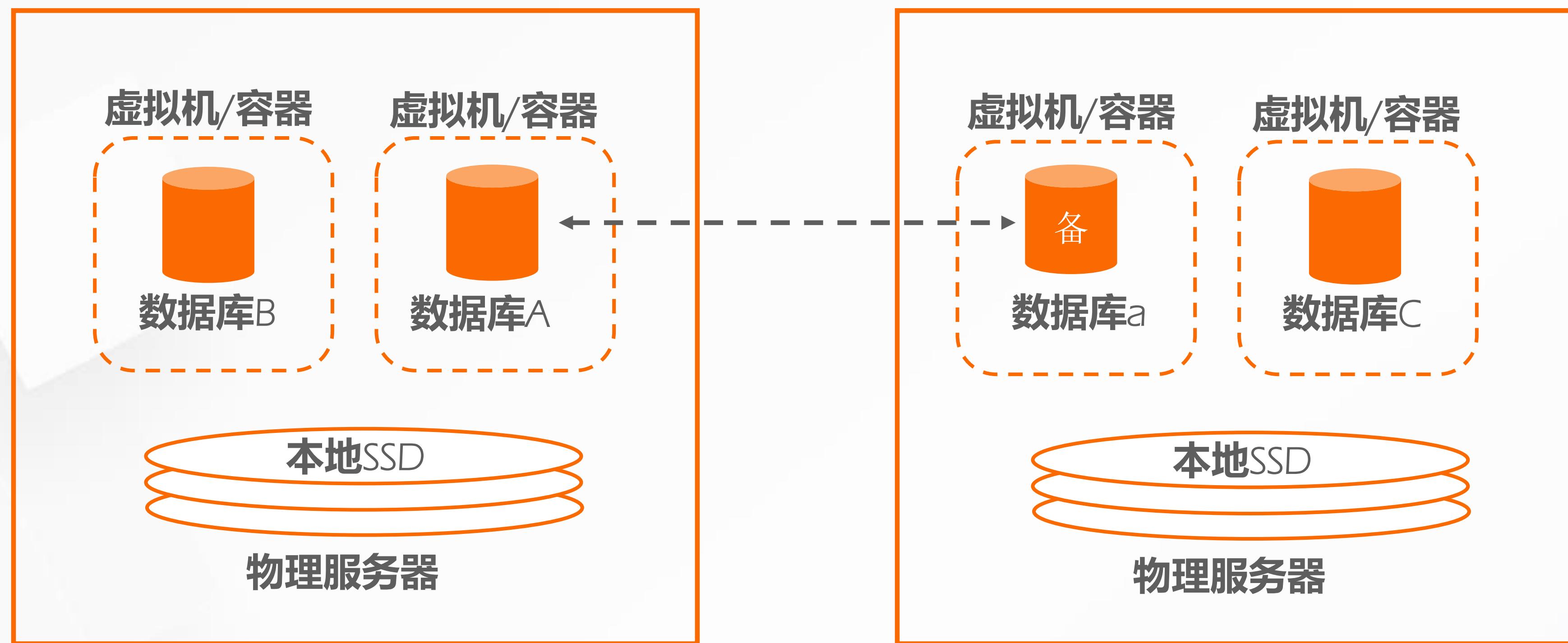
## ◎ 高可用

- ▶ Raft 协议提供金融级高可用

## ◎ 兼容生态

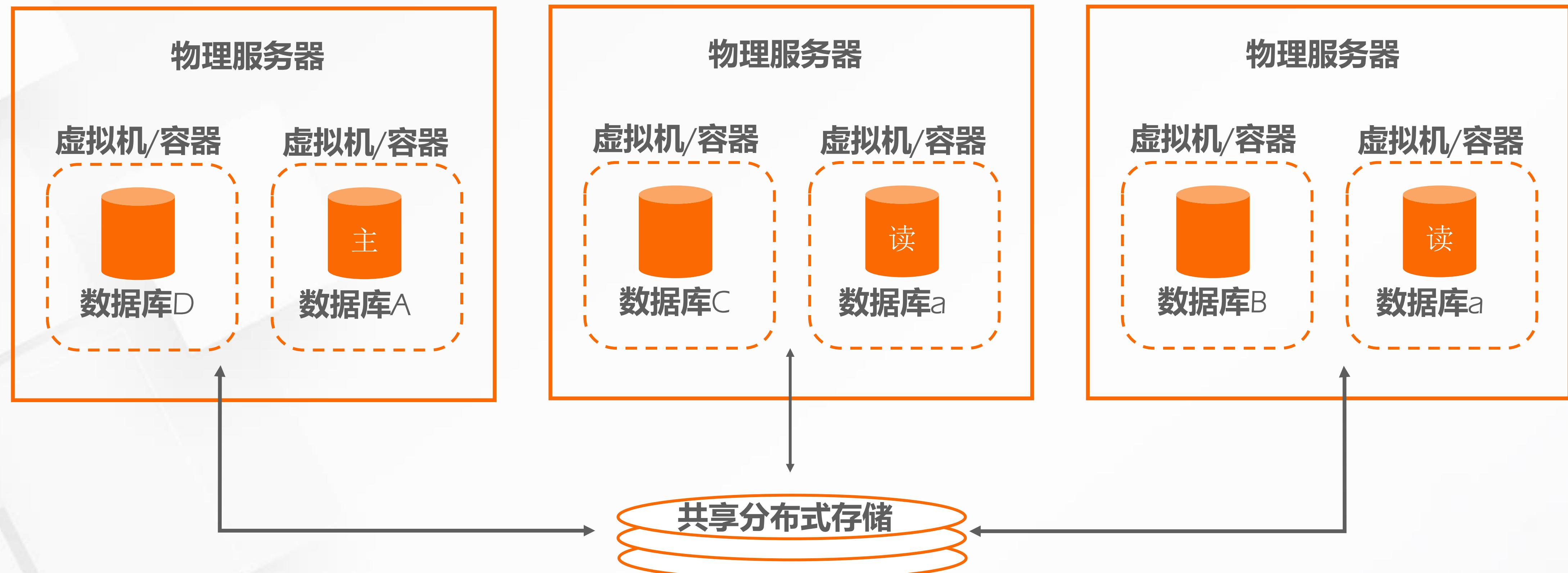
- ▶ 100%兼容 MySQL 5.6
- ▶ 2019: MySQL 8.0, Oracle, PG

## MySQL高可用



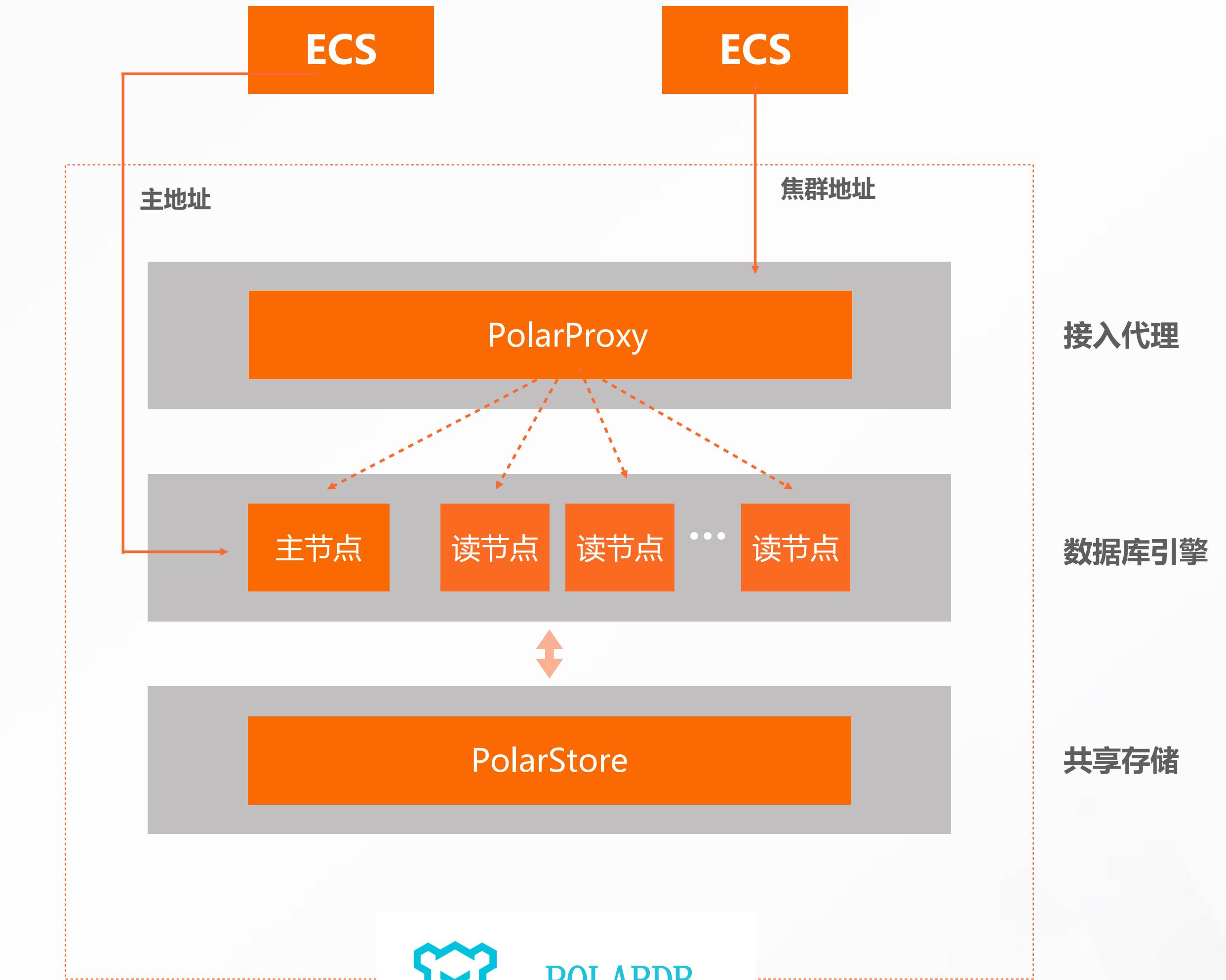
MySQL 高可用 (Active/Standby)

## POLARDB架构细节

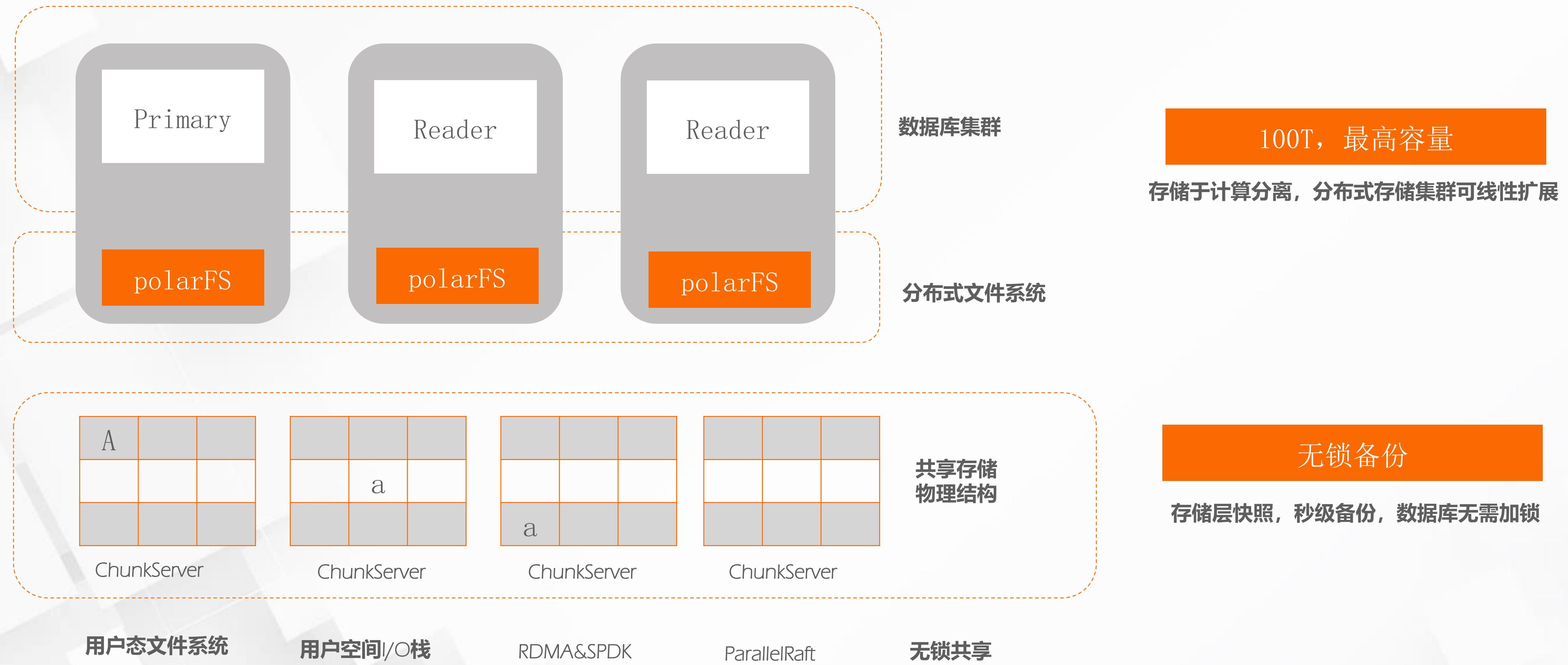


Active-Active + 读写分离 + Serverless 存储

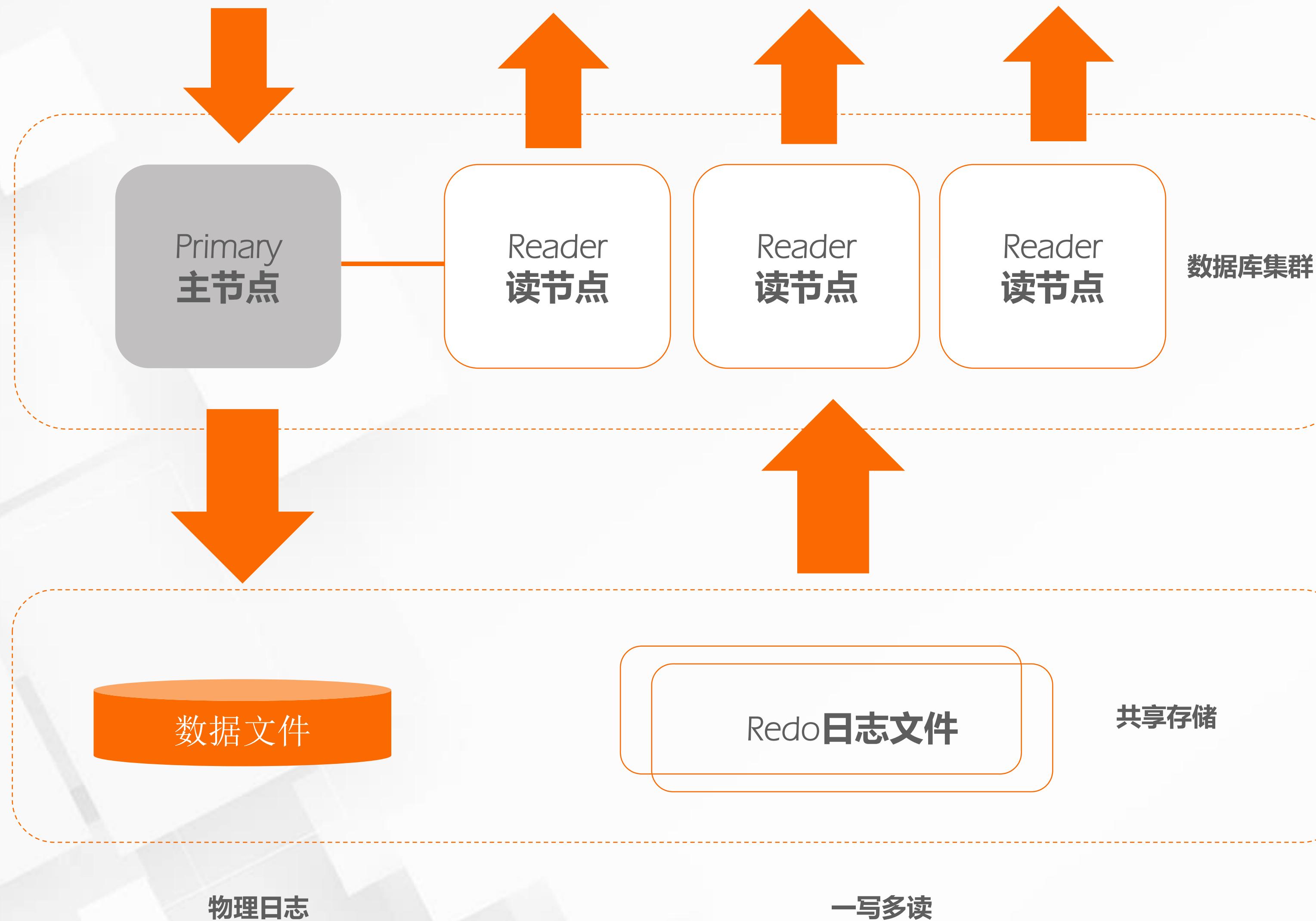
# POLARDB架构细节



# POLARDB架构细节: 共享存储-分布式+三副本



# POLARDB架构细节: 一写多读集群



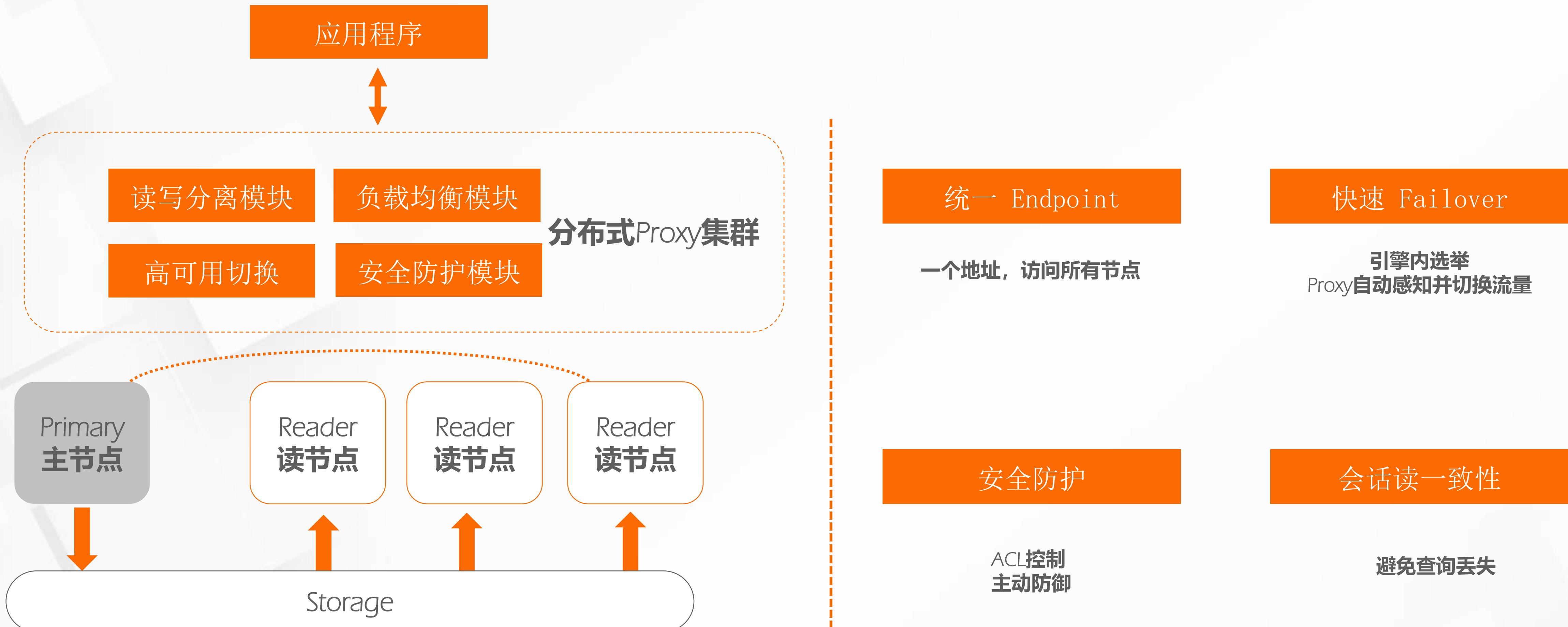
5分钟，快速伸缩

2核vCPU升级到32核，5分钟生效！  
双节点扩展到4节点，5分钟生效！

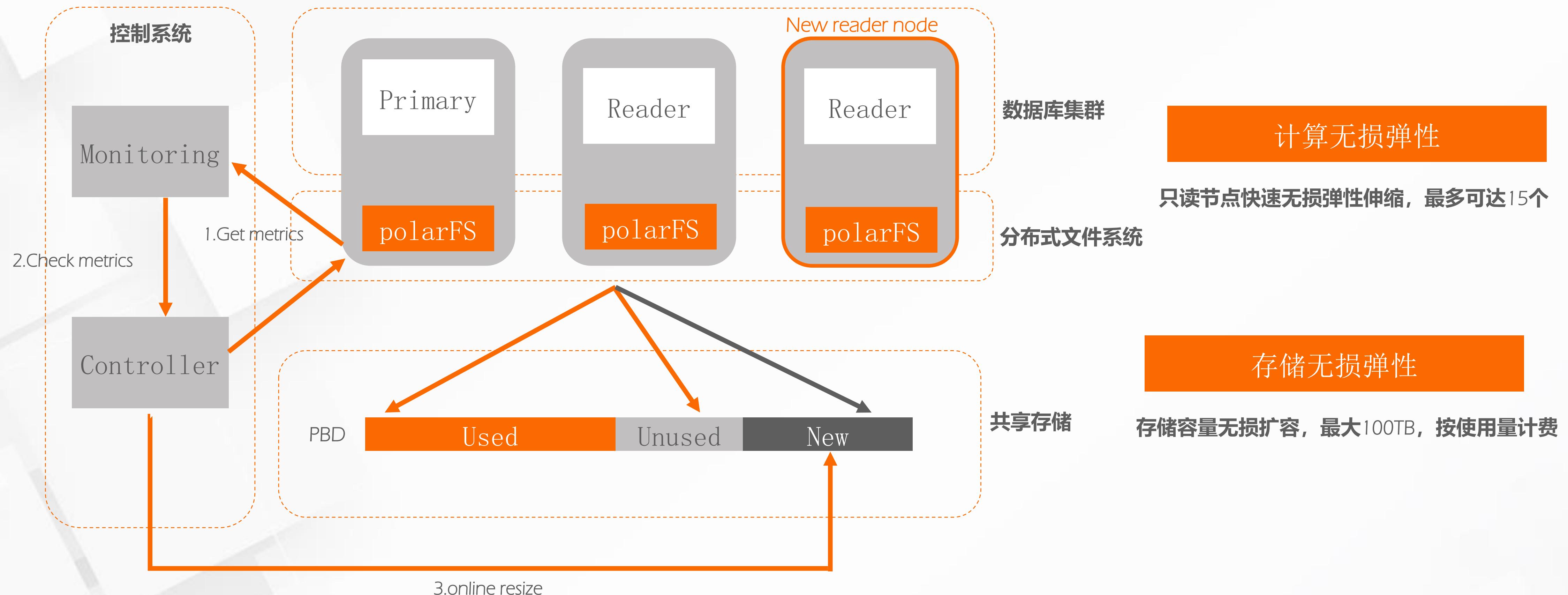
成本降低30%~50%

Serverless计费  
16个节点，只需1分钟存储  
统一存储池，减少资源碎片

# POLARDB架构细节: 接入代理-调度+读写分离



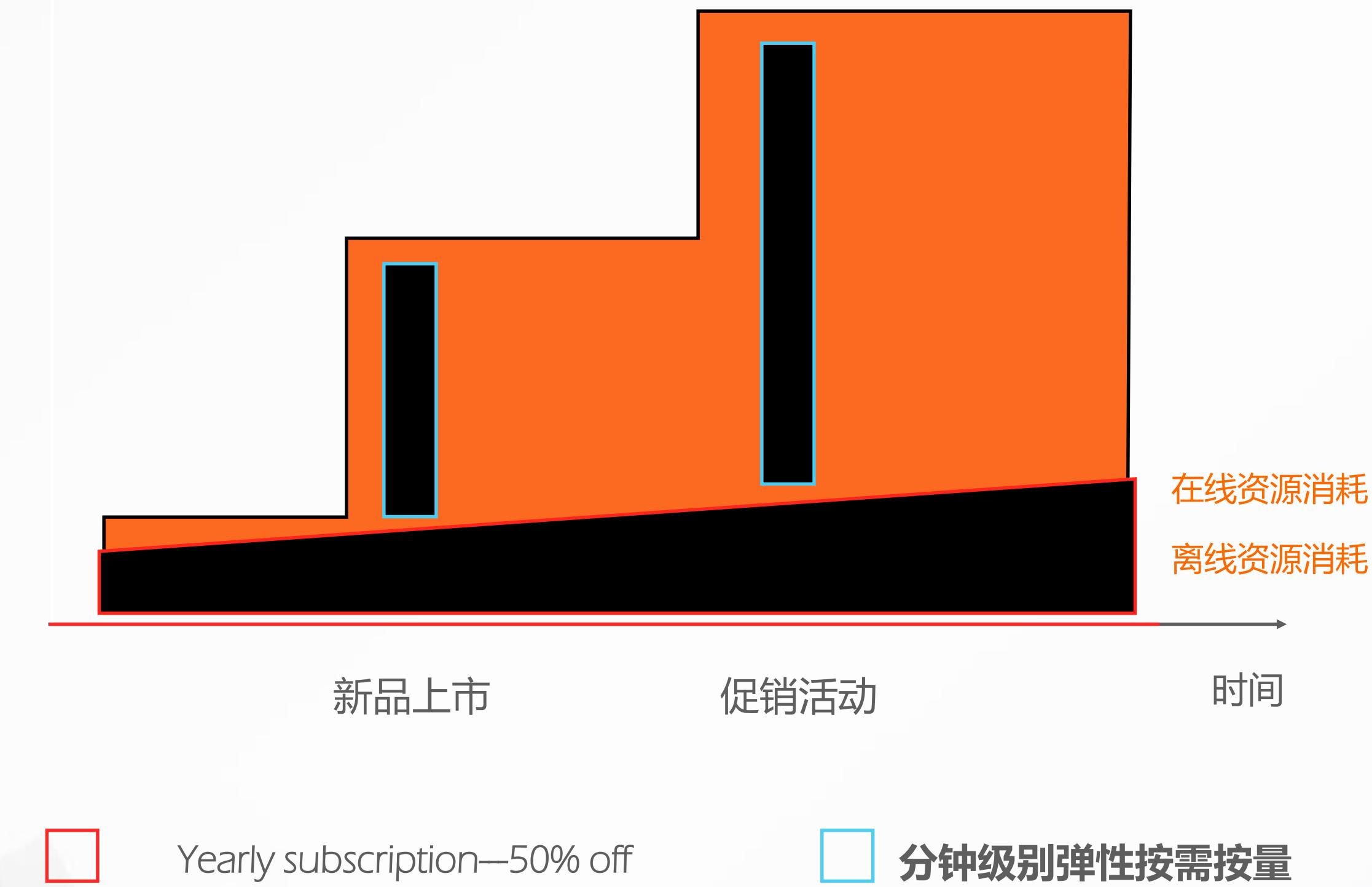
# POLARDB架构细节: 无损弹性



# 云上极致弹性

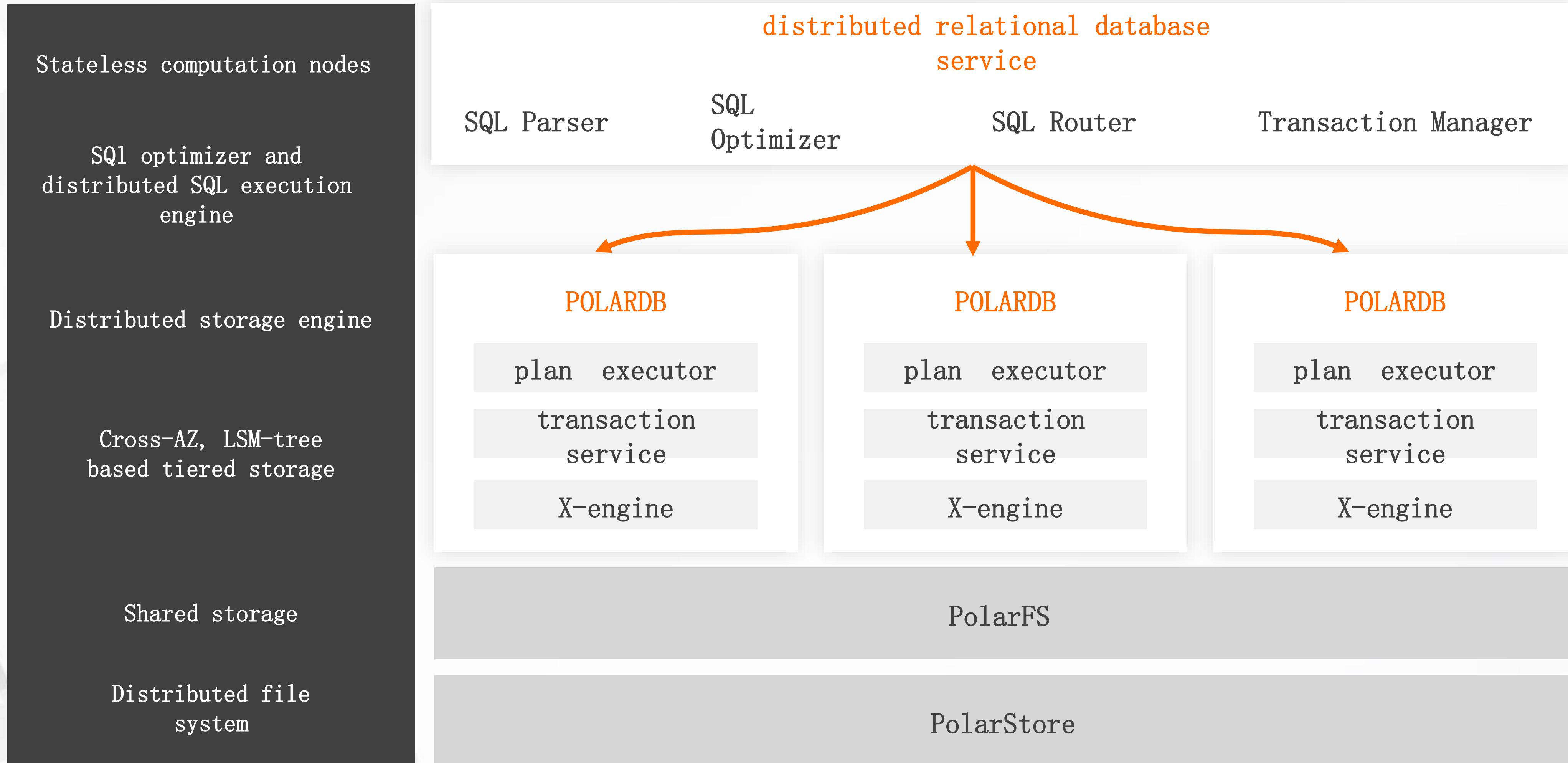
on-demand usage and elastic billing

计算节点资源需求



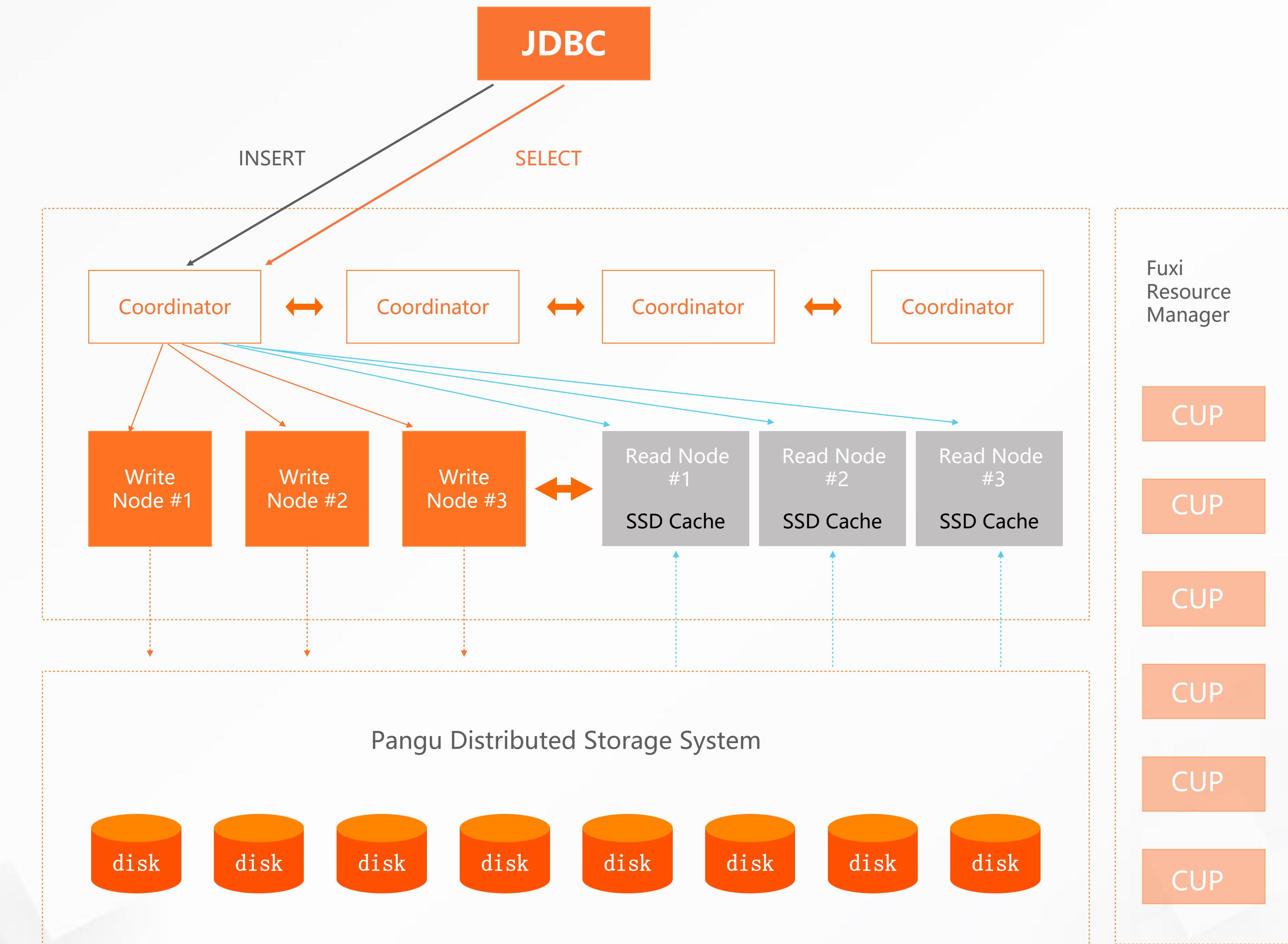
# POLARDB-X: 分布式版本-支持水平拓展 (SIGMOD 2019)

SQL (DML/DDL/DCL)

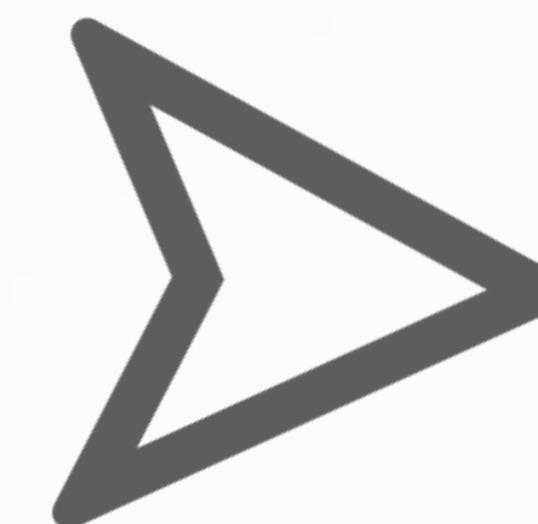
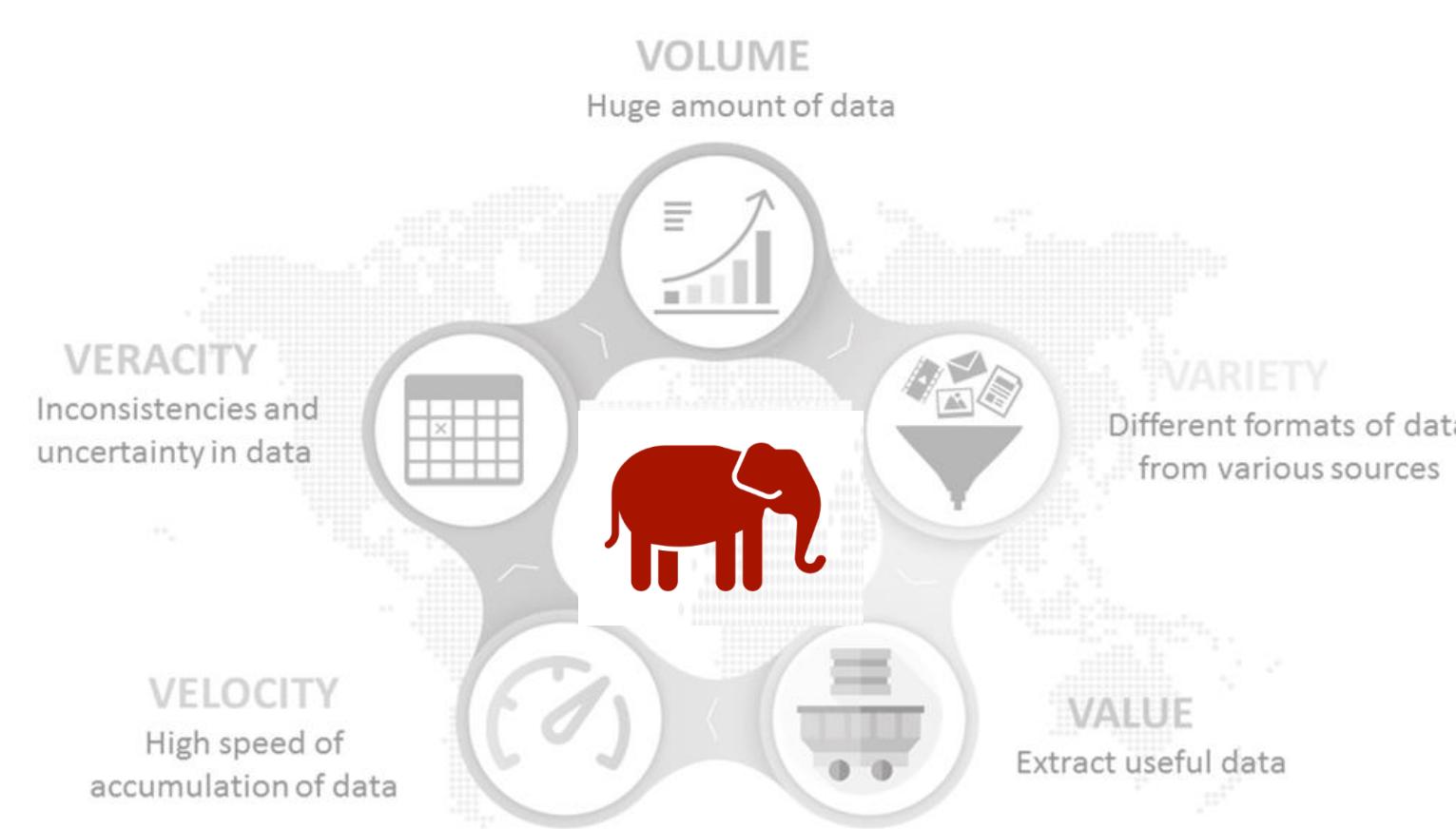


# OLAP: AnalyticDB-下一代实时交互式数仓 (VLDB 2019)

- 兼容生态
    - ▶ MySQL版本兼容MySQL生态
    - ▶ PostgreSQL版本兼容PG生态
  - 行列混存引擎
    - ▶ 高吞吐写入 (140,000 rows/s)
    - ▶ 高并发查询
  - 海量数据处理
    - ▶ PB级数据存储+分析
  - 查询优化
    - ▶ CBO+全量代码生成
    - ▶ 低延时分析 (毫秒级别响应)
  - 高可用
    - ▶ 99.99 HA
  - 向量分析
    - ▶ 结构化+非结构化数据联合分析

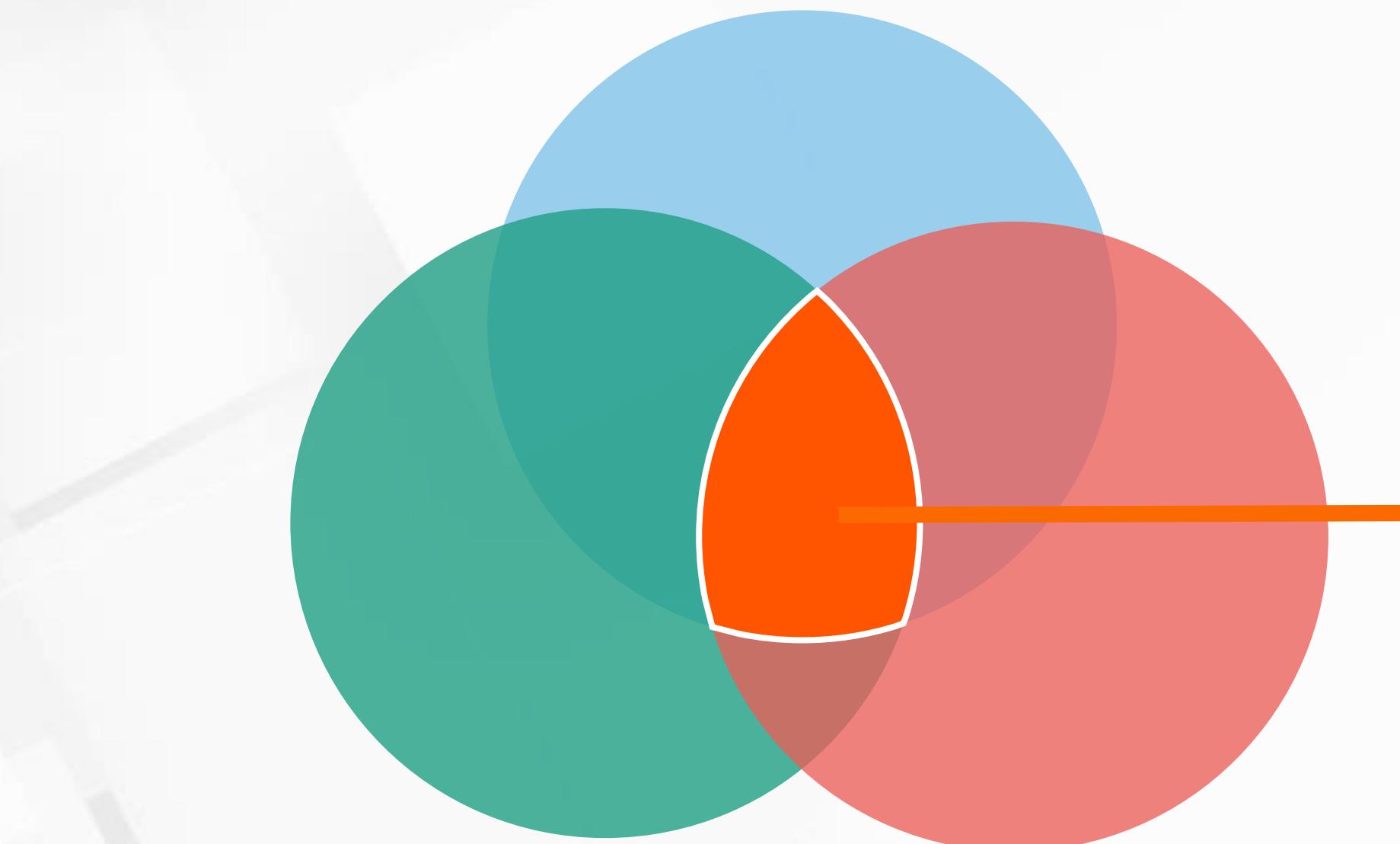


# Big Data 生态演进，Fast Data 是未来



寻求买家...

# AnalyticDB-一个系统一套存储兼顾多种场景



## 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

## 多维分析

- ✓ 任意列 Join
- ✓ 复杂长计算任务、ETL



## 行列混存 (Hybrid Store)

表组/列簇/分区组/Co-located  
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

# AnalyticDB-典型场景



达摩盘

阿里妈妈旗下产品

营销平台

13000+  
表数目

70T+  
总数据量

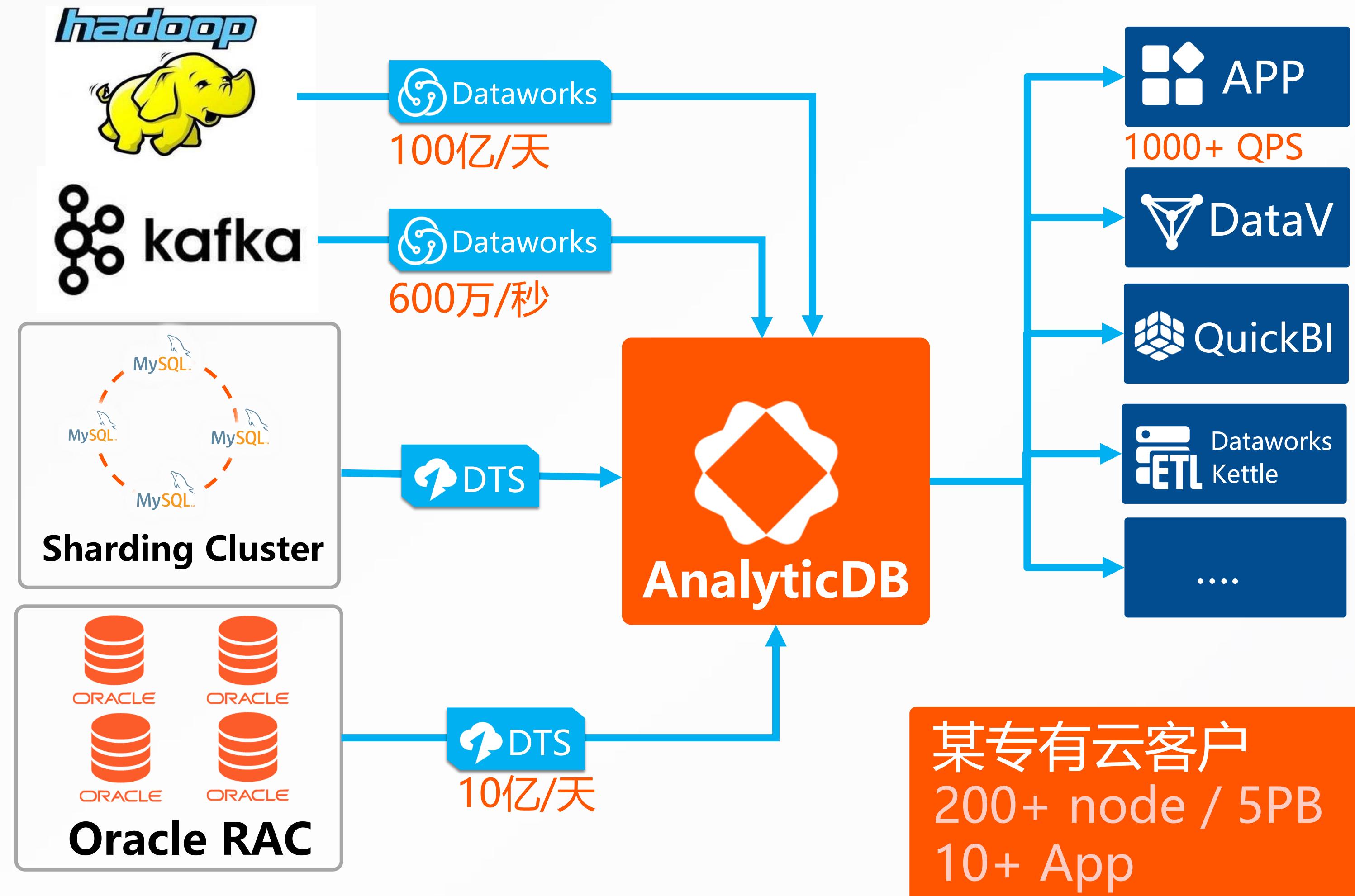
10+  
百亿大表  
Join数量

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



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，全局开放分析



## 无服务器化、扩容能力极强，助力分析极低成本

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



## 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)



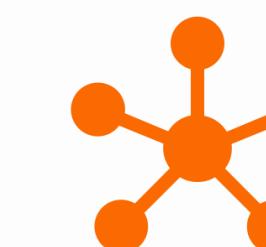
## 统一元数据，双计算引擎按需切换

- ✓ 极速分析引擎XIHE TPC-H、TPC-DS 领先 3x
- ✓ Spark引擎 支持共享内存加速 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

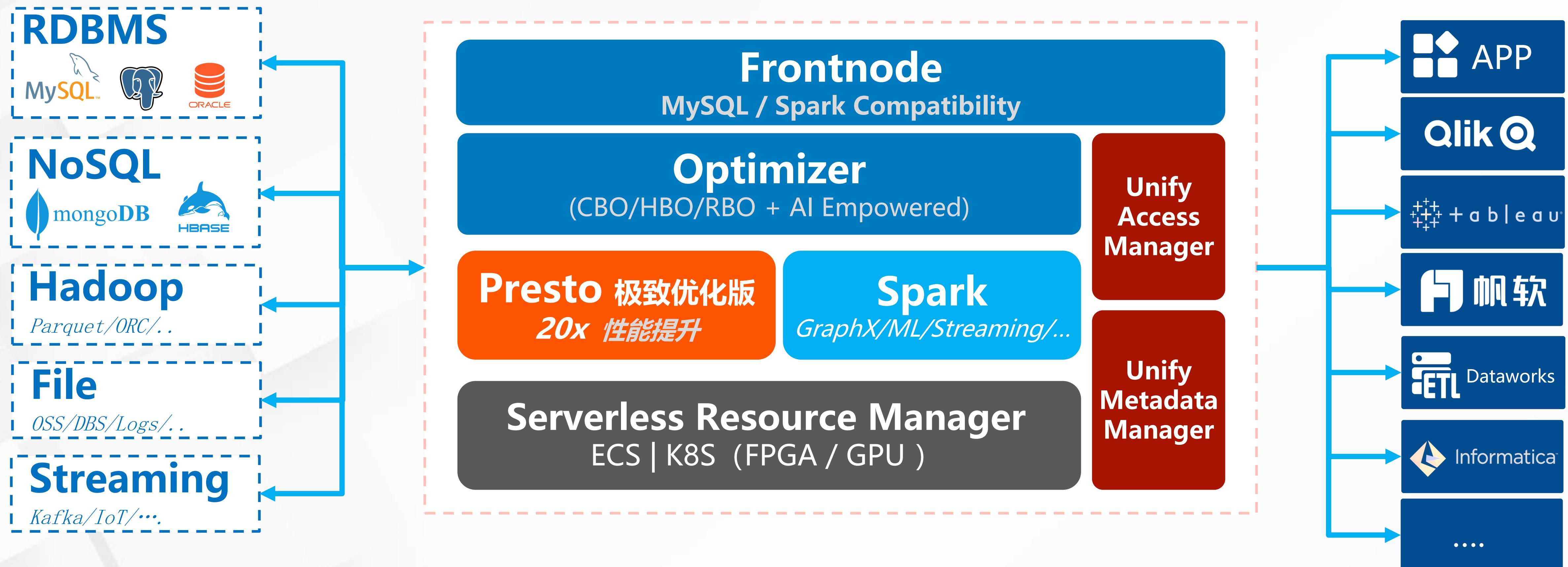


## 无需迁移即可统一分析全域数据

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

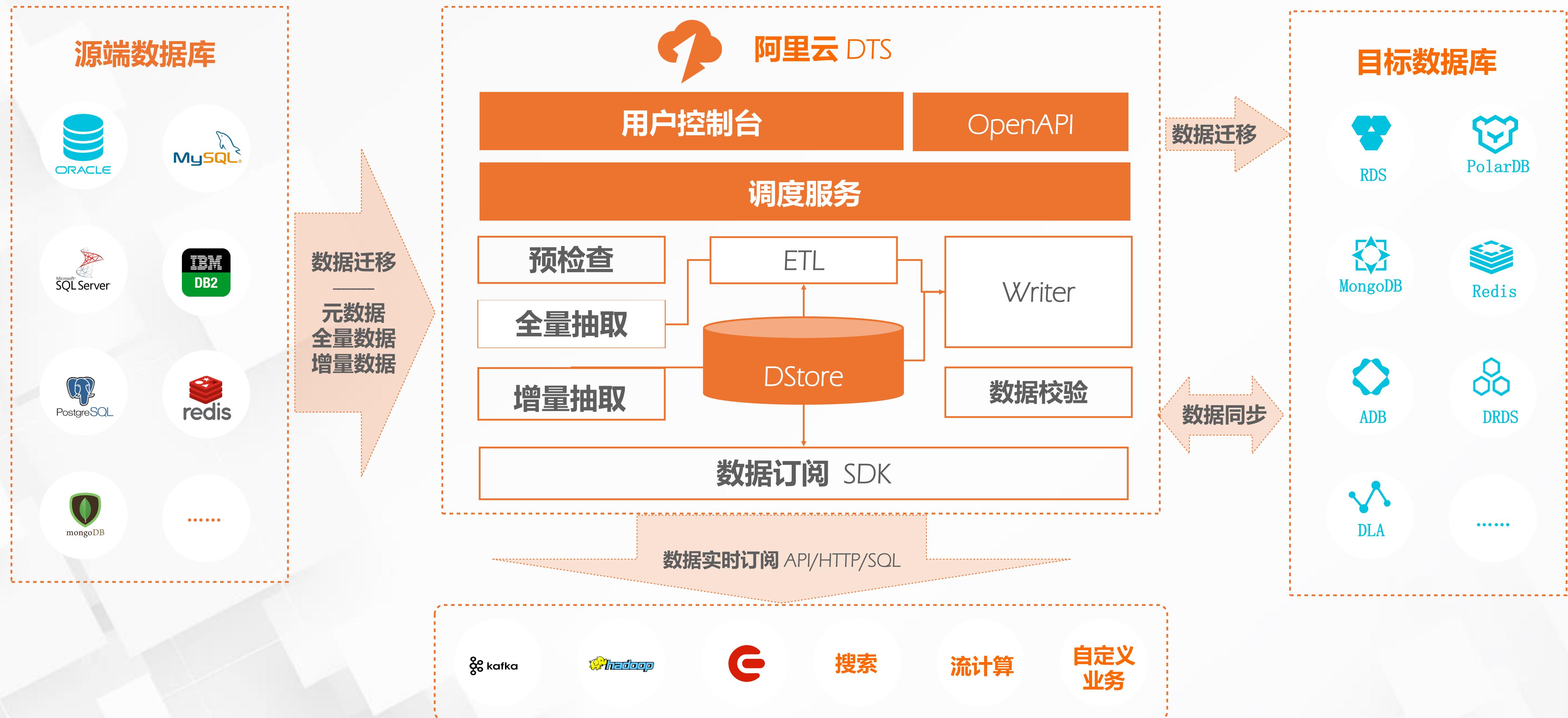
15+  
数据源

# Data Lake Analytics - 全域数据，全局开放分析

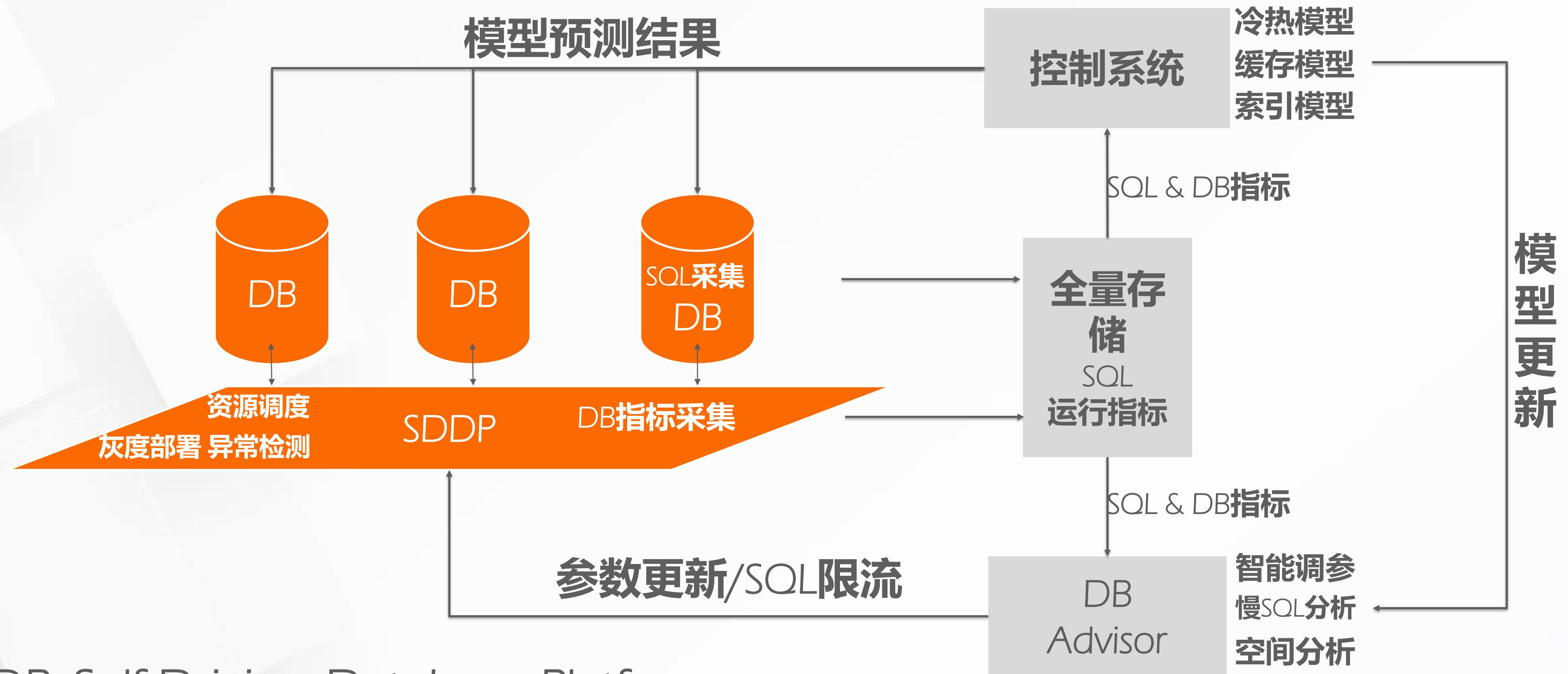


# 数据传输云服务 DTS

## 数据高速公路



# 智能数据库管控系统

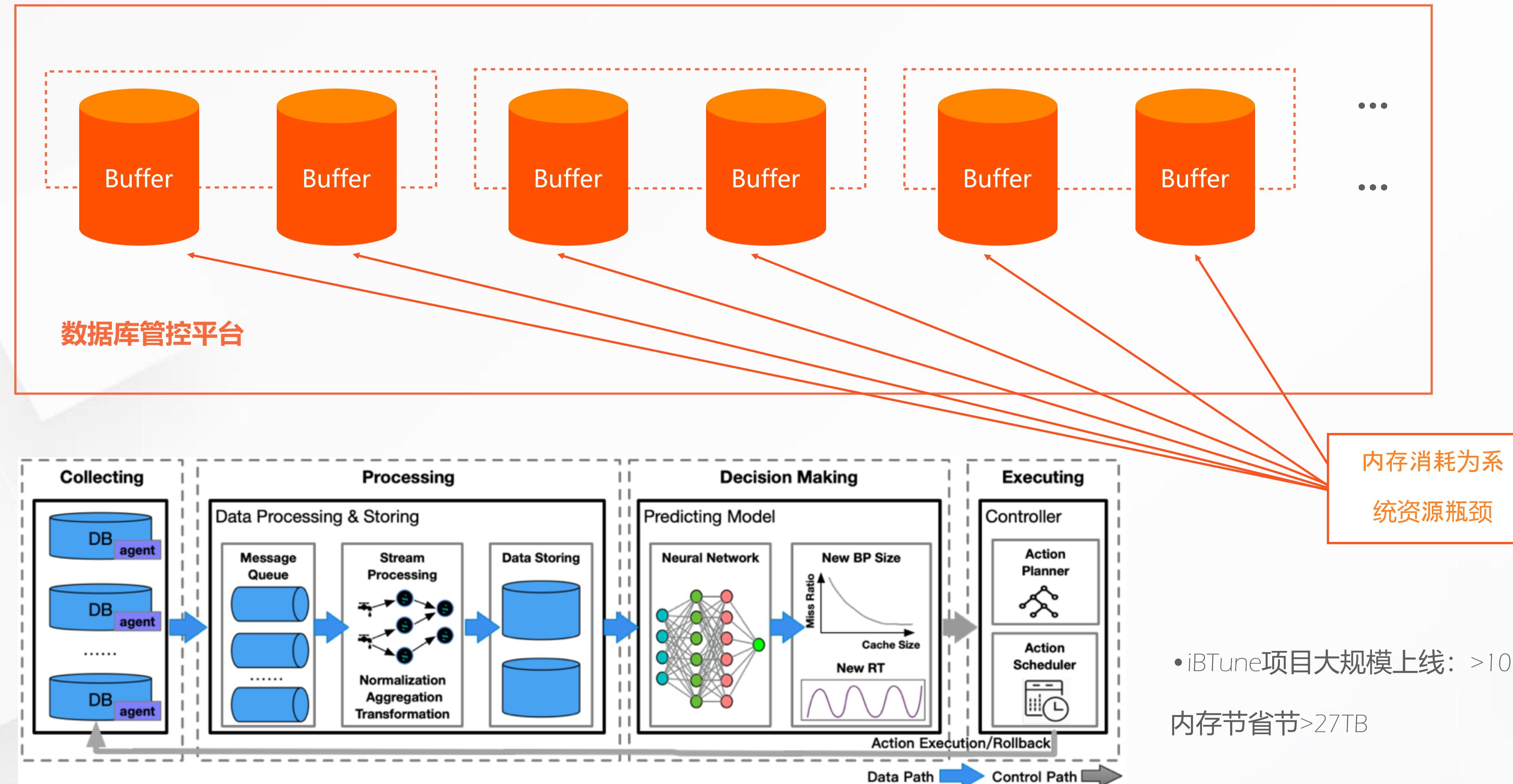


SDDP: Self-Driving Database Platform

智能数据库闭环生态体系

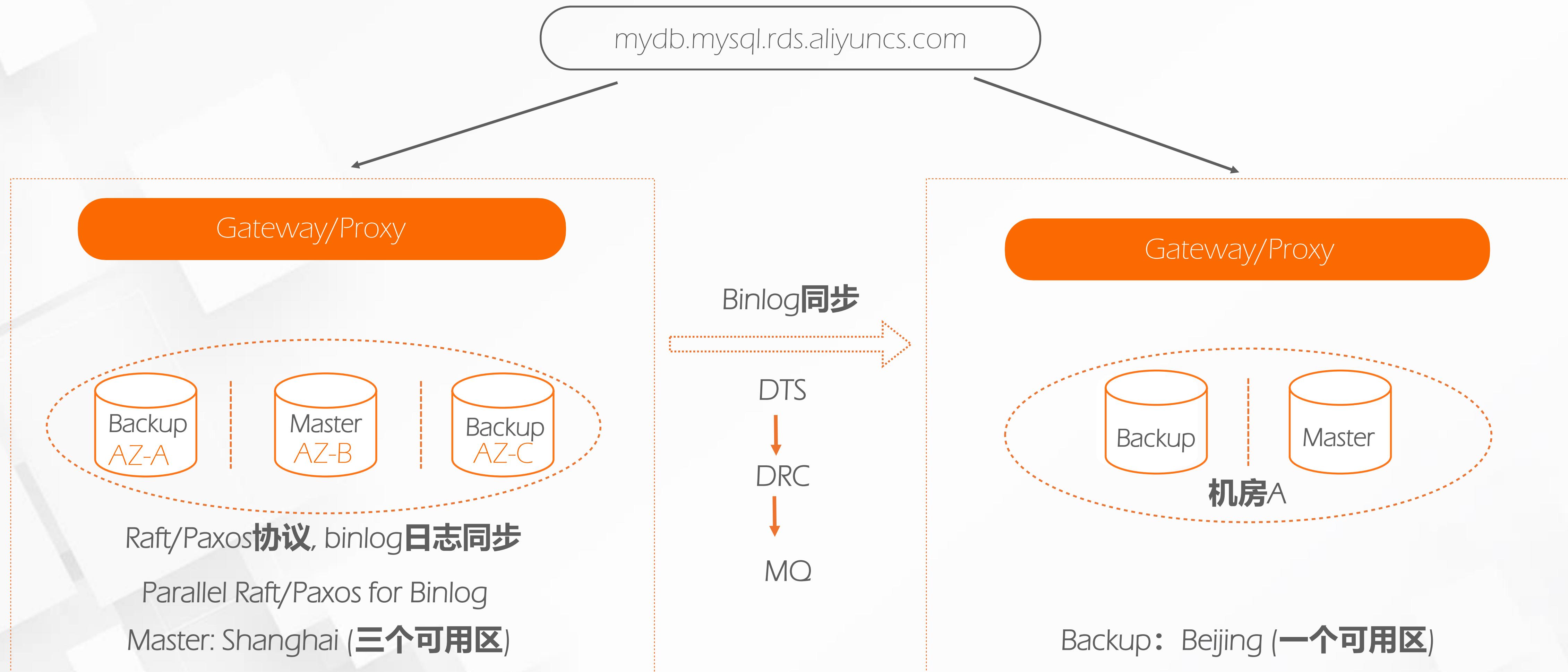
# 智能数据库案例：

智能调参上线iBTune (individualized Buffer Tuning, VLDB 2019)

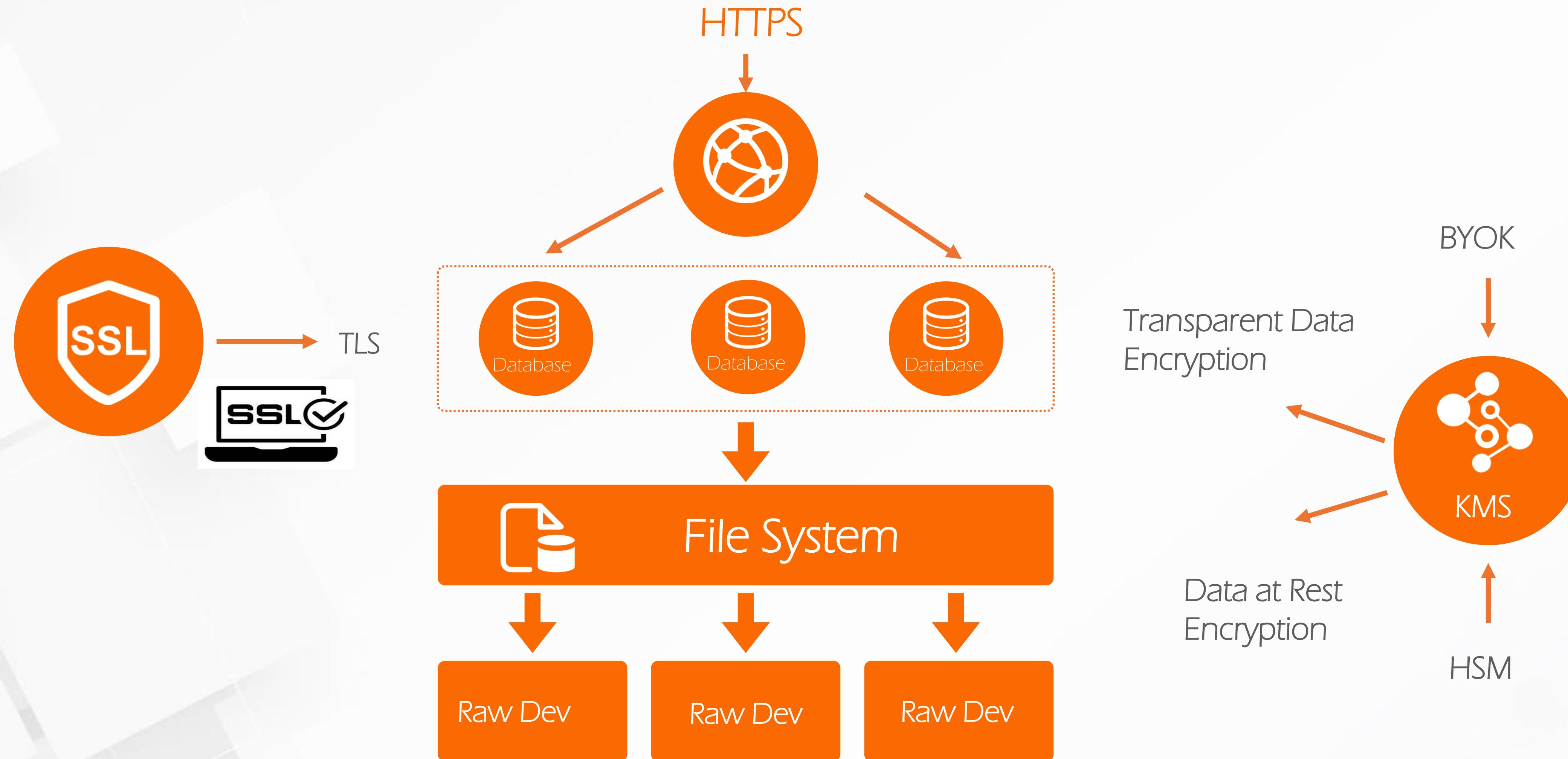


# 高可用: 跨中心、跨可用区

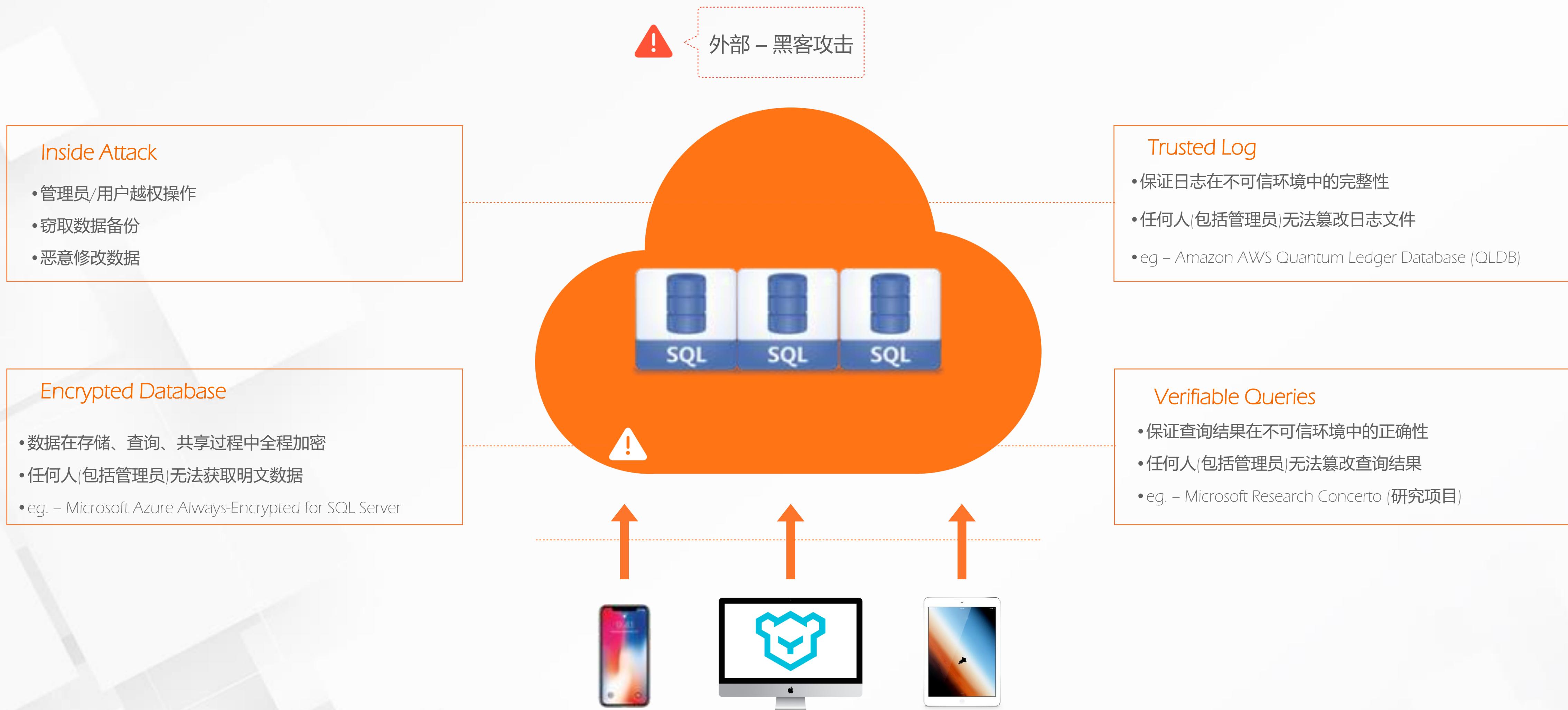
## 用户应用



# 数据安全



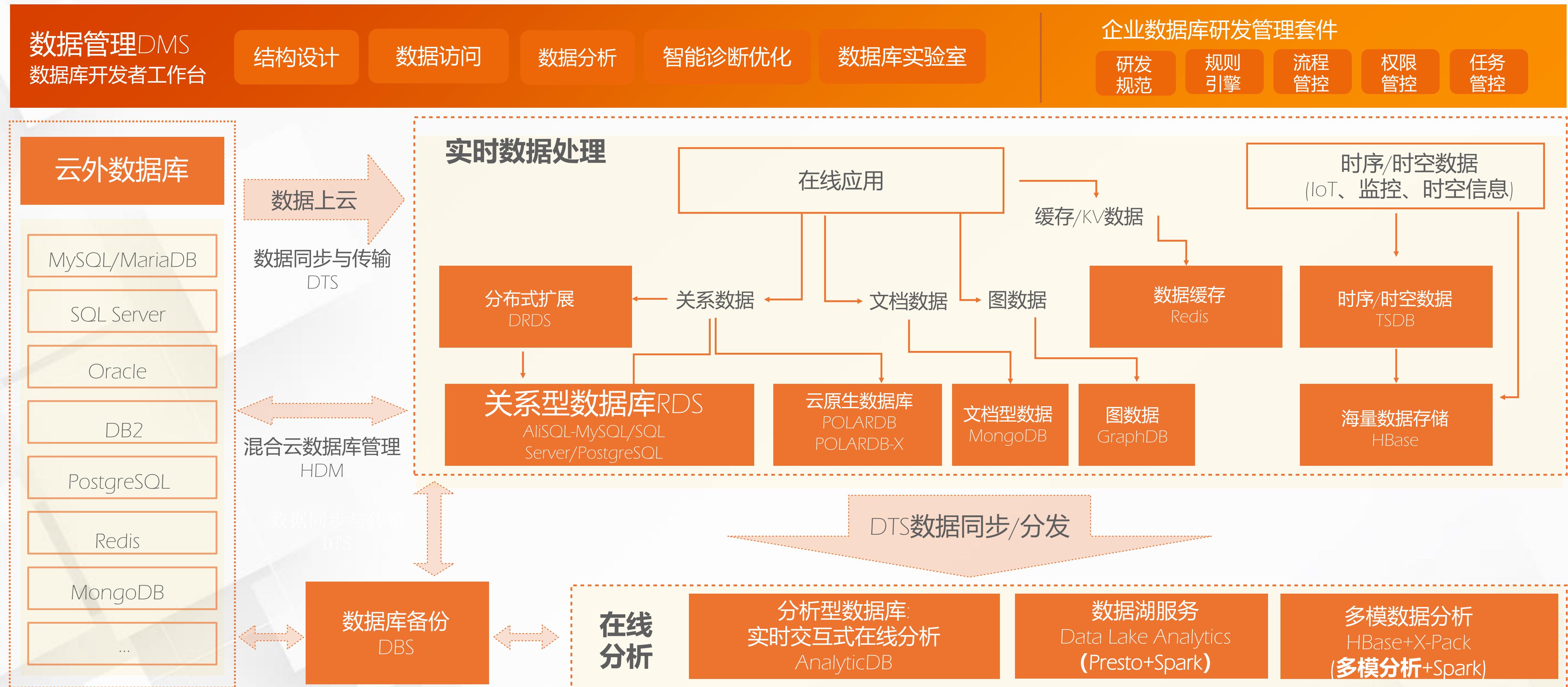
# 云上数据库安全



# 企业级数据库云服务



# 云数据库产品体系与数据链路



# 国际分析师报告

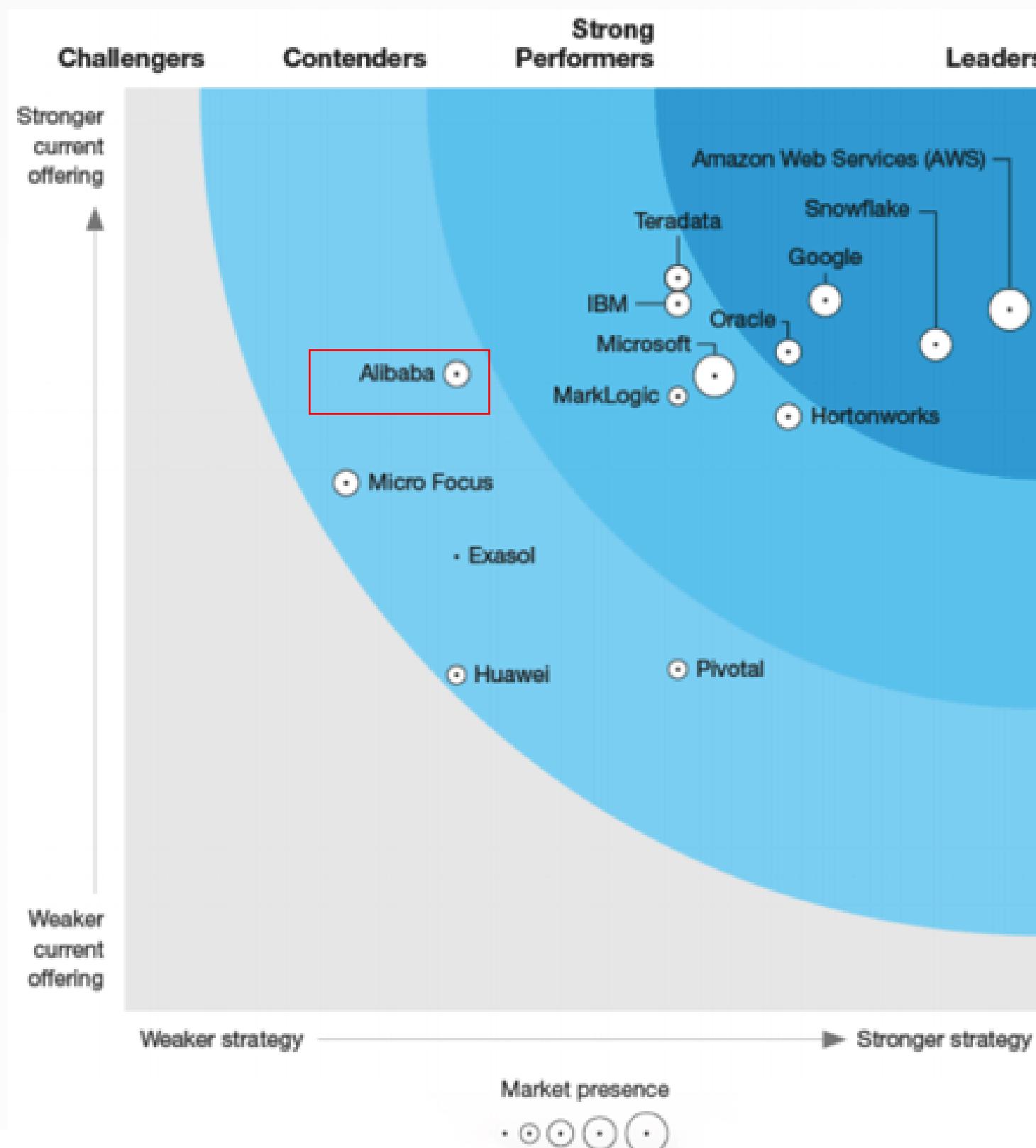
Gartner®

OPDBMS MQ 2018.10



FORRESTER®

Cloud Data Warehouse, Q4 2018



OLTP

OLAP

FORRESTER®



DBaaS

## 云数据库营收排行榜-Gartner

Table 1. DBMS Cloud Services Revenue (2016-2018)

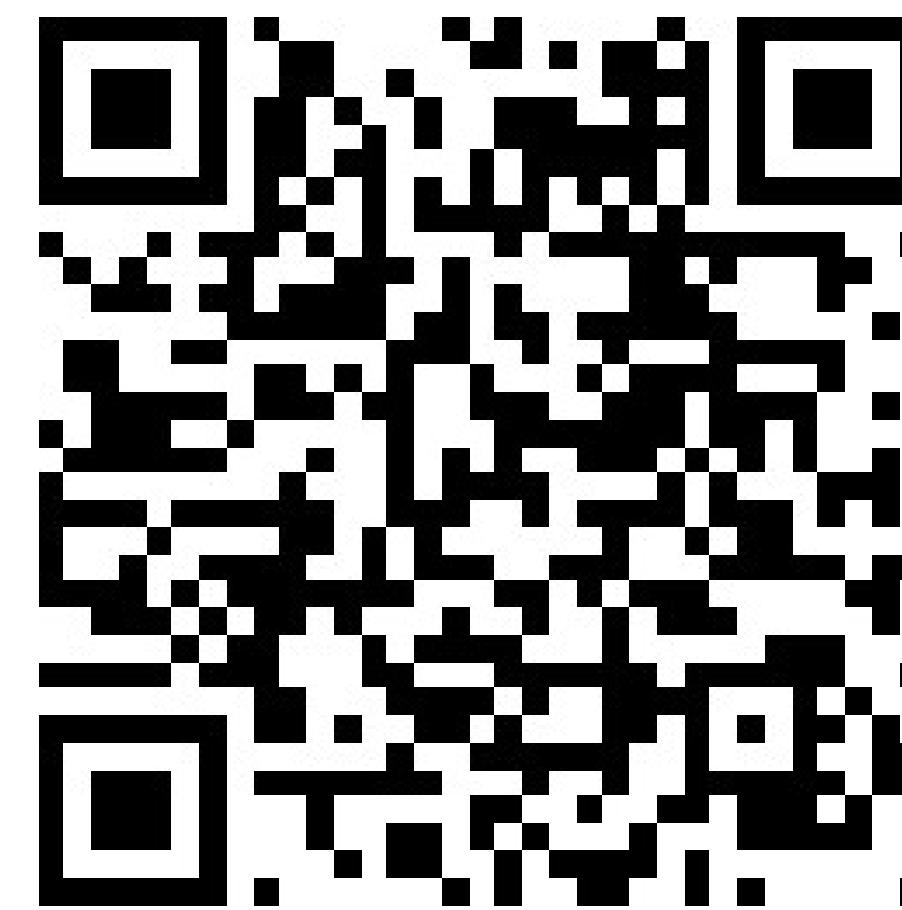
	Revenue			Revenue Growth	
	2016	2017	2018	2017	2018
Amazon	1,700.64	3,615.90	6,319.11	112.62%	74.76%
Microsoft	53.38	918.27	2,149.40	1620.11%	134.07%
Alibaba	96.93	213.44	460.55	120.21%	115.77%
Oracle	100.16	224.76	373.12	124.40%	66.01%
Google	101.47	164.36	285.49	61.98%	73.70%
Tencent	21.87	110.85	247.30	406.96%	123.09%
Huawei	77.42	70.99	137.87	-8.32%	94.22%
IBM	57.28	73.35	120.22	28.06%	63.90%
Cloudera	23.85	45.35	79.21	90.15%	74.66%
MongoDB	9.12	8.77	65.70	-3.90%	649.33%
Other	127.07	171.69	250.74		
<b>Grand Total</b>	<b>2,369.19</b>	<b>5,617.73</b>	<b>10,488.70</b>	<b>137.12%</b>	<b>86.71%</b>
% of Total DBMS	6.87%	14.43%	22.75%		

Data sourced from "Market Share: Enterprise Platform as a Service, Worldwide, 2018"

# 城市大脑案例



ADB、TSDB、POLARDB、DRDS、DMS、DTS



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



Thanks