

无锡锡商银行基于 Flink + Doris 构建实时数仓实践

施金才

无锡锡商银行 大数据技术经理

分享嘉宾 – 无锡锡商银行



施金才

信息科技部 大数据技术经理

10 年的大数据研发经验

现主要负责大数据实时数仓开发相关工作，主导建设了锡商银行飞流实时计算平台，并从 0 到 1 搭建数据服务体系。

目录

01 业务背景

02 实时计算架构演进

03 基于 Flink+Doris 架构下的应用实践

04 总结与展望

01

业务背景

无锡锡商银行介绍



无锡锡商银行坚持恪守信用、稳健经营的理念，坚定科技普惠、错位发展的方向，牢记普惠金融、八方共赢的使命，以移动金融和物联网等技术为手段，以服务中小微企业、服务科技创新、服务城乡居民为己任，让产品易知、服务易得、实惠易享，打造具有物联网特色的科技型银行。

需求场景

风控

信贷反欺诈
多渠道在贷分析
客户交易行为分析
还款计划和还款表现

运营

运营监控
实时报表
贷后管理



营销

年终账单
自动化触达

后管查询

订单查询
客户信息查询
授信信息查询
借据信息查询

需求特点

时效性要求

- 希望获取实时或准实时数据，便于运营和决策
- 大部分需求时效在分钟级即可接受
- 少部分需求可接受小时级

接口要求

- 接口调用或者查询响应要足够快，响应结果秒级返回，体验要好
- 满足高并发点查需求
- 满足多条件组合查询
- 满足客户圈选需求

稳定性要求

- 部分数据查询嵌入业务流程，接口响应必须稳定，不能影响正常业务
- 变更过程需要做到业务系统无感

02

实时计算架构演进

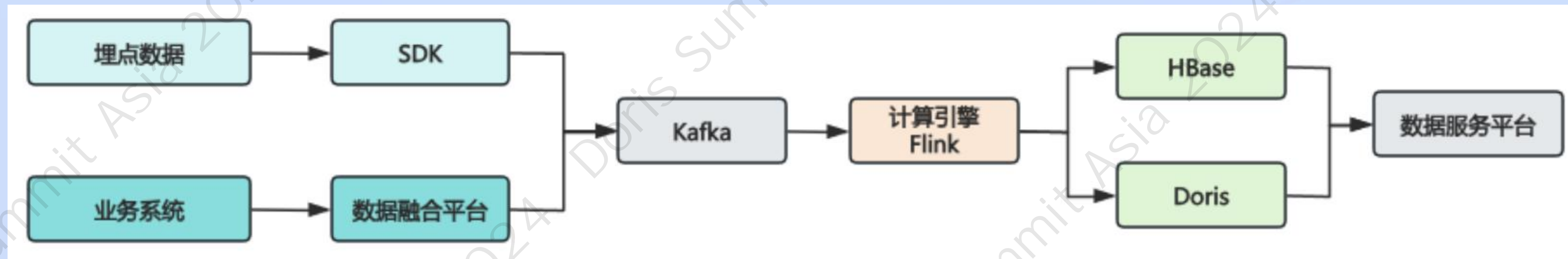
锡商银行实时计算演进



实时计算早期架构与痛点

逐步引入 Flink、HBase、Phoenix、Dinky、Doris 等平台或组件，实现实时开发能力。在未建设实时数仓之前，为了满足业务方实时数据需求，采用按需开发的策略，不同的研发组都有自己的实时任务维护，Doris 在该阶段仅作为数据服务的数据库使用。早期架构的局限迫使我们尽快改造升级：

- 1、运维复杂：组件过多，维护难度大，架构不统一
- 2、重复开发：按需开发，开发成本高，数据无法复用，响应周期长
- 3、资源利用：任务独立运行，资源利用率不高
- 4、数据质量：不同需求独立开发，数据一致性、准确性难以保证，数据稽核难度大



Apache Doris 在实时开发中的优势

1 优秀的架构设计

- 架构简洁，只有 FE 和 BE 两个进程
- 支持滚动升级和在线扩缩容，运维成本低
- 丰富的数据同步方式，与 Flink 引擎契合度好
- 支持联邦查询，与Hive数据交互更加便捷

2 低学习成本

- 兼容 MySQL 协议，支持标准 SQL，极大降低学习成本和使用成本

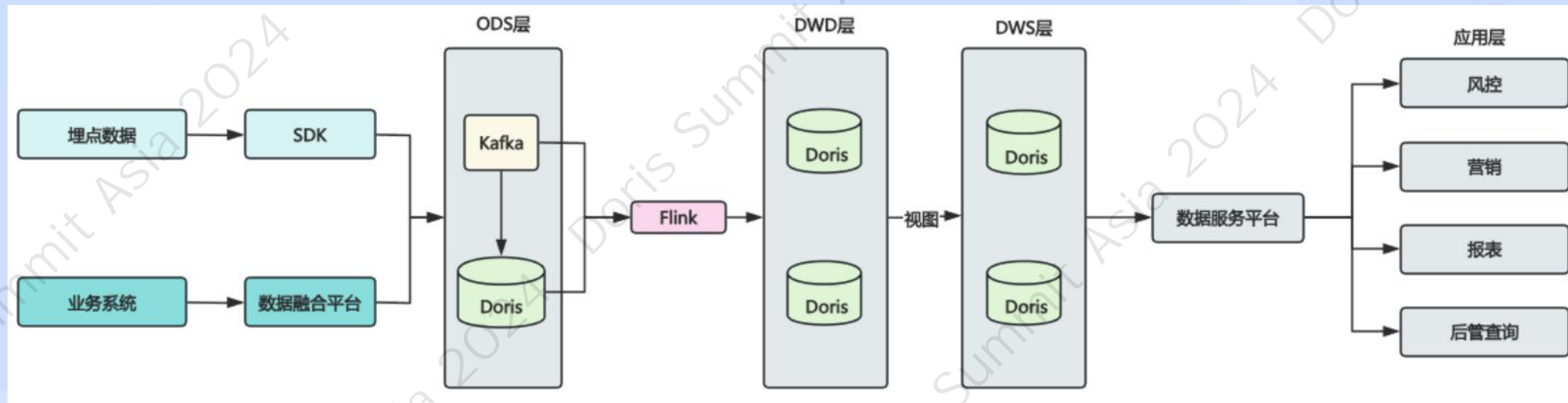
3 高查询性能

- MPP架构、向量化引擎、支持灵活构建索引、列式存储引擎保证了优秀的查询性能
- 支持高吞吐、高并发、低延迟的查询服务，契合大部分查询场景

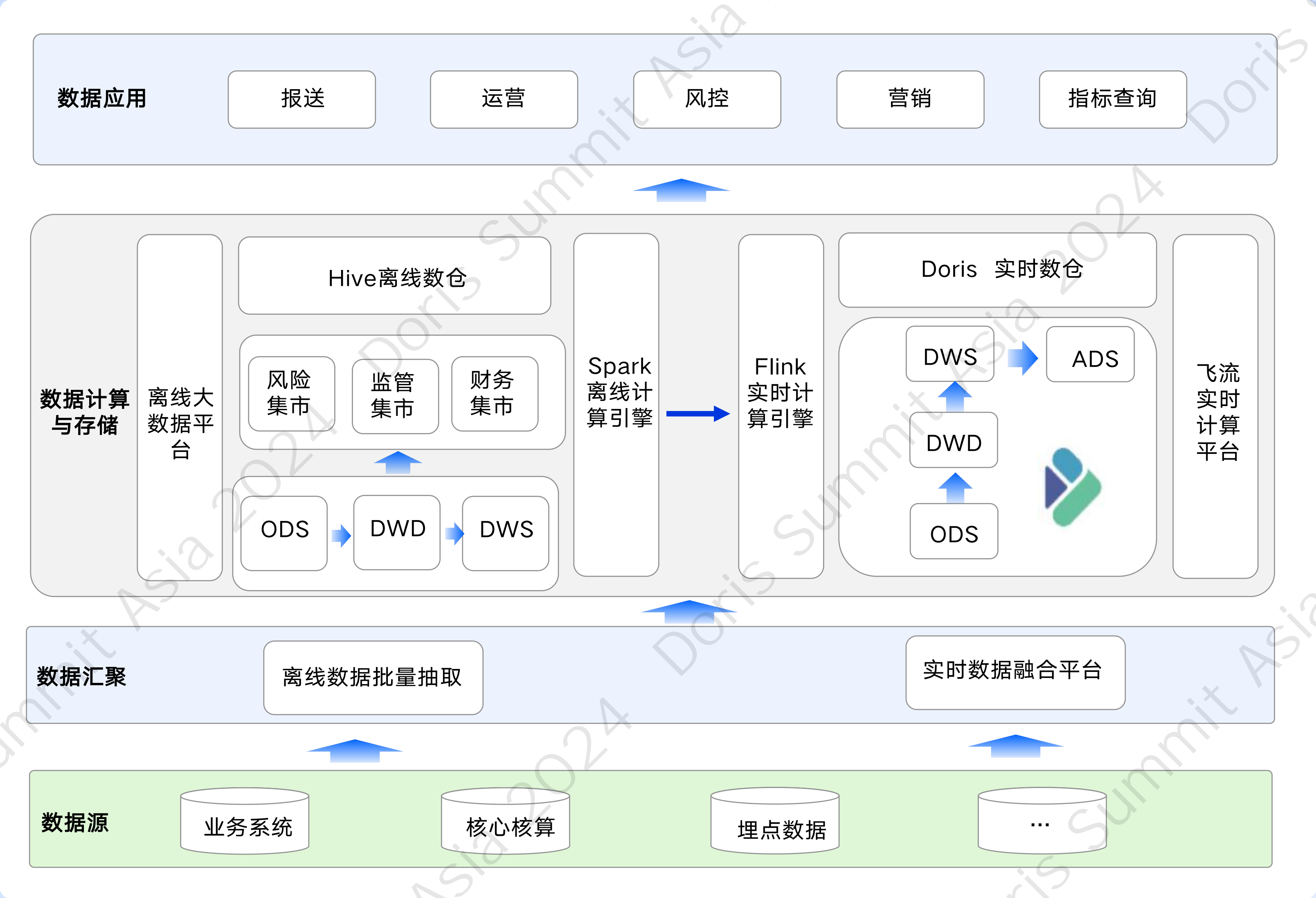
实时计算2.0架构

为解决早期架构面临的问题，开始基于数据融合平台+Kafka+Flink+Doris+数据服务平台搭建实时数仓，以应对早期架构面临的问题：

- 精简架构：统一技术架构，下线HBase和Phoenix，全部基于Doris提供存储和查询服务
- 搭建数仓：基于数仓理念并结合行内实际业务场景，规划主题，统一实时数据开发流程
- 完善服务：引入指标平台，完善数据服务能力，满足下游多样性查询需求
- 保障稳定：推动商业化升级，维护Doris的稳定，并及时有效应对突发情况



整体架构图

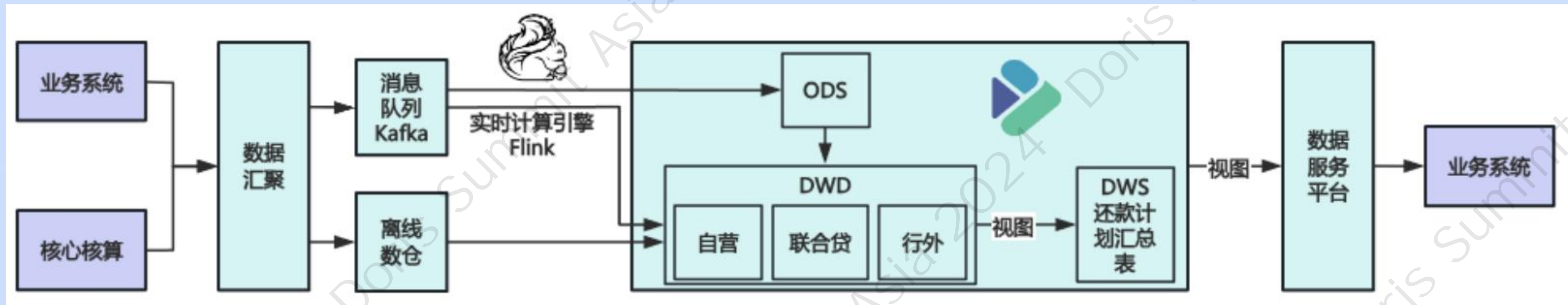


- 实时任务平台化管理：实时任务由实时计算平台统一提交和运维，实现一站式平台化管理
- Lamada 架构兼顾数据时效性和准确性
- 实时数据幂等写入，保证端到端数据一致性
- 实时数仓持久化存储，方便数据回查和复用

03

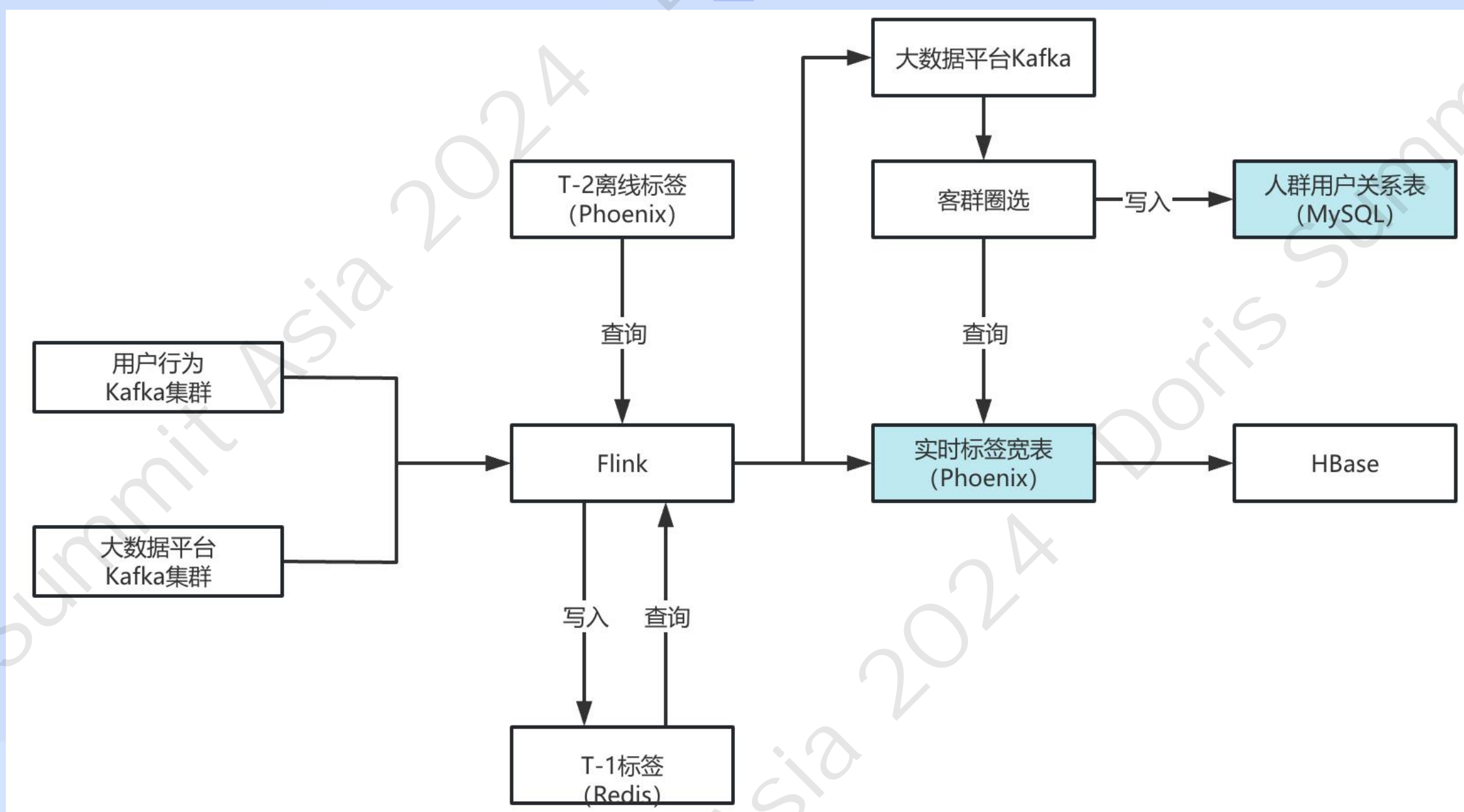
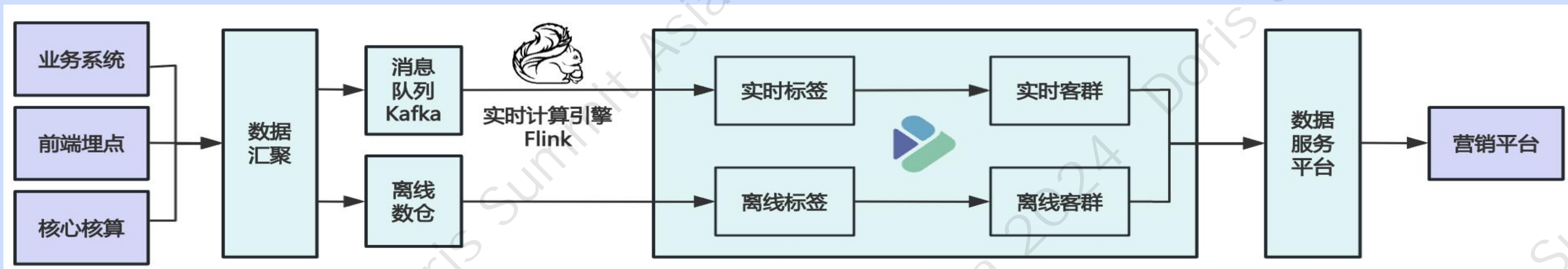
Flink+Doris 架构应用实践

风控-还款计划和还款表现



- ◆ **明细层：**充分利用 Flink 的状态管理机制，分摊 Doris 查询压力，数据清洗和加工均在 Flink 端完成，为了便于开发和维护，根据产品类别，划分了自营、联合贷、三方产品三张分表，均采用聚合模型。
- ◆ **汇总层：**基于 Doris 视图实现简单汇总，将 DWD 层的三张分表 Union 在一起，便于下游使用，不实际存储数据。
- ◆ **应用层：**同样通过视图实现，该层面向于实际业务需求场景，基于 DWS 层进一步加工汇总，根据应用系统查询需求的不同，提供不同的接口服务。
- ◆ **任务监控：**自定义监控、系统监控
- ◆ **整体响应时间：**平均 78ms
- ◆ **数据时效性：**<1min

客群圈选



- **实时链路：**基于实时数仓，采集、加工、计算、存储实时标签，最终圈选实时客群
- **离线链路：**基于离线数仓，采集、加工、计算、存储离线标签，最终圈选离线客群
- **营销平台：**通过接口获取符合规则条件的客群信息

Doris表结构设计

主键及索引设计

- DWD层为业务数据明细表，通过联合主键确定唯一一条数据
- 查询字段使用布隆过滤器创建二级索引，加速查询

```
`term_repay_status` varchar(10) REPLACE_IF_NOT_NULL NULL
`is_deleted` tinyint(4) REPLACE_IF_NOT_NULL NULL
`etl_time` datetime MAX NULL DEFAULT CURRENT_TIMESTAMP
) ENGINE=OLAP
AGGREGATE KEY(`bill_no`, `term_no`, `prod_code`, `cust_id`, `a_b_flag`, `fund_flag`)
COMMENT '个人贷款还款计划和还款表现表-核算自营mb'
DISTRIBUTED BY HASH(`cust_id`) BUCKETS 32
PROPERTIES (
  "replication_allocation" = "tag.location.default: 3",
  "bloom_filter_columns" = "cust_id",
  "in_memory" = "false",
  "storage_format" = "V2",
  "disable_auto_compaction" = "false"
);
```

分桶设计

- 0-10M 的数据量:1
- 10-50M 的数据量:2
- 50M-2G 的数据量:4
- 2-5G 的数据量:8
- 5-25G 的数据量:16
- 25-50G 的数据量:32
- 超过 50G 的数据量:64

表模型设计

- 为保证幂等写入，明细层统一采用聚合模型
- 为实现部分列更新，需更新字段声明为“replace if not null”

| | | | |
|--------------------|-------------|---------------------|------|
| `biz_prod_code` | varchar(50) | REPLACE_IF_NOT_NULL | NULL |
| `prod_name` | varchar(50) | REPLACE_IF_NOT_NULL | NULL |
| `credit_cont_no` | varchar(40) | REPLACE_IF_NOT_NULL | NULL |
| `cust_name` | text | REPLACE_IF_NOT_NULL | NULL |
| `cert_type` | varchar(20) | REPLACE_IF_NOT_NULL | NULL |
| `cert_no` | varchar(40) | REPLACE_IF_NOT_NULL | NULL |
| `total_loan_terms` | int(11) | REPLACE_IF_NOT_NULL | NULL |
| `loan_term_type` | varchar(2) | REPLACE_IF_NOT_NULL | NULL |
| `term_start_date` | date | REPLACE_IF_NOT_NULL | NULL |
| `term_mature_date` | date | REPLACE_IF_NOT_NULL | NULL |
| `grace_days` | int(11) | REPLACE_IF_NOT_NULL | NULL |
| `term_grace_date` | date | REPLACE_IF_NOT_NULL | NULL |

数据服务-统一数据服务平台

欢迎进入 招商银行大数据联机分析处理平台

欢迎您, admin

退出登录

首页

项目

接口信息

接口名称:

项目名称:

状态:

查询

重置

新增

导出

导入

| <input type="checkbox"/> | 编号 | 接口名称 | 所属项目 | 数据源 | 状态 | 创建人 | SQL | 描述 | 创建时间 | 修改时间 | 操作 |
|--------------------------|----|------------------|-----------|-----------|----|-----|-------------|--------------|---------------------|---------------------|----------|
| <input type="checkbox"/> | 1 | cc_u_data | cc_system | doris_uat | 启用 | | select ... | --接口名称: 贵... | 2024-11-14 15:33:25 | | 详情 测试 更多 |
| <input type="checkbox"/> | 2 | credit_api | credit | doris_uat | 启用 | | SELECT ... | --接口名称: x... | 2024-11-14 08:55:12 | | 详情 测试 更多 |
| <input type="checkbox"/> | 3 | xs_s_or_realtime | xs_system | doris_uat | 启用 | | select ... | --接口名称: 自... | 2024-11-07 10:46:34 | 2024-11-25 10:38:03 | 详情 测试 更多 |
| <input type="checkbox"/> | 4 | it_info | it_system | doris_uat | 启用 | | select ... | --接口名称: 供... | 2024-10-31 17:02:12 | | 详情 测试 更多 |
| <input type="checkbox"/> | 5 | xs_d_statistics | xs_system | doris_uat | 启用 | | select ... | --接口名称: 催... | 2024-09-27 08:50:03 | 2024-11-06 14:30:18 | 详情 测试 更多 |
| <input type="checkbox"/> | 6 | n_bill_info | ns_system | doris_uat | 启用 | c | select ... | --接口名称: 我... | 2024-09-19 11:19:36 | 2024-10-22 16:26:43 | 详情 测试 更多 |
| <input type="checkbox"/> | 7 | '29 | f | doris_uat | 启用 | ry | select c... | -- | 2024-07-29 10:03:26 | 2024-07-29 10:05:04 | 详情 测试 更多 |
| <input type="checkbox"/> | 8 | 3 | | doris_uat | 启用 | y | select u... | -- | 2024-07-29 09:50:27 | 2024-07-29 10:00:53 | 详情 测试 更多 |
| <input type="checkbox"/> | 9 | ac_dv_f | tail_bill | cc_system | 启用 | z | select ... | -- | 2024-07-26 15:06:40 | | 详情 测试 更多 |
| <input type="checkbox"/> | 10 | pu_rain_dep | nt_avg | cc_system | 启用 | y | select ... | -- | 2024-07-23 16:49:29 | | 详情 测试 更多 |

1-10 共98条 < 1 2 3 4 5 ... 10 > 10条/页 跳至 页

首页

项目

接口信息

调用记录

日期: -

项目:

接口:

状态:

查询

重置

导出

已选择 0项

清空



04

总结与展望

平台收益

63_↑

接口数量

86_万

每日调用次数

39_{ms}

接口响应时间

开发收益

上游数据汇聚简化

通过状态管理和幂等写入，上游数据汇聚无需特别关注数据乱序或者重复问题，开发阶段统一处理

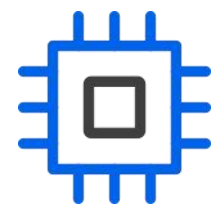
数据管理更加清晰

通过业务主题建设和分层管理，每层数据各司其职，解决了数据无法复用和重复开发的问题，真正实现一次开发，多次使用的目标

下游开发效率提高

数据服务团队可直接基于模型数据开发相应指标，不需要关注实现过程，对离线和实时切换无感

业务收益



营销

满足营销场景下客户标签和客群实时更新需求



风控

提高线上审批效率，加快审批流程，满足高并发点查



运营

报表更加及时，便于领导层及业务人员决策



后台管理

解决业务系统响应慢、查询体验的不好问题，效率更高

后续规划

集群能力升级

- ◆ 资源隔离
- ◆ 搭建双Doris集群（CCR方案）

完善实时数仓

- ◆ 完善主题，满足更多业务需求
- ◆ 完善指标平台，维护血缘关系

完善实时开发生态体系

- ◆ 灯塔平台
- ◆ 数据治理平台

Thanks for Watching!