

Apache Doris

在电信行业场景下的应用与实践

王加雷 18813009343

浩瀚深度 大数据高级研发工程师

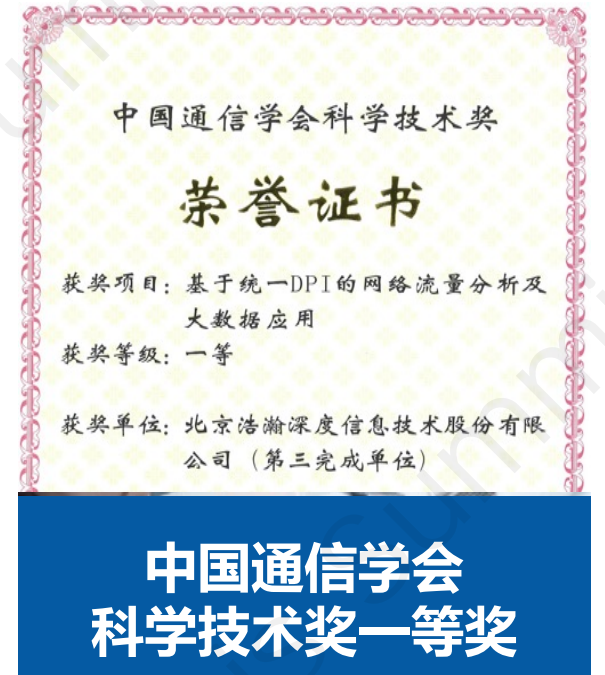
目录

1. 背景介绍
2. 架构演进
3. 基于 Doris 的应用实践
4. 未来展望

1 背景介绍

浩瀚深度公司介绍

- 浩瀚深度**成立于1994年**，20多年来坚持“自主创新、质量至上、极致卓越”的原则，敢为人先、砥砺前行，与2022年8月18日在上海证券交易所科创板上市，**股票代码：688292**。
- 作为**互联网流量和数据智能化的领航者**，浩瀚深度持续探索新技术、新业态、新模式，实现了**网络可视、资源优化、智能管控、安全防护和数据价值**，并加速推动互联网大数据的广泛应用，是一家集软硬件产品研发、生产、销售和服务于一体的高科技企业。
- 截止到2023年，监测互联网带宽超过**800Tbps**，覆盖三大运营商31个省市，**国内DPI市场份额大于40%**。



基于电信级大数据的顺水云平台

业务规模

- 单集群日均处理话单量超过800TB
- 存储规模超过100PB+
- 覆盖全国的10000+计算节点，50w+CPU逻辑核，2PB+内存

业务场景

- 海量用户上网日志存储与快速查询（工信部考核）
- 电信骨干/省网/城域网/IDC出口的流量质量分析
- 面向政企市场的用户画像&精准营销类场景

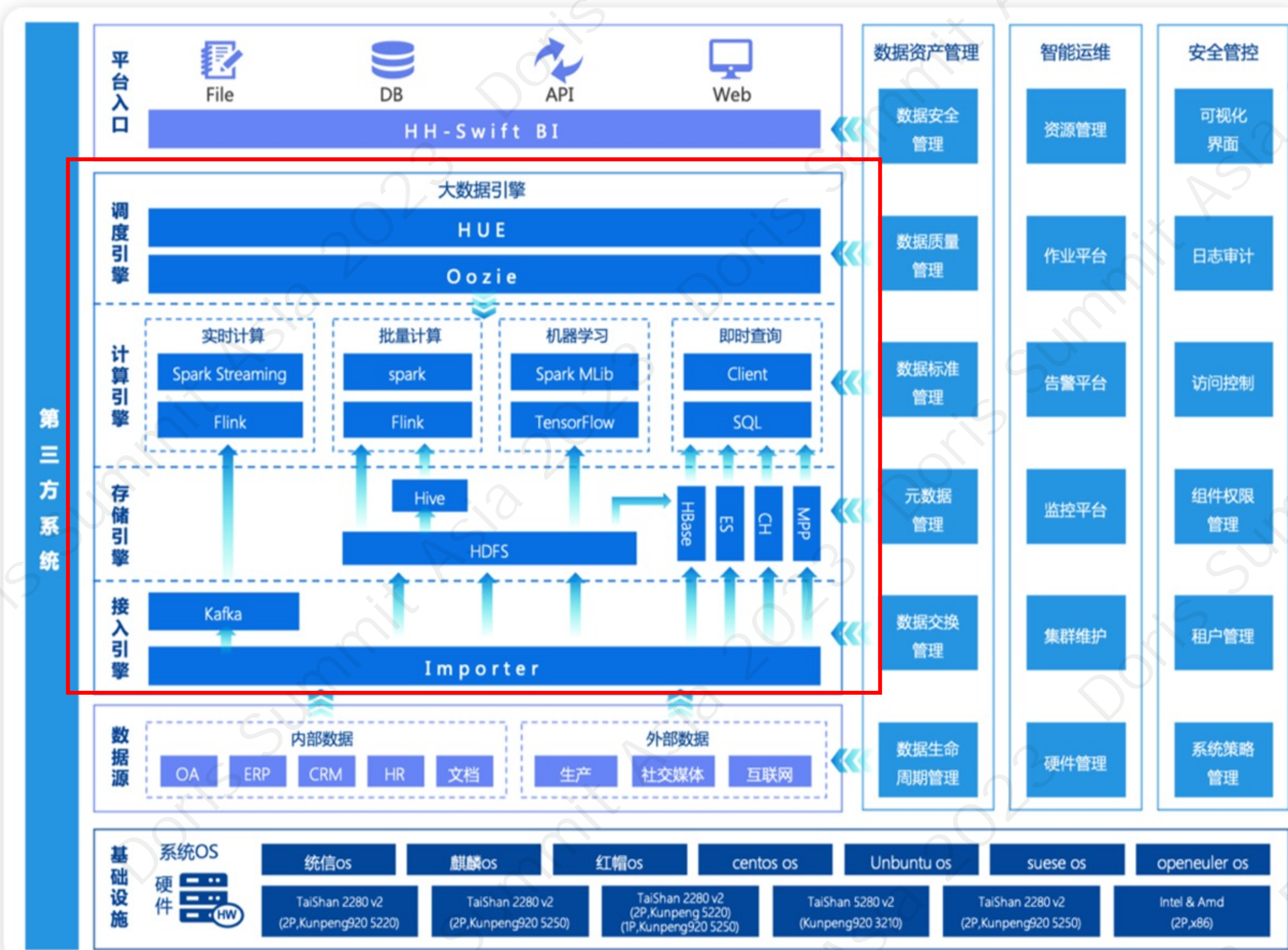
核心业务诉求

- 高吞吐的数据写入
- 海量数据的关联查询
- 数据的高压缩比与高可靠存储
- 亿级数据的秒级响应



2 架构演进

基于 Hadoop 生态的 1.0 架构



技术积累

- 开发200+的spark报表作业，每日1000+的作业调度
- 基于spark实现TB级别数据的关联查询
- 完善了大数据平台周边产品体系，包括自研的运维管理、安全管控、数据治理等产品工具

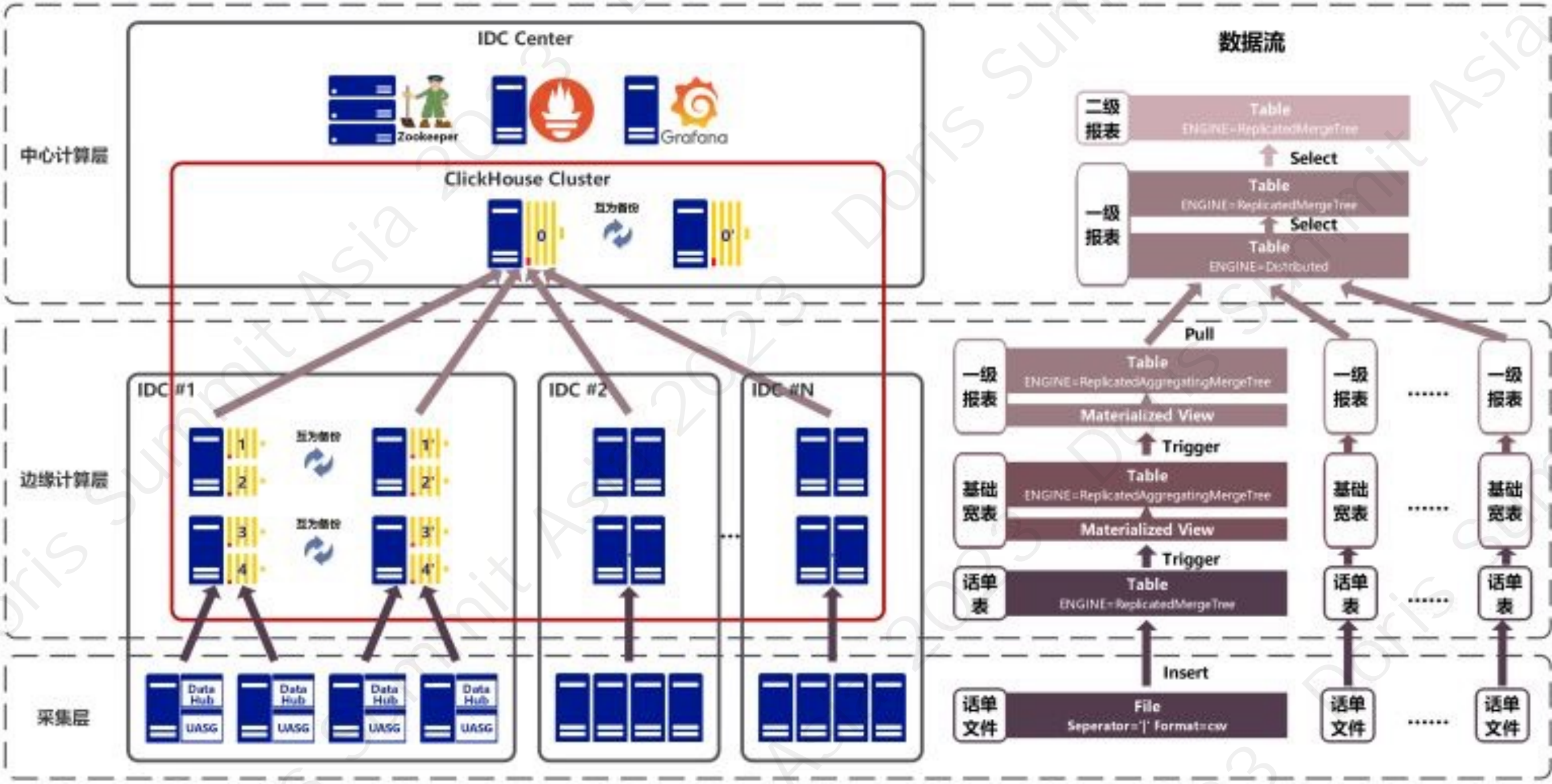
业务场景

- 实现了PB级别的用户上网日志存储和查询，满足工信部要求和移动集团规范
- 实现了TB级别数据的准实时关联合成，并共享给下游厂商

架构痛点

- Spark作业启动慢，无法满足更实时的应用场景
- 报表数据的查询响应时延较大
- Hadoop生态组件众多，维护成本过高

基于 Clickhouse 的 2.0 架构



技术积累

- 针对Clickhouse集群模式**负载均衡写入**的优化
- **多级物化视图**的统一管理
- 基于Prometheus+Grafana搭建的**统一的监控告警**体系
- 原生**内置函数**的扩展和开发，重新编译版本

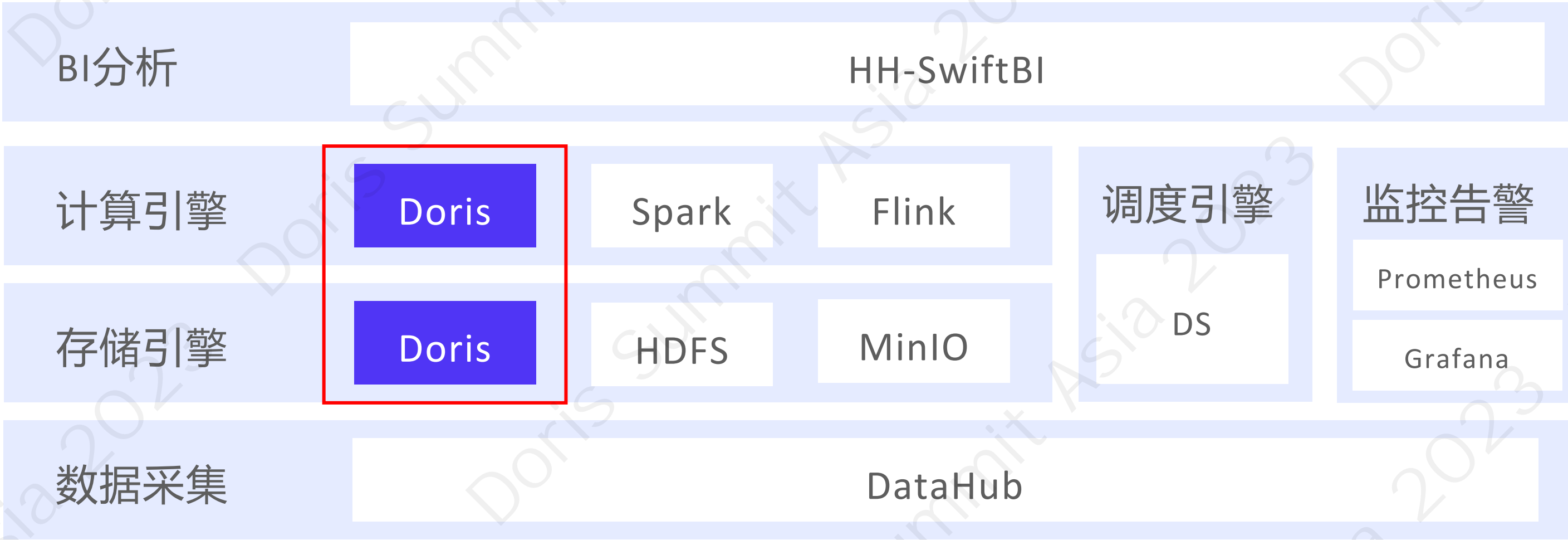
业务场景

- 覆盖某省**20+IDC机房**的**90Tbps**链路带宽，日均处理话单量**400TB**
- 利用物化视图在边缘计算节点进行预聚合，统计分析粒度降到**分钟级**，**节省一半的服务器资源**

架构痛点

- Clickhouse基于MergeTree的引擎写入吞吐量较低
- Clickhouse副本扩容之后不会自动进行数据的balance
- 在压力较大时，基于zookeeper的元数据管理机制出现故障几率较高

基于 Doris 的 3.0 架构



统一简化的系统架构



FE/BE的水平扩展能力



高速的数据查询能力

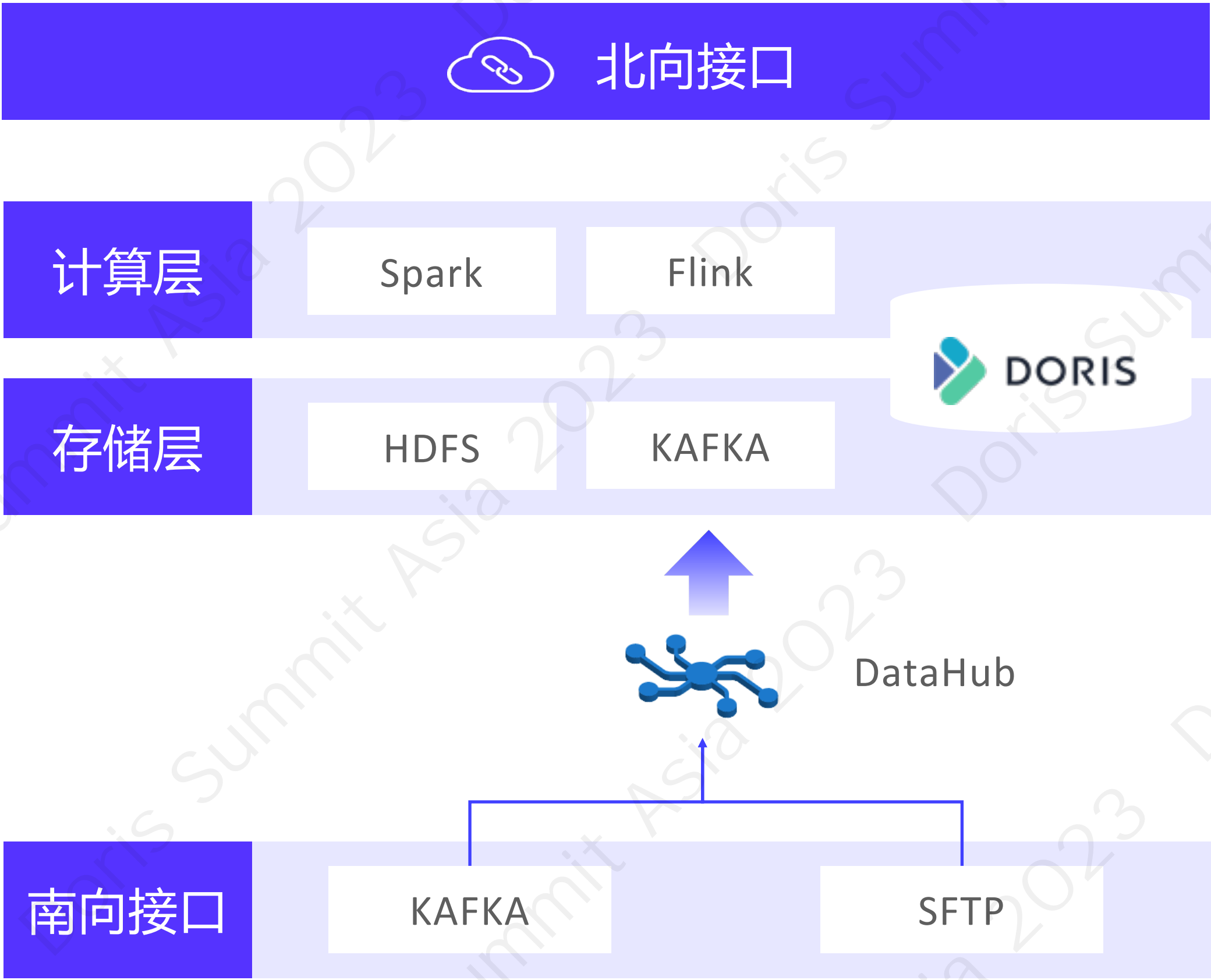


高吞吐的数据写入能力

选择 Doris 的若干原因

3 基于 Doris 的应用实践

基于 Doris 实现的实时数据合成共享平台



- 单个BE的写入使用StreamLoad多并发写入的方式，与之前CH相比在写入性能上有40%以上性能提升
- 在Doris中对于不同类型的列使用不同的压缩算法来提升压缩比，比如URL和上网账号可使用ZSTD压缩，普通数值用zlib，优化后的压缩比可以达到6:1甚至更高
- 结合电信行业使用场景，将常见的IP地址转换、IP地址段查找、资源匹配等逻辑用外部UDF的方式进行了实现，INSERT+SELECT效率与之前的CH相比不相上下
- 通过Doris提供的快速关联查询能力，在合成率保持不变的情况下将话单合成时延由之前的15分钟降到5分钟以内

4 未来展望

携手同行，共筑数字大厦的根基

使用 建议

- **进一步丰富内置函数**：丰富的内置函数可以极大降低SQL编写的复杂度，并增加数据库的使用场景。
- **增强物化视图的能力**：可以在物化视图中使用关联查询等语句，满足更多实时计算的场景。
- **探索DB For AI的能力**：可以进一步降低AI和LLM的使用门槛，与业界其它数据库能力对齐。

未来 展望

- 尽快将基于Doris的方案在现网项目进行推广，带来实际的项目收益。
- 将公司自研数据集成工具DataHub与Doris进行适配。
- 积极参与Doris社区活动，贡献最佳应用实践，协助解决产品问题。
- 与Doris社区合作，共同打造优秀的行业案例。



获取更多社区动态与最佳实践

Apache Doris 官方平台:

- Apache Doris 官网: doris.apache.org
- Apache Doris GitHub: github.com/apache/doris/

获取更多峰会资料:

- Doris Summit 峰会官网: doris-summit.org.cn
- Doris Summit 峰会回放: <https://space.bilibili.com/1196172099/channel/collectiondetail?sid=1824324>