

基于阿里云 SelectDB 构建低成本、高性能的 湖仓一体化架构

阿里云技术解决方案团队

数据仓库的核心目标是助力商业智能（BI）决策，
将分散于各业务系统的结构化数据抽取、转换、加载（ETL）至集中存储库。

高效的
结构化
数据分析

- 严格的 Schema 设计
- 列式存储
- 专用数据结构
- 存算一体

成熟的
SQL
查询支持

- 基于关系模型的数据仓库完美适配 SQL 语言

数据
一致性保证

- 集中式的存储与管理
- 严格的事务机制

基于 Hadoop 的技术架构的数据湖

超大规模 数据处理

- 分部署计算
- 并行计算

多模态数据 与 低成本存储

- 以原始格式存储
- Schema-on-Read 机制
- 对象存储

多模态计算

- 共享数据集
- 多样化的数据处理需求

数据湖：原始数据汇聚成“湖”，湖上构建多样的数据分析系统

存储层

- 分布式文件系统或对象存储
- 无限存储容量
- 以原始格式存储

计算层

- 多引擎支持
- 批处理、流处理、交互式分析

元数据层

- 统一元数据服务
- 数据共享与协同处理

湖仓一体的演进：湖和仓的对比

数据仓库	功能	数据湖
高度结构化的数据类型	数据类型	结构化、半结构化和非结构化
专有格式、高度定制化	数据格式	开放格式，如 Parquet、ORC
预定义的 Schema（Schema-on-Write）。不可变更、写入成本高、读取成本低	Schema	无预定 Schema（Schema-on-Read）。写入成本低，读取灵活、性能低
ETL	数据摄入	ELT
专有存储系统。存储成本高	数据存储	对象存储、HDFS。存储成本低
SQL	用户接口	API、各种大数据处理系统
报表、BI	应用场景	数据分析、机器学习、探索式数据查询

湖仓一体的演进：数据处理的新挑战



数据仓库

- 数据实时性匮乏
- 半结构化与非结构化数据支持乏力



数据湖

- 性能瓶颈
- 事务性保证缺失
- 数据治理危机

湖仓一体新范式



存储层：坚实基座

- 以 HDFS 或云对象存储为基础：低成本、高可用
- 高效文件读写：列式存储、索引



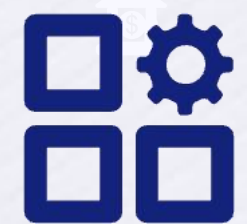
开放数据格式：灵活互通

- 开放文件格式：Parquet、ORC
- 开放表格式：Iceberg、Paimon、Hudi、Delta Lake



计算引擎：多元协同

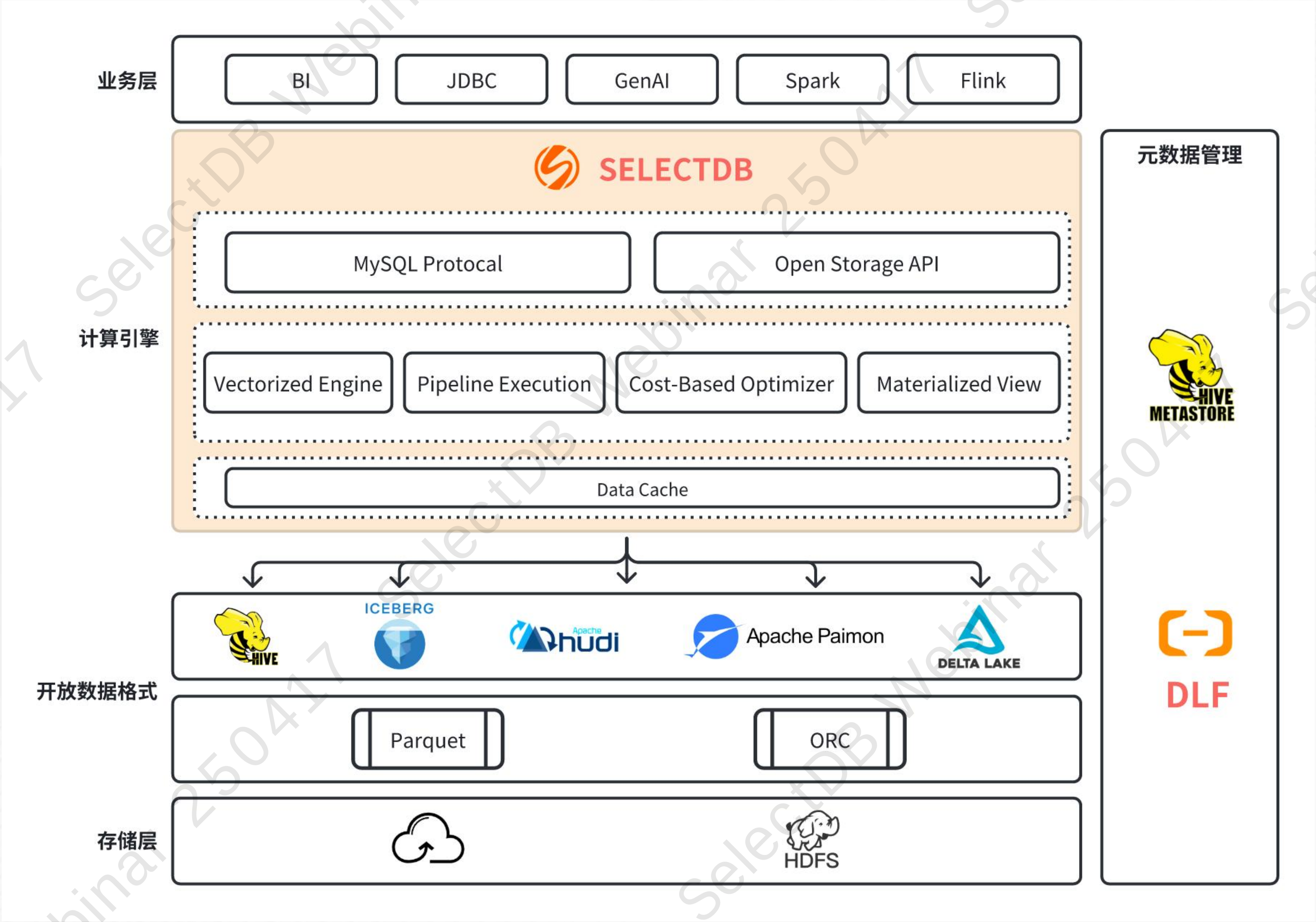
- 批处理：Spark
- 流处理：Flink
- 交互式分析：SelectDB



元数据管理：统一中枢

- 统一的命名空间
- 全局的数据目录
- 统一权限管理

SELECTDB 湖仓一体新范式



阿里云 SelectDB 版 对比

Trino/Presto 方案

- 增量更新 & 透明加速
- 复杂查询精准优化
- 开箱即用
- 适用场景丰富

Trino/Presto	VS	SelectDB
功能单一，依赖外部组件 仅支持手动和全量的物化视图刷新，并且依赖开放湖格式提供存储能力。	物化视图	✓ 增量更新和透明加速 高效支持各个数据源的物化视图构建，以及物化视图的增量更新和查询透明加速能力，极大提升查询效率。
复杂查询优化效果差 基于代价和规则的查询优化器，支持复杂SQL优化，但不支持统计信息的自动收集。	查询优化器	✓ 复杂查询精准优化 基于代价和规则的查询优化器。拥有丰富的统计信息自动收集和采样能力，能够获得更精准的查询规划。
依靠第三方组件 依赖外部组件如 Alluxio 提供缓存能力。	缓存能力	✓ 开箱即用 内置基于内存和本地高速磁盘的数据缓存能力，并拥有丰富的缓存预热与淘汰策略。
适用场景单一 仅作为查询引擎。	应用场景	✓ 适用场景丰富 云原生数据仓库，内置高性能表格存储，既可支持低延迟的高并发查询，又可作为湖仓查询引擎进行交互式查询分析。

SelectDB 湖仓一体：打破数据边界

灵活的数据接入

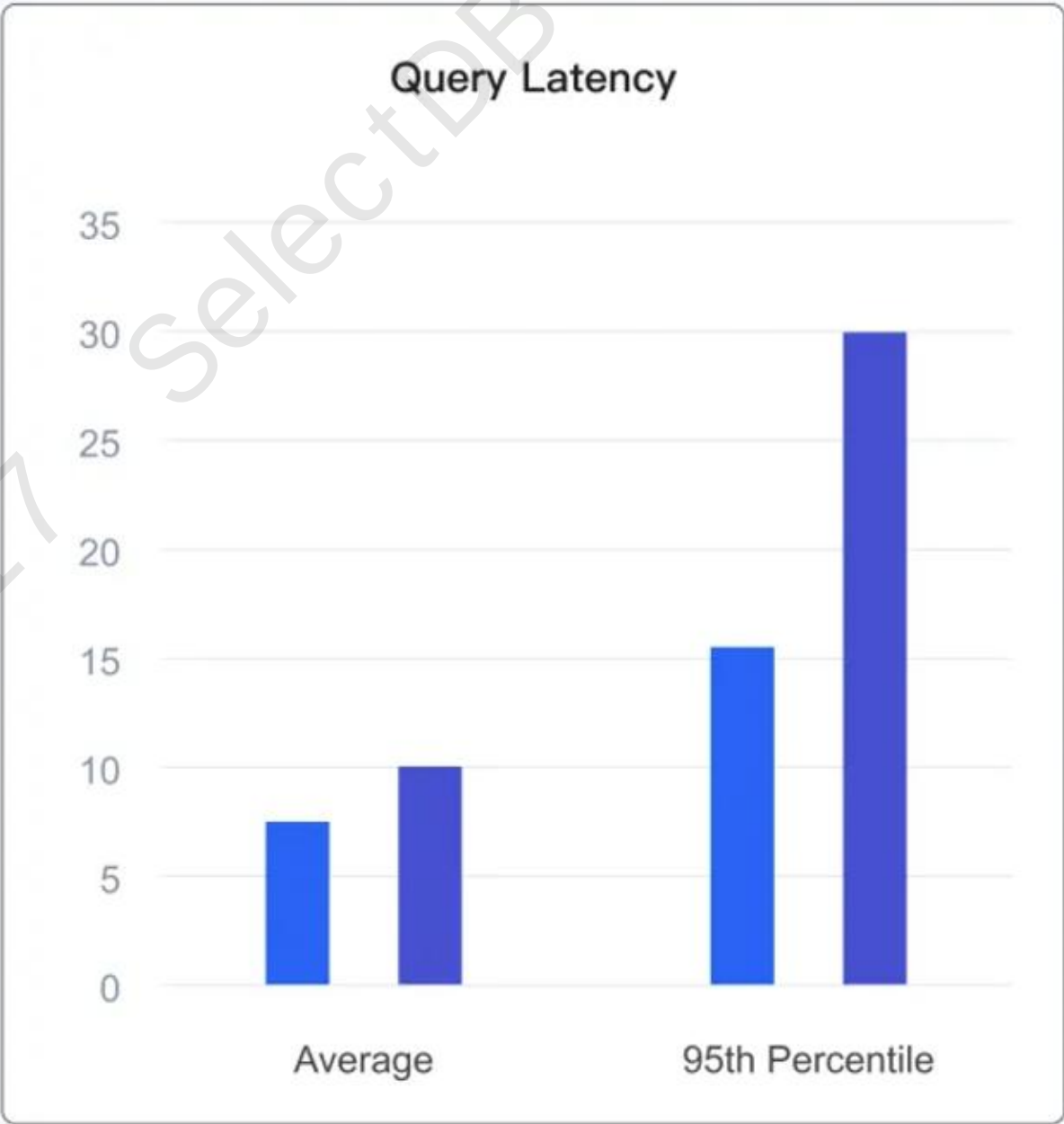
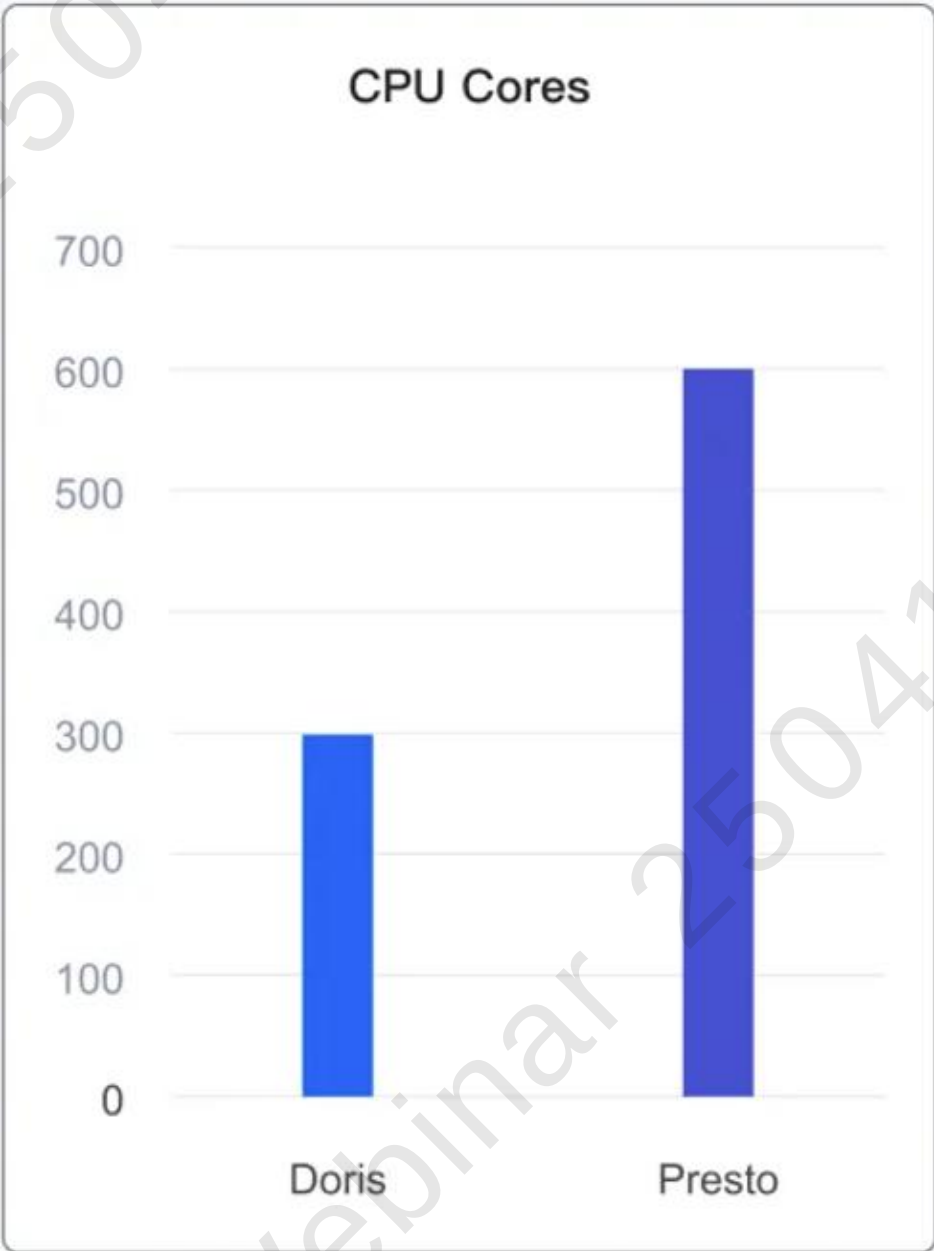
- 丰富的数据源连接器
- 跨源联邦分析

湖仓系统	数据库	文件格式	元数据服务	数据存储
- Hive - Iceberg - Hudi - Paimon - Delta Lake - Kudu - Max Compute ...	- Aliyun RDS - MySQL - PostgreSQL - Oracle - SQL Server - IBM Db2 - ClickHouse - SAP HANA - OceanBase - Elasticsearch ...	- Parquet - ORC - Text - Json ...	- Hive Metastore - Aliyun DLF - Iceberg Rest Catalog - Filesystem ...	- HDFS 及兼容存储 - 阿里云 OSS ...

SelectDB 湖仓一体：打破数据边界

高性能的数据处理

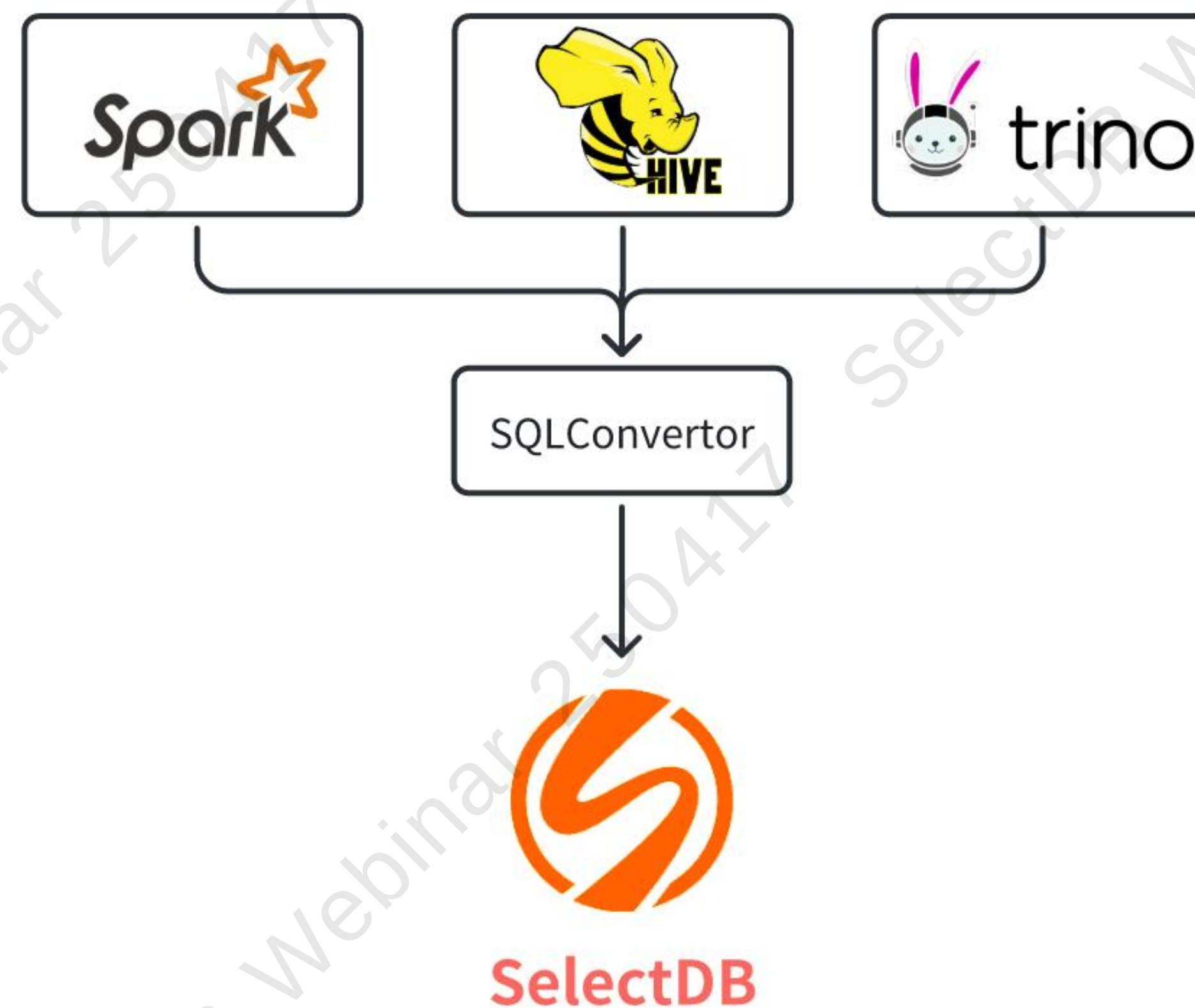
- 向量化执行引擎
- 现代化查询优化器
- 缓存加速与 IO 优化
- 物化视图与透明加速



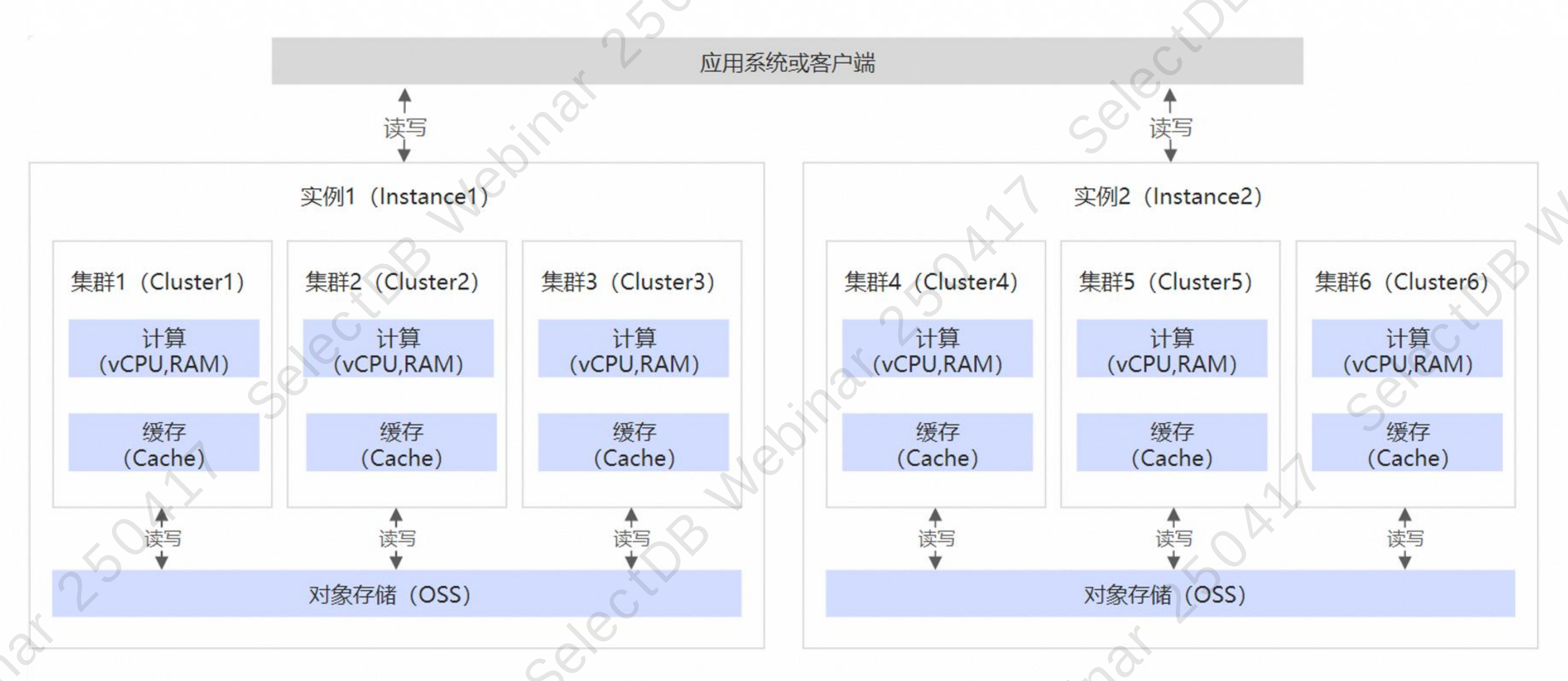
SelectDB 湖仓一体：打破数据边界

便捷的业务迁移

- **SQL 方言转换：**无需业务修改 SQL。
- **支持多种方言：**Hive、Spark、Trino、...



SelectDB 湖仓一体：打破系统边界



现代化的部署架构

- **低成本存储：** 存储和计算资源可独立按需扩展
- **唯一可信来源：** 同一份数据供不同的计算集群访问和处理
- **负载多样性：** 动态调配计算资源

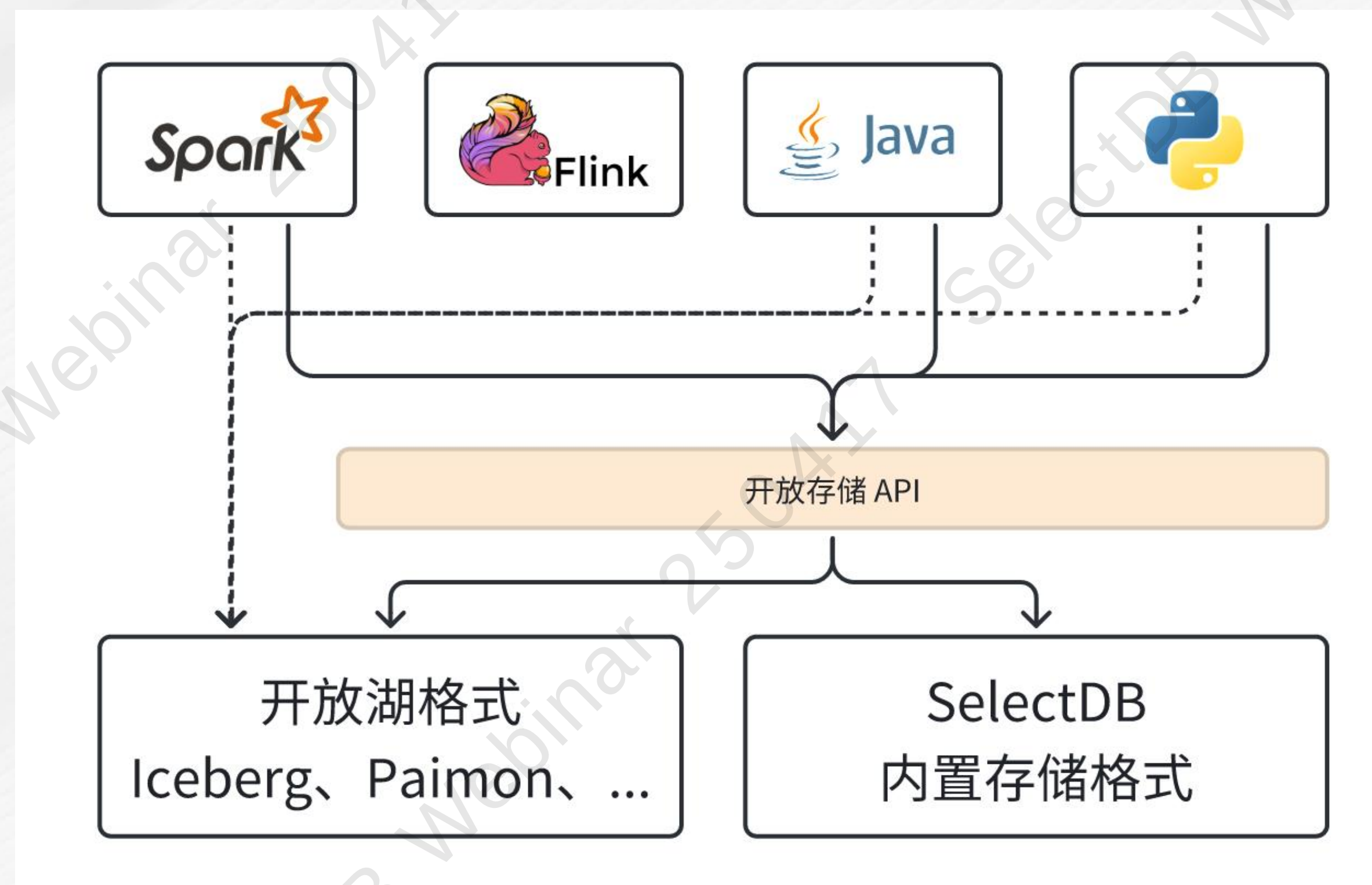
丰富的数据存储和管理功能

- **数据湖格式与内置存储：**同时支持开放湖格式和高性能内表格式的读写操作。
- **半结构化数据支持：**原生 Variant 类型，提供 Schemaless 的数据管理体验。
- **数据更新的支持：**近实时的数据更新能力。基于部分列更新能力，支持多流合并宽表的操作。
- **丰富的数据索引：**前缀索引、倒排索引、跳数索引、Bloom Filter。
- **实时数据和批数据写入：**行高频数据写入或批量加载。

SelectDB 湖仓一体：打破系统边界

开放存储 API

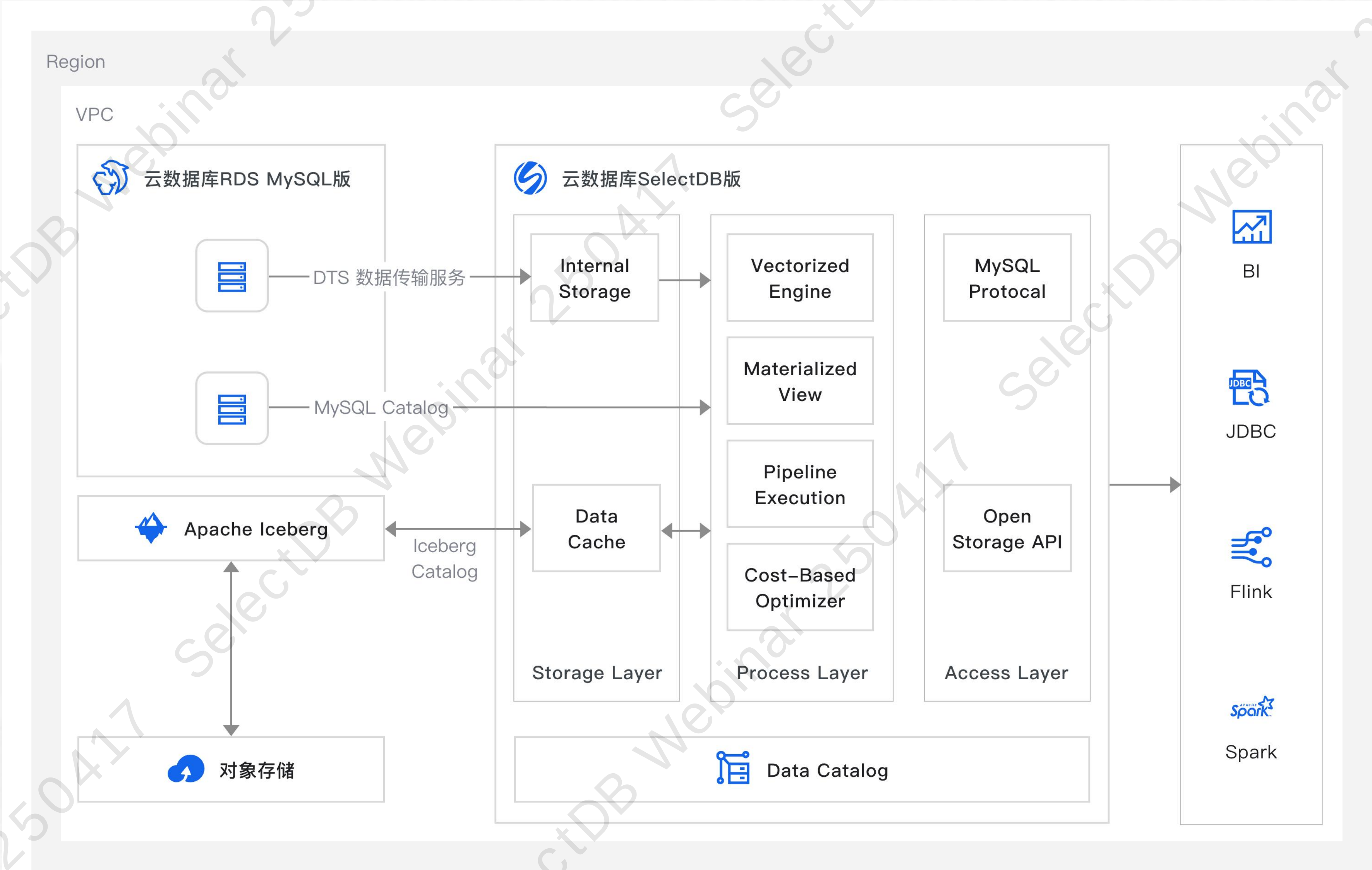
- **高速数据链路：**基于 Arrow Flight
- **开放性：**多引擎、多语言支持
- **一致性：**屏蔽底层格式细节，统一读写行为



典型场景基于 SelectDB 湖仓一体的最佳实践

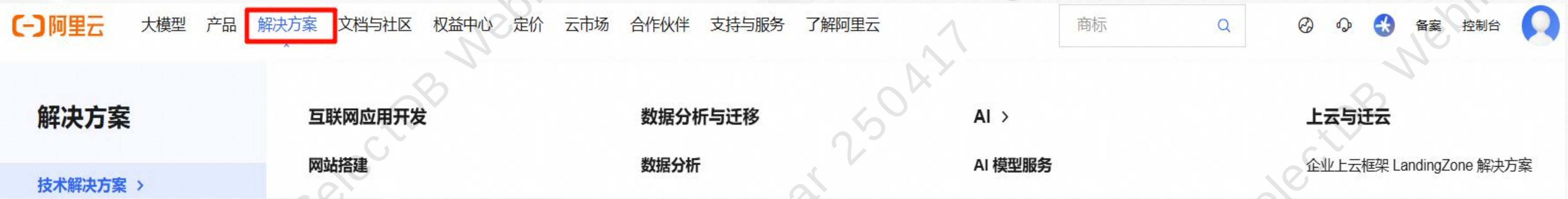
三大典型场景最佳实践

- 湖仓分析加速
- 湖仓数据处理
- 多源联邦分析

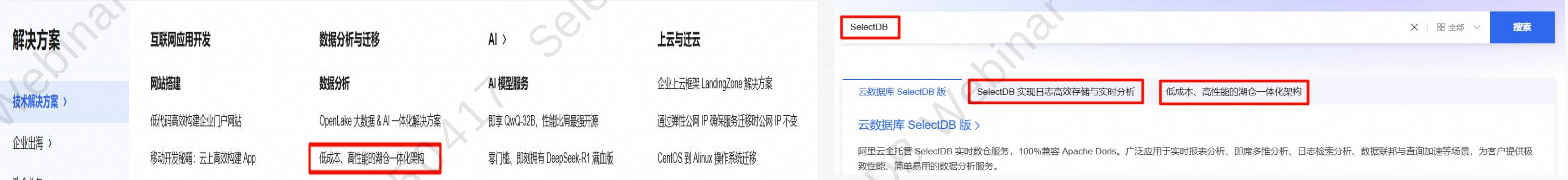


立即部署体验阿里云 SelectDB 解决方案

01 进入 aliyun.com，点击“解决方案”



02 在解决方案下拉菜单中找到“低成本、高性能的湖仓一体化架构”，您也可以直接在搜索框中输入“**SelectDB**”直接搜索。



03 进入解决方案，点击“立即部署”，即刻体验阿里云 SelectDB 版。

< 技术解决方案

SelectDB 实现日志高效存储与实时分析

企业级日志数据具有数据量巨大、写入和查询速度快、结构多样的特点，本方案基于阿里云云数据库 SelectDB 版构建高性能、低成本、分析能力强大的日志存储与分析解决方案，覆盖运维监控、安全审计、业务分析等场景，并通过智能索引与分级存储实现数据亚秒级检索。

立即部署

联系咨询

适用客户

✓ 需要大幅降低日志存储分析成本的用户

✓ 需总量PB级、日增量百TB或更多海量日志存储的用户

✓ 需要高效存储分析日志中的 JSON 动态数据的用户

立即部署体验阿里云 SelectDB 解决方案



奥运会全球指定云服务商

扫码立即畅享技术方案



湖仓一体解决方案



日志存储与分析解决方案

Thanks !