

阿里云数据库 SelectDB 版

-100% 兼容 Apache Doris 的全托管云原生实时数仓

CONTENT 目录

01 阿里云SelectDB版介绍

总体介绍

02 阿里云SelectDB版核心特性

核心亮点

03 阿里云SelectDB版案例介绍

数百家客户的选择

Apache Doris 是全球流行的开源实时数仓产品



Open-Source Real-Time
Data Warehouse

2013
Project Creation

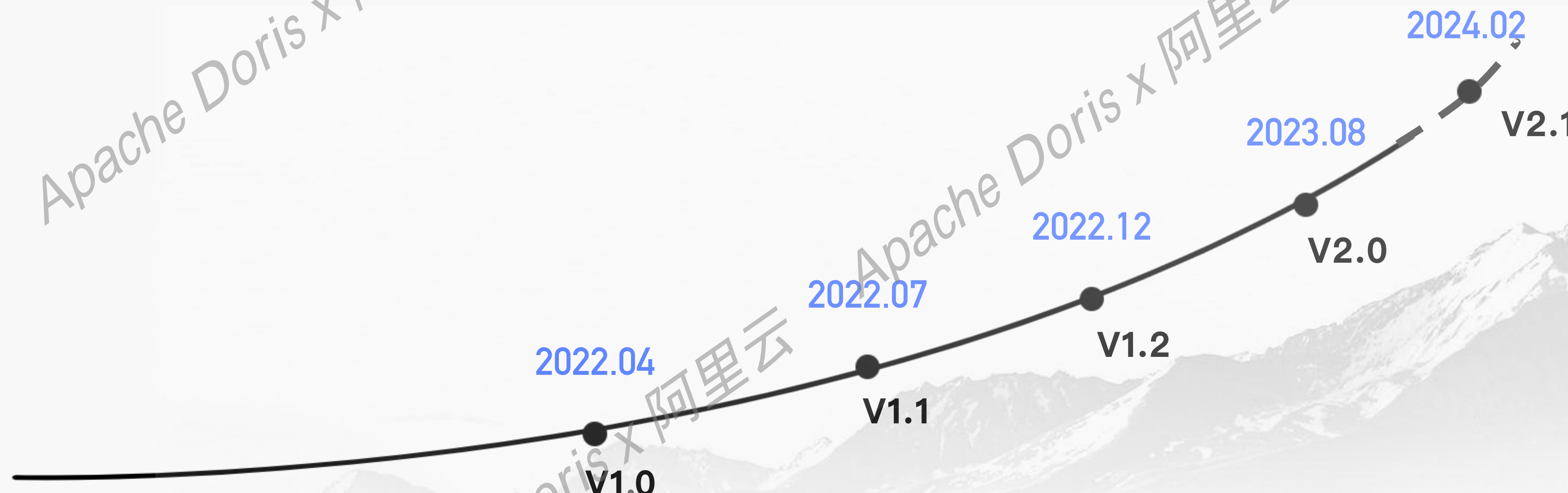
2017
Open Source

2022
ASF Top Project

10.6k +
GitHub Stars

550+
Contributors

4000+
Enterprises



Part 1

阿里云SelectDB版

总体介绍

阿里云 SelectDB: 基于 Apache Doris 构建的现代化实时数据仓库



云原生架构



实时极速



融合统一



开放生态



阿里云数据库 SelectDB版 – 阿里云和飞轮科技战略合作产品

飞轮科技 Apache Doris 产品的商业化公司



可靠服务

云数据库 SelectDB 是阿里云一方产品，
阿里云提供服务及 SLA 保障



强大支撑能力

云数据库 SelectDB 由阿里云产研及 Apache Doris 主创团队
(完成90%新增代码) 联合开发和支持



完善生态对接

云数据库 SelectDB 完整适配
Apache Doris 开源生态以及阿里云产品生态



Part 2

阿里云SelectDB版特性介绍

核心亮点

阿里云数据库 SelectDB 四大核心特性



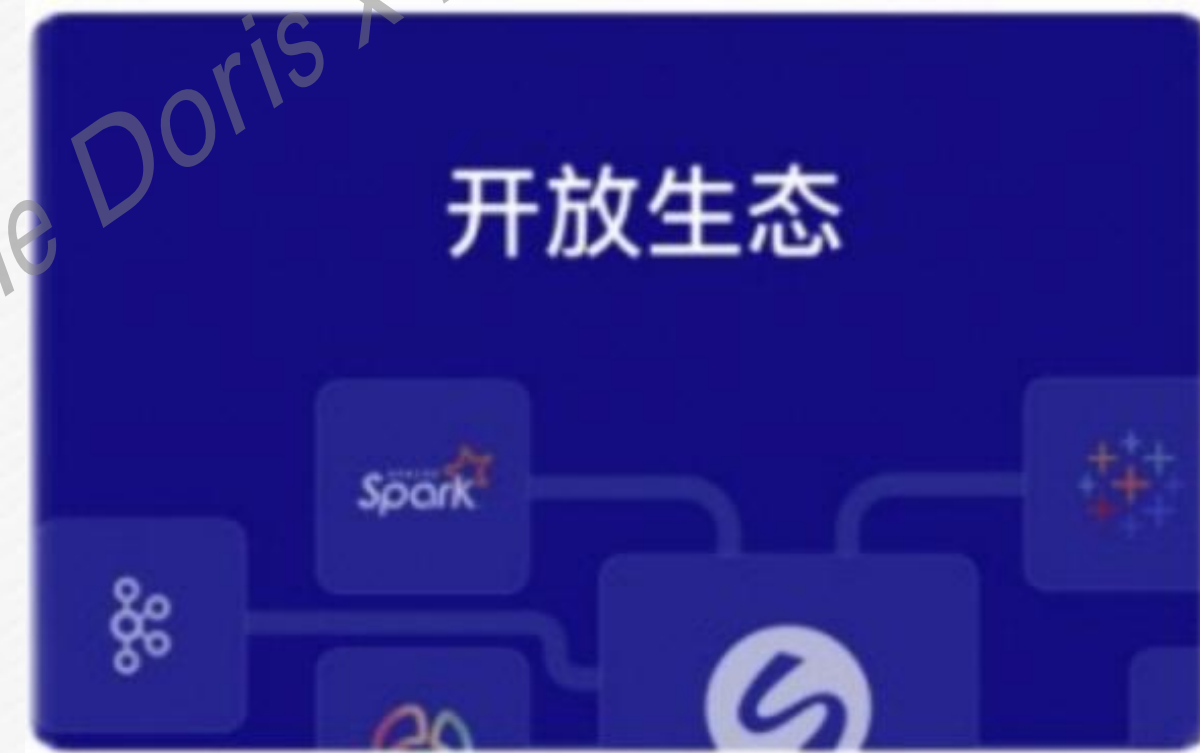
极致性价比与灵活的弹性能力



实时数据写入与极速查询响应



单一系统可支持各种分析负载



开放设计简化与外围系统集成

云原生架构 - 主要特性

存算分离

存储卸载到低成本高可靠的对象存储服务

计算弹性

计算呈现波峰波谷，希望计算更加弹性

云原生

计算负载隔离

读写分离，业务隔离等

数据共享

同一份数据被多个计算业务使用

云原生架构 - 存算分离带来极致性价比



存算分离与弹性计算

- 核心存储：对象存储 OSS，实时扩容
- 计算集群：分钟级弹性，支持多集群
- 缓存存储：通过云盘缓存加速，秒级扩容

极致性价比

- 存储：数据分片单副本，相比存算一体降低 66%~；对象存储单价相比云盘低 80%+
- 计算：单副本消除冗余数据写入、合并开销，结合弹性，计算资源可降低 50%~

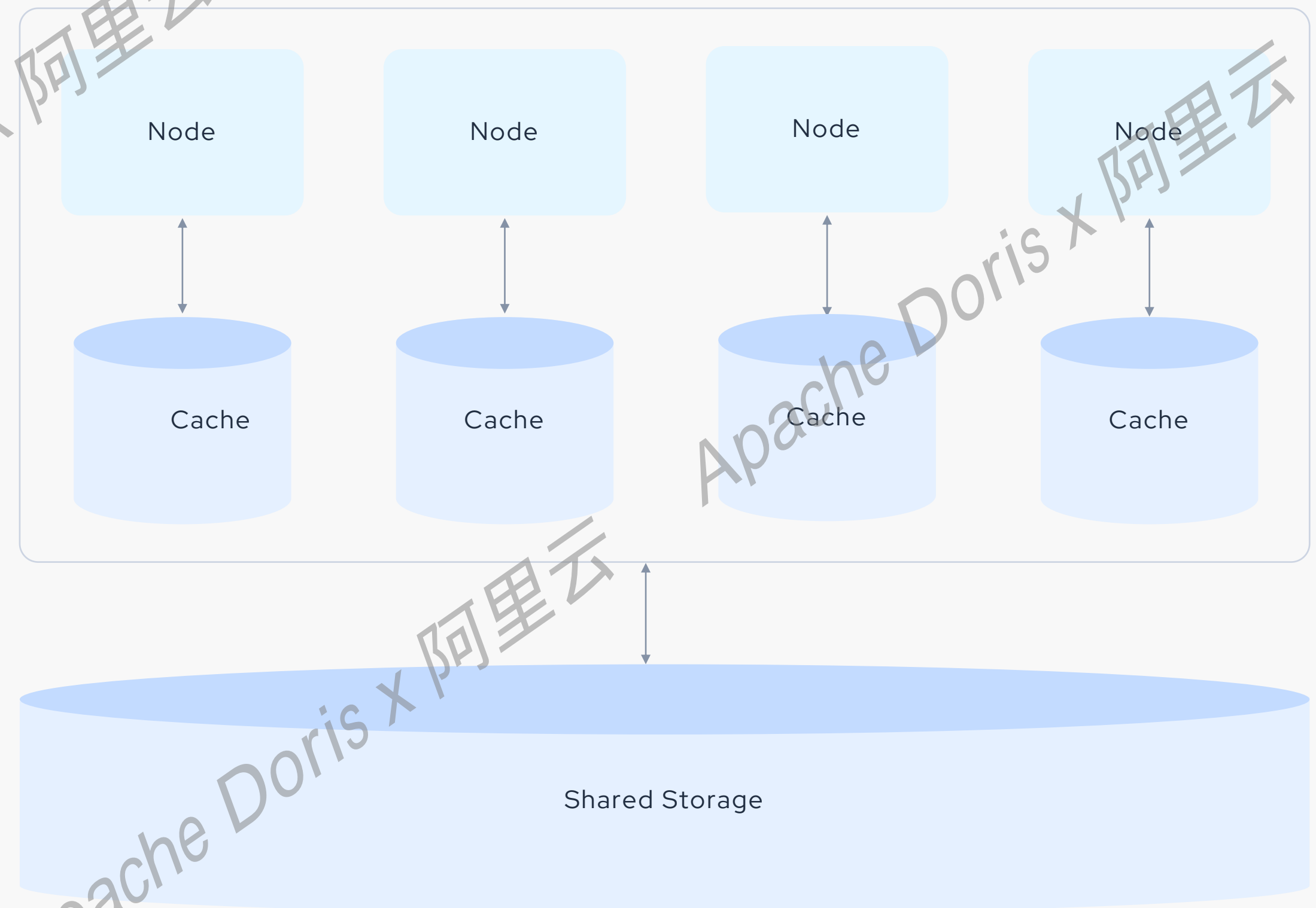
阿里云 SelectDB 云原生特性：存算分离降低 60%+ 成本

存储资源用量优化 30%+

- Doris：3副本冗余
- SelectDB：1副本+部分缓存（10%~100%）

存储资源单价优化 68%+

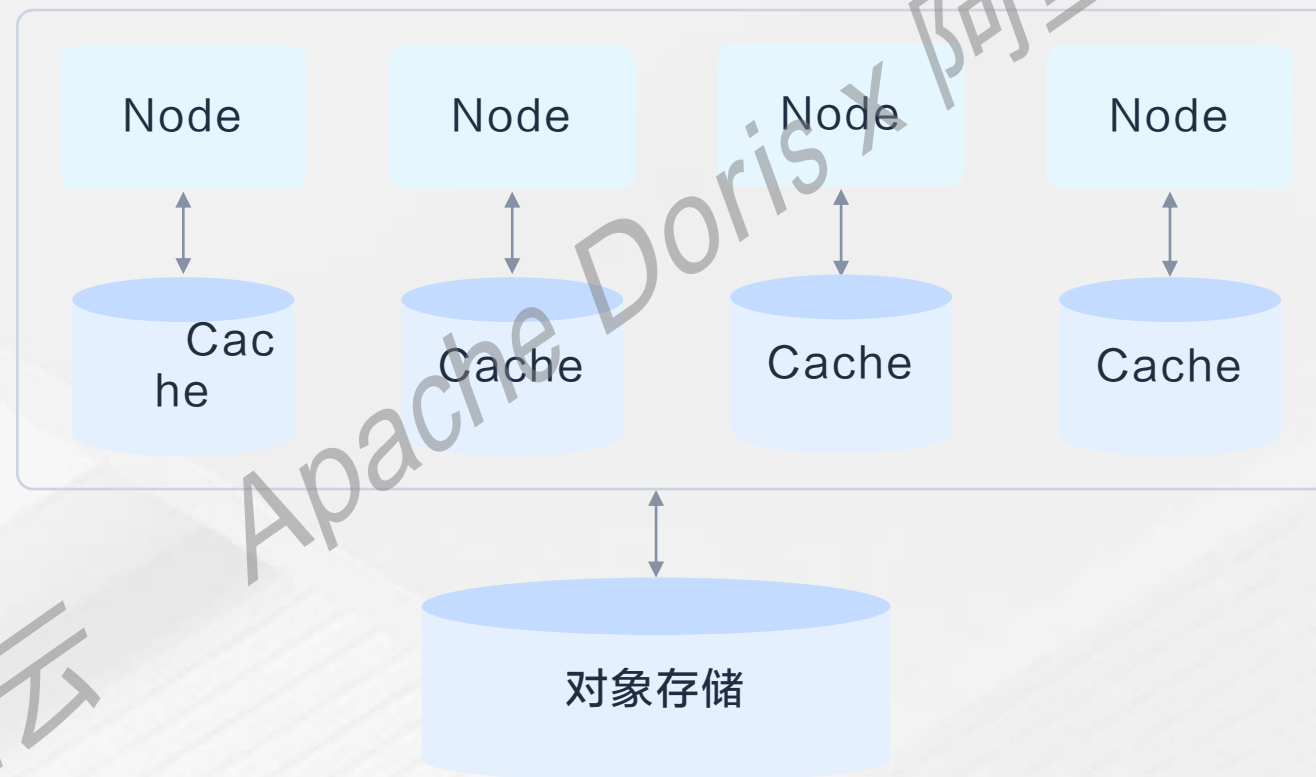
- Doris：标准云盘价格
- SelectDB：持久化对象存储，80%单价成本缩减



云原生架构 - 灵活弹性与资源隔离

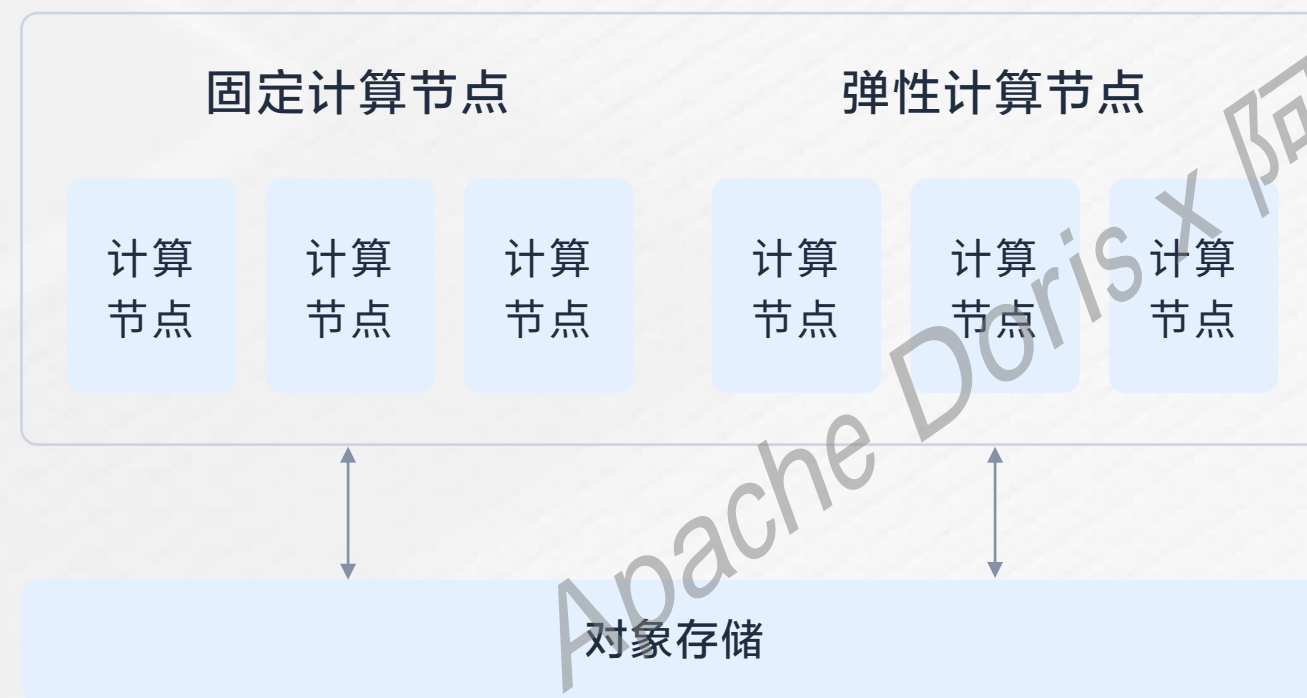
共享存储与本地缓存

- 对象存储：全部数据，按需使用，持久化
- 数据缓存：热数据，弹性扩容，查询加速
- 伸缩平滑：缓存的预热与迁移



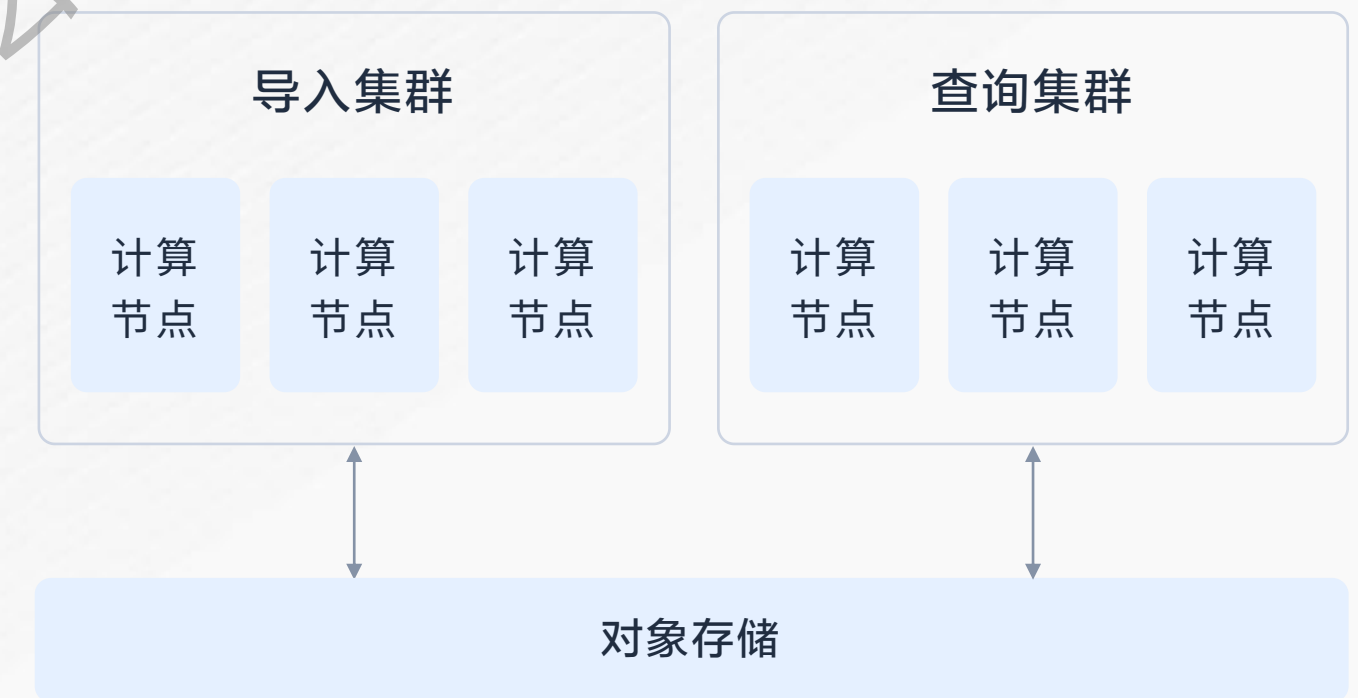
计算节点弹性伸缩

- 按需伸缩：计算和缓存
- 按需启停：节省计算成本
- 支持混合计费：包月、按量集群混合



多计算集群

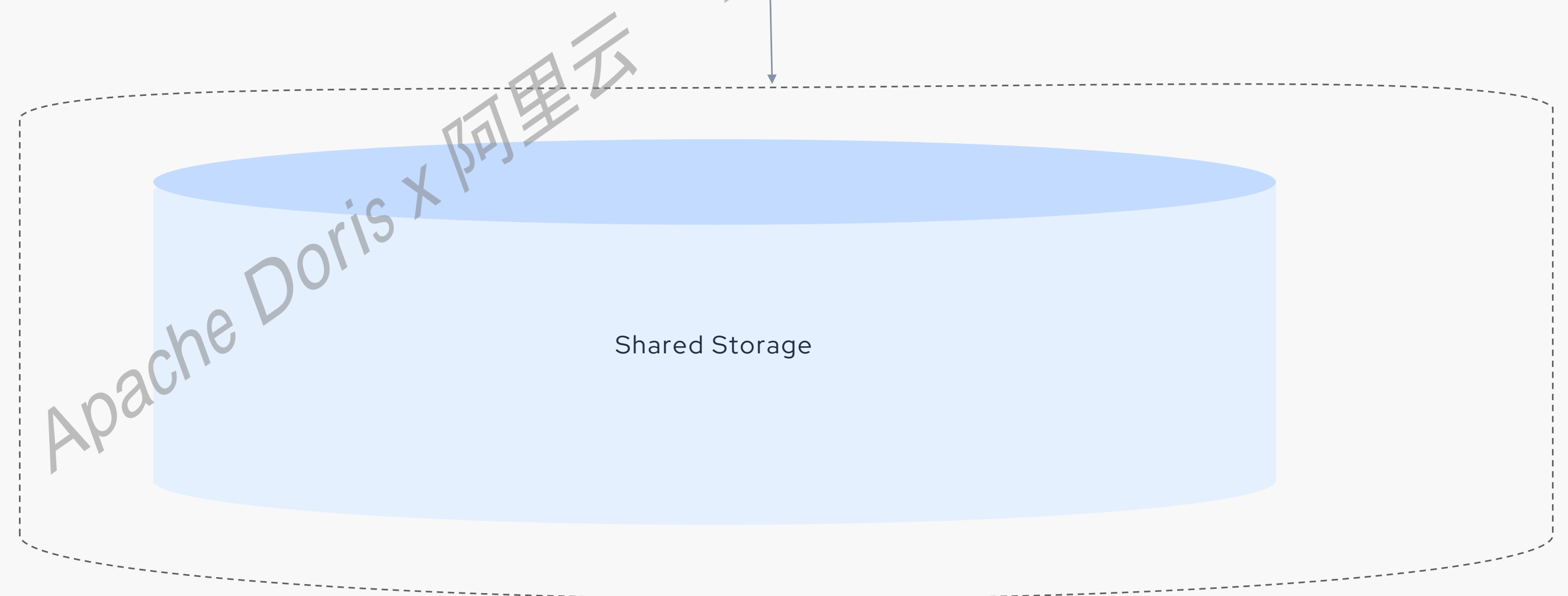
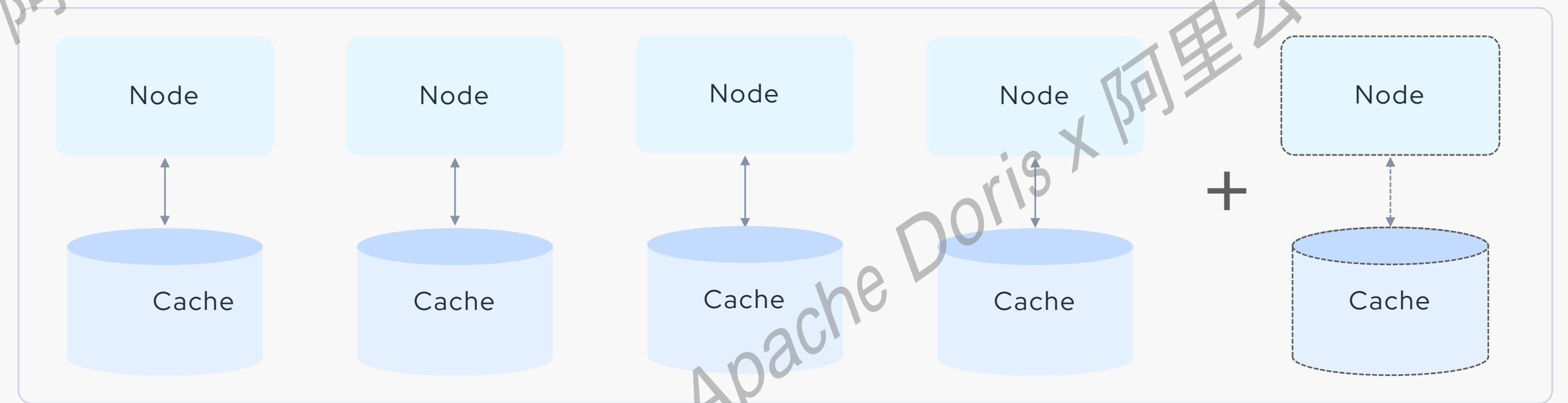
- 数据共享：元数据和数据均共享，强一致
- 多读多写：多集群均可以读写
- 计算隔离：读写隔离、在离线隔离等



阿里云SelectDB 云原生特性：无全量迁移的实时弹性

实例各层资源实时弹性表现

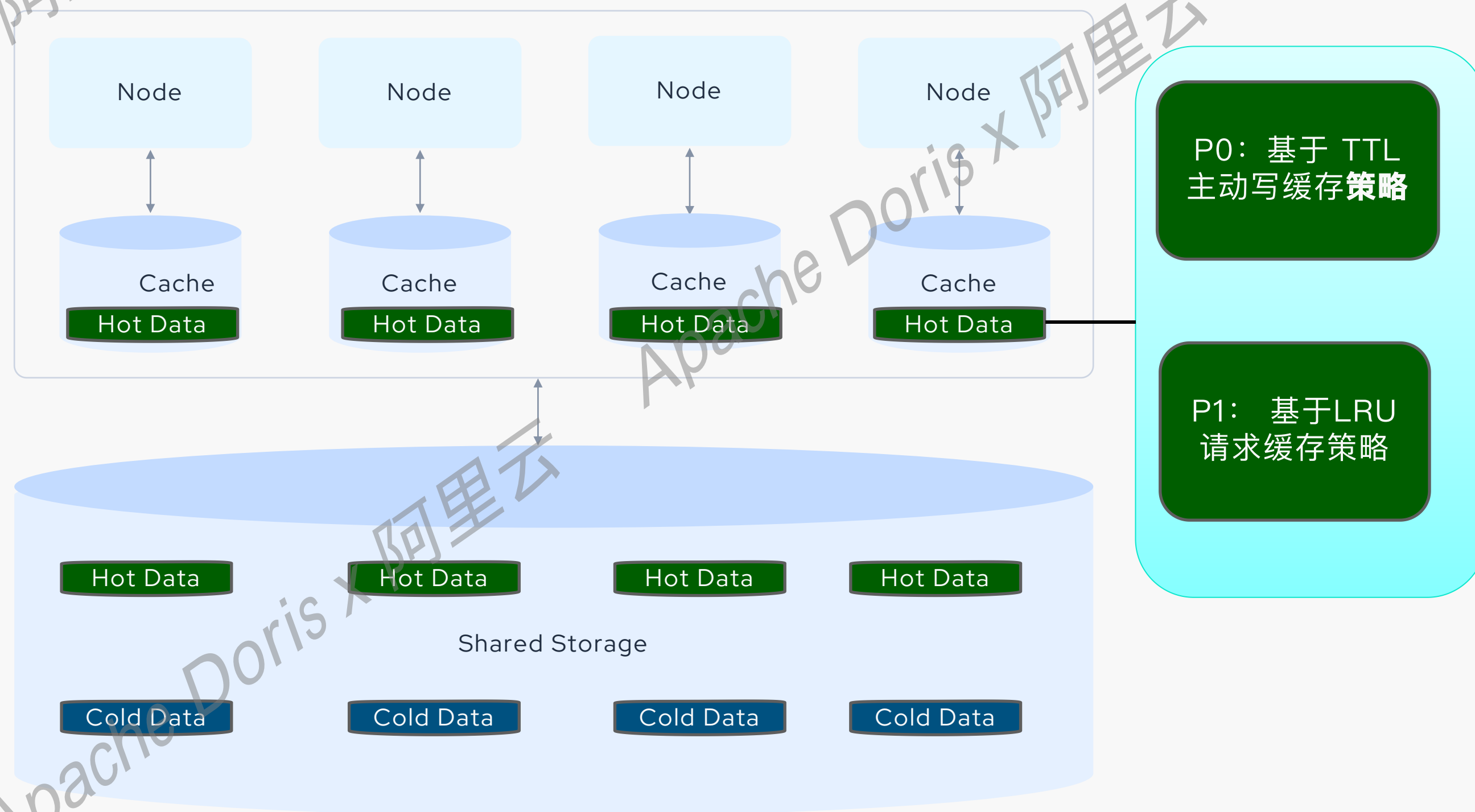
- 计算层：分钟级完成水平扩展和缓存迁移
- 存储层：Serverless 按需实时扩容
- 缓存层：基于云盘扩容，无感秒级



阿里云SelectDB 云原生特性：单副本写&读写Cache 加速性能

SelectDB 读写缓存策略

- 基于TTL 缓存策略，指定写入数据缓存时间
- 基于LRU的默认读写缓存策略，加速读写
- 指定表和分区预热，提高查询效率
- 单副本写入，降低IO，加速写入性能



阿里云 SelectDB 云原生特性：单实例多计算组资源隔离



隔离性和一致性保证

- 多集群元数据和数据共享，数据强一致
- 多个计算集群都可以读写
- “热数据”缓存数据预热

典型使用场景

- 导入工作负载与查询负载隔离
- 大查询负载和在线点查询的负载隔离

阿里云 SelectDB 开放生态和产品协同

基于 Apache Doris 构建，100% 兼容 Doris

- 与 Apache Doris 存储格式兼容，支持 SelectDB 与 Apache Doris 随时切换
- 与 Apache Doris 使用接口一致，兼容 Apache Doris 所有开源生态

业务数据库协同

- 支持从RDS, PolarDB, MongoDB秒级别导入到SelectDB, 实现HTAP一体化解决方案。
- 支持 DTS / DataX 实时和批量同步工具

兼容MySQL 协议，语法生态

- 支持 MySQL Client、JDBC、ODBC 等的连接
- 支持 Navicat、DBeaver 等工具连接
- 各类对接 MySQL 的 BI 工具都可以连接

SelectDB

大数据系统协同

- 大数据工具打通：Flink, Spark, Kafka, Dataworks, MaxCompute, Hive
- BI工具打通：QuickBI
- 机器学习：依托 SelectDB 基于 Arrow Flight 的高速数据读取接口

云数据库SelectDB版 - 4.0版本新特性概览

开箱盲测性能领先 (1+ 倍)

- 自适应的并行查询
- 统计信息和规划规则增强
- 导入性能优化

存储增强

- 自增列
- Auto Partition
- 实时 Insert

多表物化视图 (异步物化视图)

- 物化视图全量和增量刷新
- 物化视图直查和透明改写
- 内表物化视图和外表物化视图

查询增强

- 智能负载管理
- Decimal 增强
- Spill To Disk (实验特性)

半结构化数据分析

- Variant 数据类型
- 复杂数据类型分析函数完善

数据湖分析增强

- 数据湖回写
- Arrow Flight 高速数据读取
- 多 SQL 方言兼容 (Trino, Presto, Hive 等)

Part 3

阿里云SelectDB版案例介绍

数百家客户的选择

实时报表分析



头部在线教育公司的统一报表分析平台

- 面向客户的报表平台，典型的 **高并发报表分析 (customer facing analytics)**，对并发和响应延时要求高
- 采用 SelectDB 作为高并发报表分析解决方案，替代 **MySQL、Apache Doris**，快速解决客户线上并发痛点

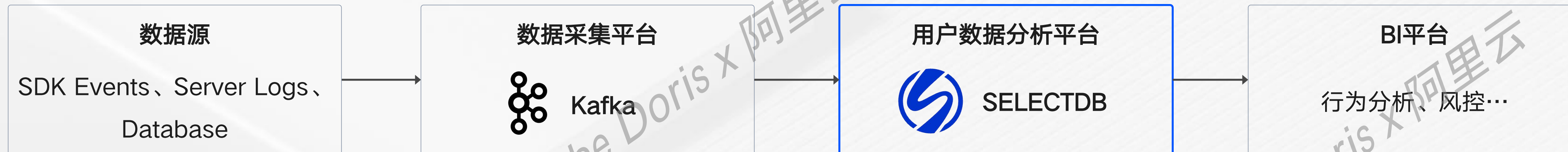
查询：**500w+ 查询/天**，99分位查询延时 **100ms**，查询峰值 **QPS 500~**，压测峰值 **1000+**

导入：数据流 **秒级** 延时，持续追加写入

头部互联网金融公司的实时特征平台

- 面向 ToC 客户的实时特征服务，要求实时更新背景下，保持高并发、低延时的查询能力
- 采用 SelectDB 作为实时特征查询服务，查询 **QPS 4000~**，响应延时 **20ms**，支持实时更新

即席分析



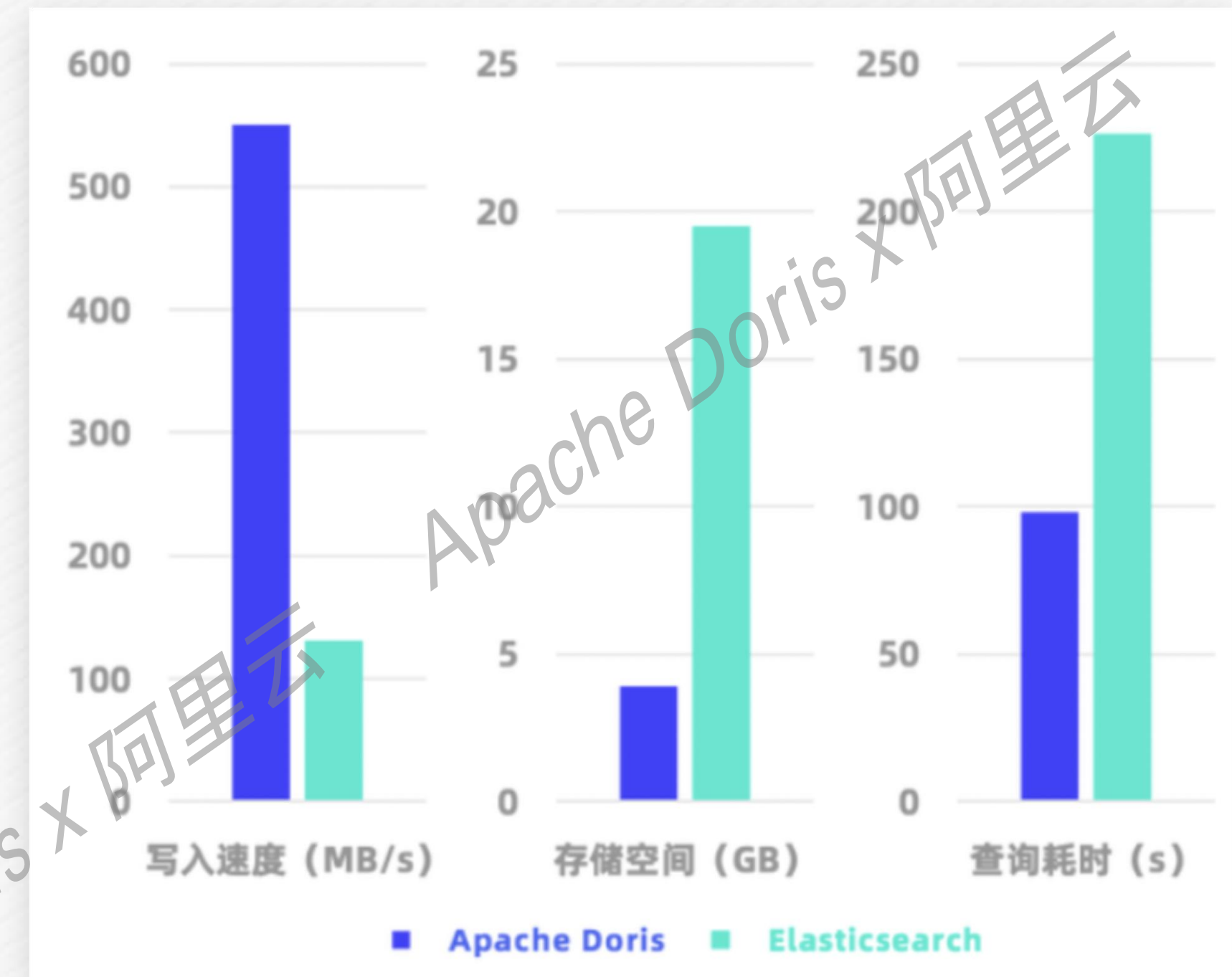
头部互联网内容分发公司的用户数据分析平台

- 面向公司全平台的用户行为与画像分析平台，满足事件、留存、漏斗、圈人等用户行为分析需求
- 使用 SelectDB 作为实时解决方案，替换原有离线架构(Kafka + Hive + Spark + MySQL)，“颠覆式提升数据分析效率”
 - 需求交付方式从 开发离线作业 转变为 BI工具自助分析，交付周期从 **7天** 下降至 **0天**
 - 十亿级用户的用户行为数据，数据延迟从 **1小时** 降低至 **15s**，查询耗时从 **分钟级** 降低至 **秒级**
 - 助力客户一个季度对接公司内 **20+** 条新业务，助力平台在全公司内快速扩张
- 目前已在海内外已扩展至 **10~** 集群、1000+核，业务仍在快速增长中，已成为公司统一的 用户数据分析平台

日志搜索分析

SelectDB 提供相比 Elasticsearch 10 倍性价比的日志搜索分析方案

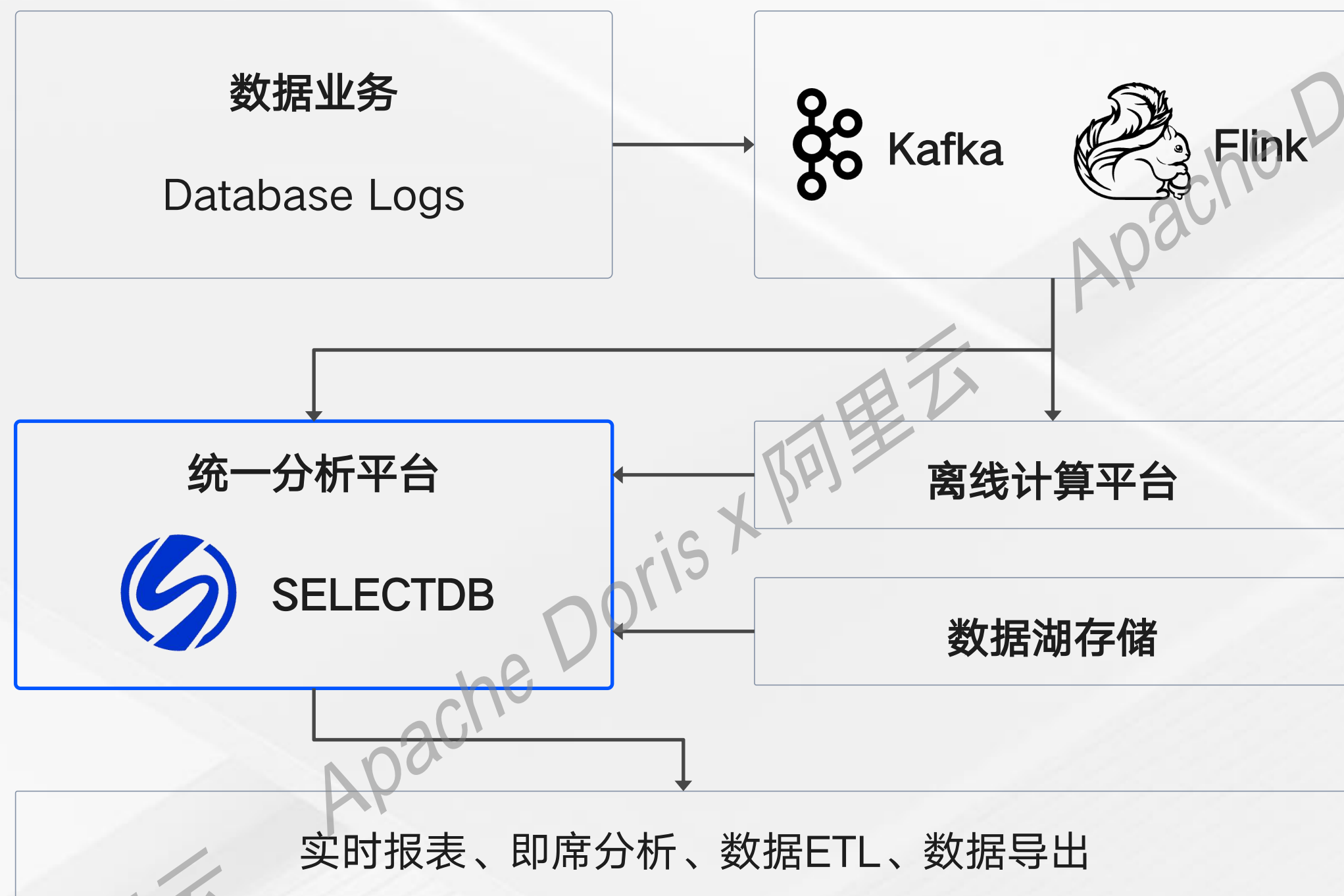
<p>2 倍 查询性能提升</p>	<ul style="list-style-type: none"> SQL 查询语言，支持 Join，相比 DSL 简单，表达能力强 高效的执行引擎和优化器，高效的索引与文本分析
<p>5 倍 写入吞吐提升</p>	<ul style="list-style-type: none"> 向量化指令，提升数据解析，索引构建的性能 简化去掉正排等索引结构，降低索引构建开销
<p>80% 存储成本降低</p>	<ul style="list-style-type: none"> 简化去掉正排等索引结构，减少倒排数据量 30% 列式存储和 ZSTD 高效压缩，提供 5-10 倍压缩比 冷热分层，大幅降低冷数据成本
<p>稳定性 大幅提升</p>	<ul style="list-style-type: none"> 资源隔离和大查询限制 高效内存管理，避免 Java GC 影响



用户场景从日志搜索分析从 Elasticsearch 到 SelectDB/Doris 的效果

- 某头部可观测性公司：使用 SelectDB 作为日志检索分析方案，提供完整的可观测性产品，资源成本降低 70%
- 某 AI 独角兽公司：每天新增数据量 400TB，总数据量 10PB+，支持公司全量用户访问日志检索与统计分析

统一分析平台



头部服装零售公司的统一分析平台

- 公司统一的大数据分析平台，支撑公司生产、销售、库存、渠道等全链路数据分析需求，包含高并发实时报表、即席查询、数据ETL处理、数据导出等需求
- 采用 **SelectDB** 作为统一分析平台，替代其原有组合式架构，大大**降低架构复杂性**，性价比提升 **3倍+**

实时报表：性能提升 **3倍+**，并发达 **6000**，响应延时 **20ms**

即席查询：性能提升 **1倍+**，响应延时稳定在 **1秒** 以内

ETL 处理：性能提升 **2倍+**，相比Spark类传统系统提升 **8倍+**

数据导出：性能提升 **2倍+**，10GB数据全量导出耗时 **30s级**

混合负载：持续稳定运行，报表可达 **4000+ qps**，即席查询平均响应时间在 **1秒** 内

- 支持客户千亿级的统一分析平台，可轻松扩展满足客户未来 10 倍级的业务增长