

银联商务基于ApacheDoris 构建新一代实时数仓实践

杨劲雄

银联商务股份有限公司

目录

1. 公司介绍
2. 架构演进
3. 基于 Doris 的应用实践
4. 展望未来

1 公司介绍

银联商务简介

主营业务

支付银商

线下支付服务
线上支付服务
国际业务

增值银商

结算增值服务
数字营销
客户及终端专业化服务
数字终端
普惠金融服务

科技银商

商业科技服务
大数据服务
IDC及云服务
信息化解决方案

市场规模大

累计服务商户 > 2500万家
累计铺设终端 > 3900万台

市场份额高

收单交易笔数连续十年蝉联亚太收单机构榜首
2022年综合支付交易清算笔数超140亿笔，
交易金额超19万亿元



公司架构

36家一级分公司
10家一级全资子公司
6家一级控股子公司

数据应用场景



业务诉求



更快的数据分析速度



更灵活的数据分析模式



更稳定的数据服务



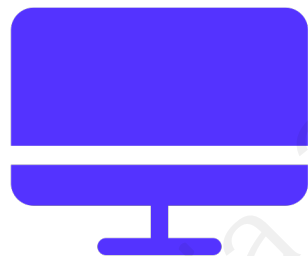
更实时的数据时效

2 架构演进

离线数仓架构



架构升级目标



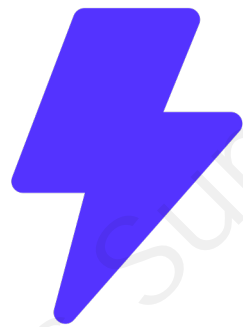
统一、简洁

一套系统
统一加工
统一服务



准确、实时

数据准确、不丢不重复
支持数据更新
支持实时数据接入与加工



高效、稳定

数据加工效率高
高性能查询
集群稳定



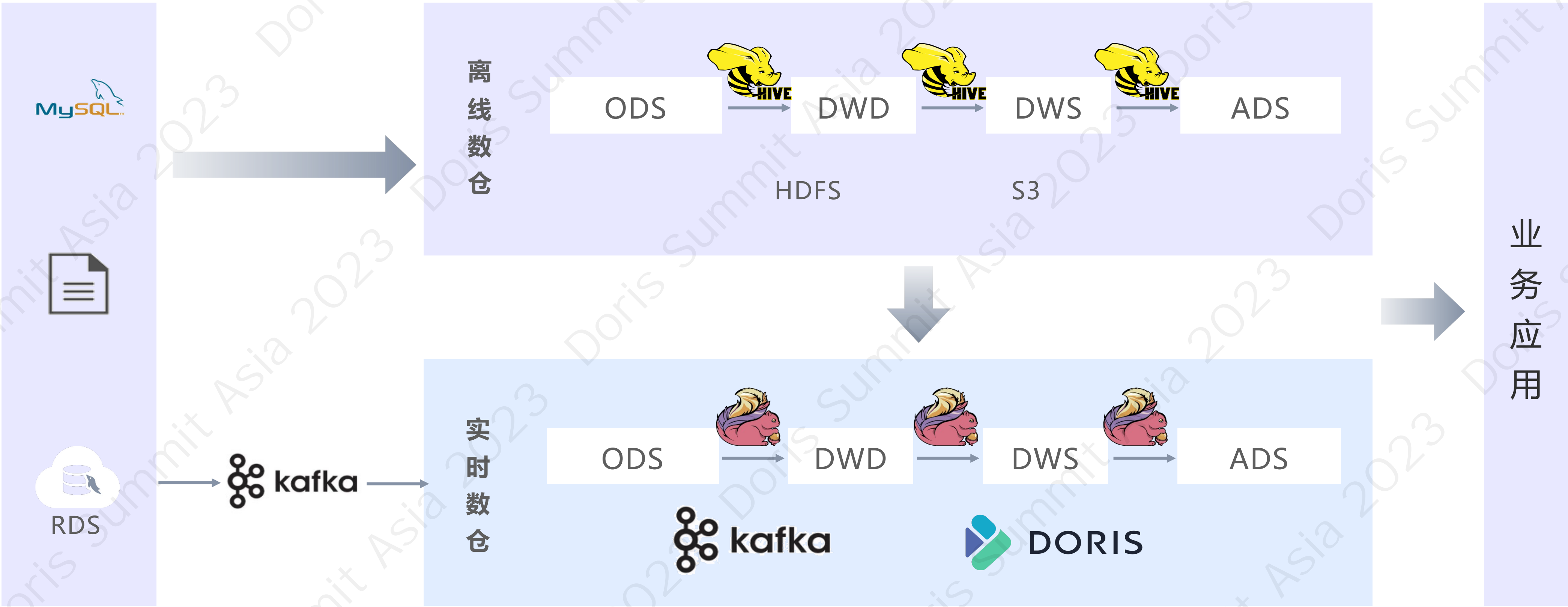
安全、可靠

数据访问安全
数据存储安全
集群灾备

为什么选择 Apache Doris

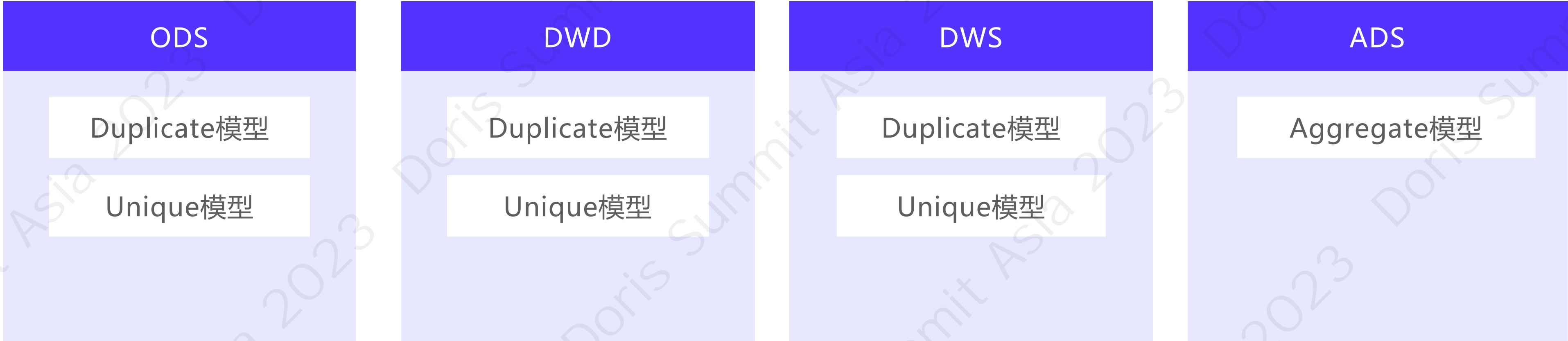
极简易用	架构统一	高性能	多租户管理	安全
<ul style="list-style-type: none">兼容 MySQL 协议，学习成本低高效运维	<ul style="list-style-type: none">湖仓一体数据服务统一出口	<ul style="list-style-type: none">高性能计算高性能查询高并发查询	<ul style="list-style-type: none">资源隔离权限控制	<ul style="list-style-type: none">跨集群数据复制容灾备份

新的实时数仓架构



3 基于 Doris 的应用实践

数据仓库统一规划--数据模型



Duplicate模型

明细数据场景
任何维度的即席查询

Unique模型

数据更新场景
大宽表多流Upsert
标签数据生成

Aggregate模型

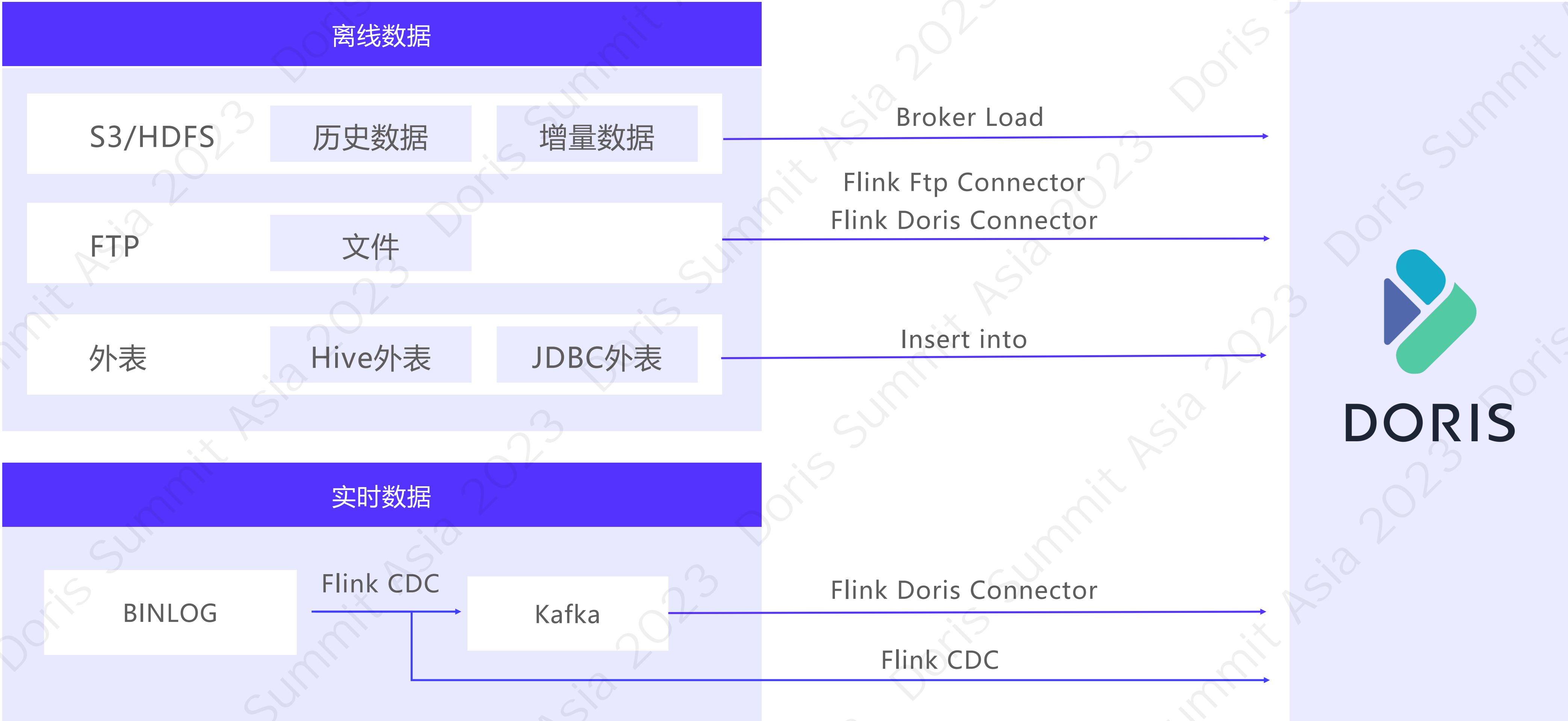
聚合场景
预计算

数据仓库统一规划--分桶数量

建议分区分桶，Tablet 总数量 = 分区数 * 桶数
单个tablet大小建议在1G-10G之间。

数据大小	Bucket数量
0M-500M	4
500M-5G	8
5G-25G	16
25G-50G	32
>50G	数据大小/1.5

数据同步与加工——丰富的数据同步方式



数据同步与加工--数据同步策略

全量同步

- 清空临时表
- 导入临时表
- 使用临时表原子替换正式表

```
ALTER TABLE t1 REPLACE WITH TABLE t2;
```

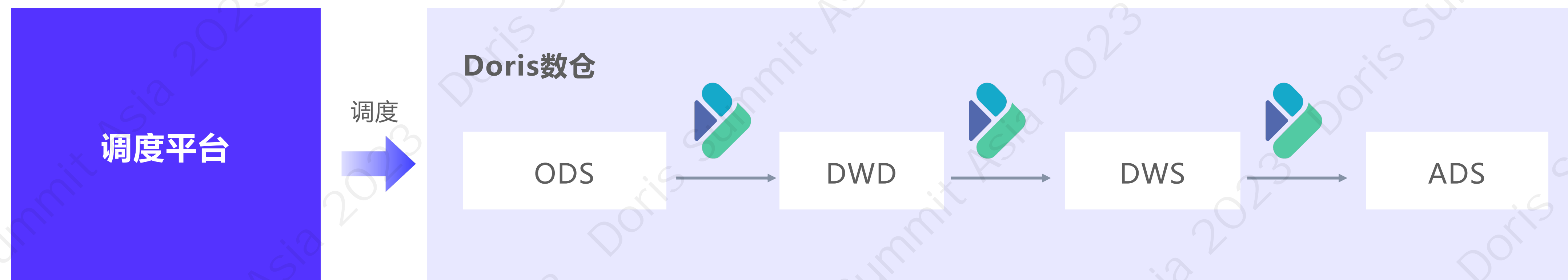
增量同步

- 删除增量分区
- 新建增量分区
- 导入数据到增量分区

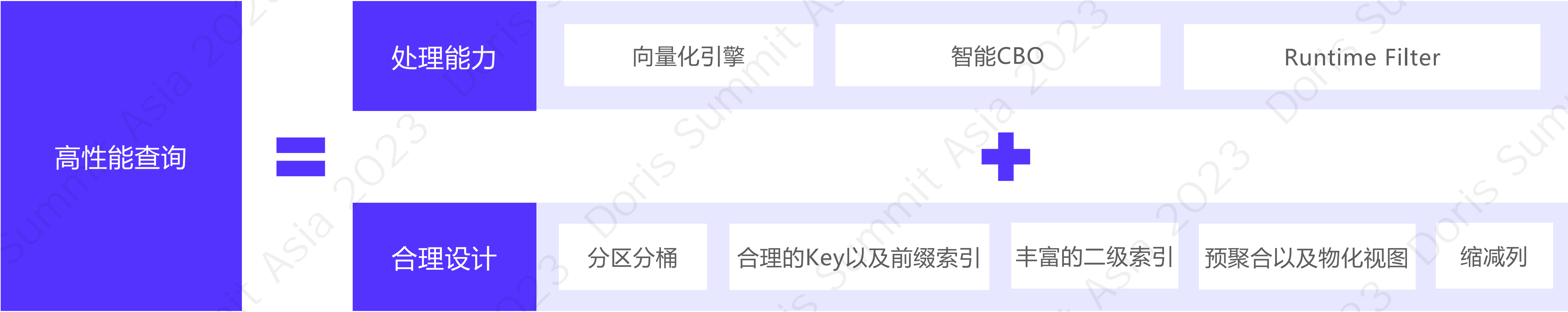
```
alter table ${DB_NAME}.${TBL_NAME} drop partition IF EXISTS p${P_DOWN_DATE};
ALTER TABLE ${DB_NAME}.${TBL_NAME} ADD PARTITION IF NOT EXISTS p${P_DOWN_DATE} VALUES [('${P_DOWN_DATE}'), ('${P_UP_DATE}')];
```

```
LOAD LABEL ${TBL_NAME}_${load_timestamp} ...
```

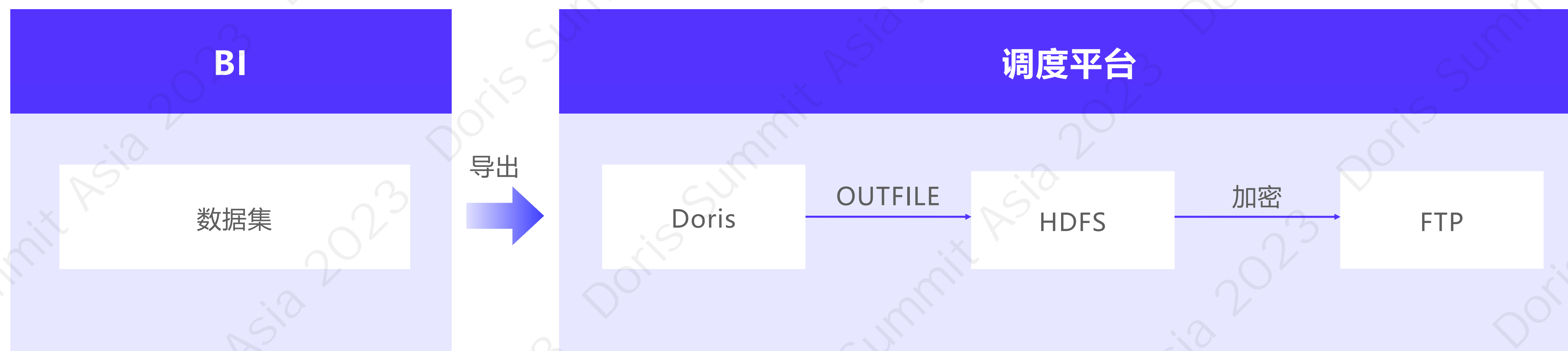

数据同步与加工--数据加工



应用价值



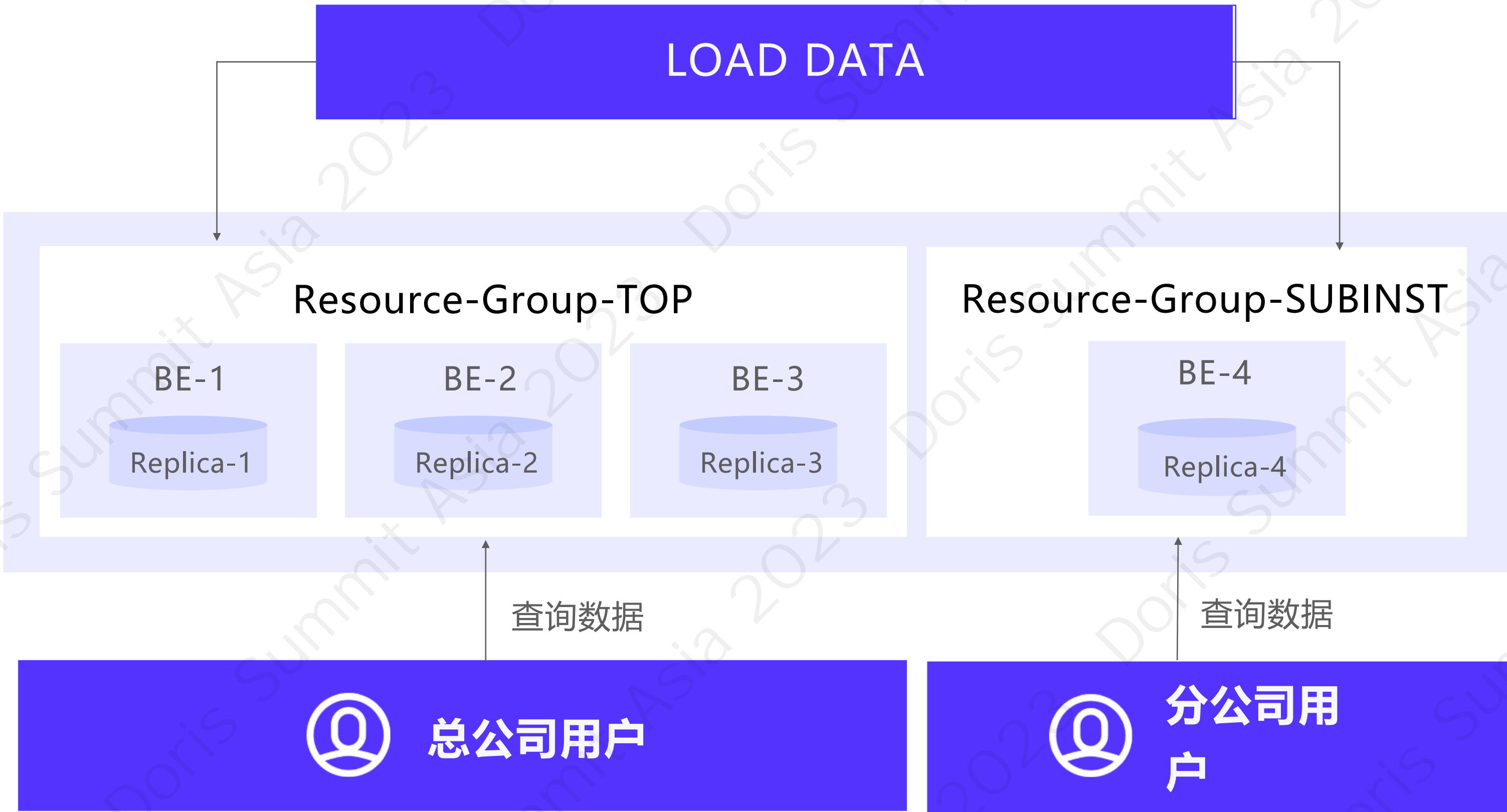
数据查询与分析——自助取数场景



多租户管理——单查询资源限制

分类	cpu_resource_limit	max_user_connections
etl类	20	300
点查类	30	300
看板类	10	100
分析类	5	100

多租户管理——资源组硬隔离



设置BE节点标签

- 分配总公司资源组和分公司资源组

设置数据分布

- 建表时设置好replication_allocation
- 总分公司比例为3：1

设置用户资源组

- 总公司使用总公司资源组
- 分公司使用分公司资源组

多租户管理-- workload group 软隔离

```
create workload group if not exists etl_wg
properties (
  "cpu_share"="30",
  "memory_limit"="20%",
  "enable_memory_overcommit"="true"
);
```

```
set experimental_enable_pipeline_engine = true;
set property 'default_workload_group' = 'etl_wg';
```

设置workload分组

开启pipeline执行引擎
设置用户默认的workload group

多租户管理--用户权限管理

每个分支机构，不同场景，不同用户

```
create user sh_query identified by 'password' ;
```

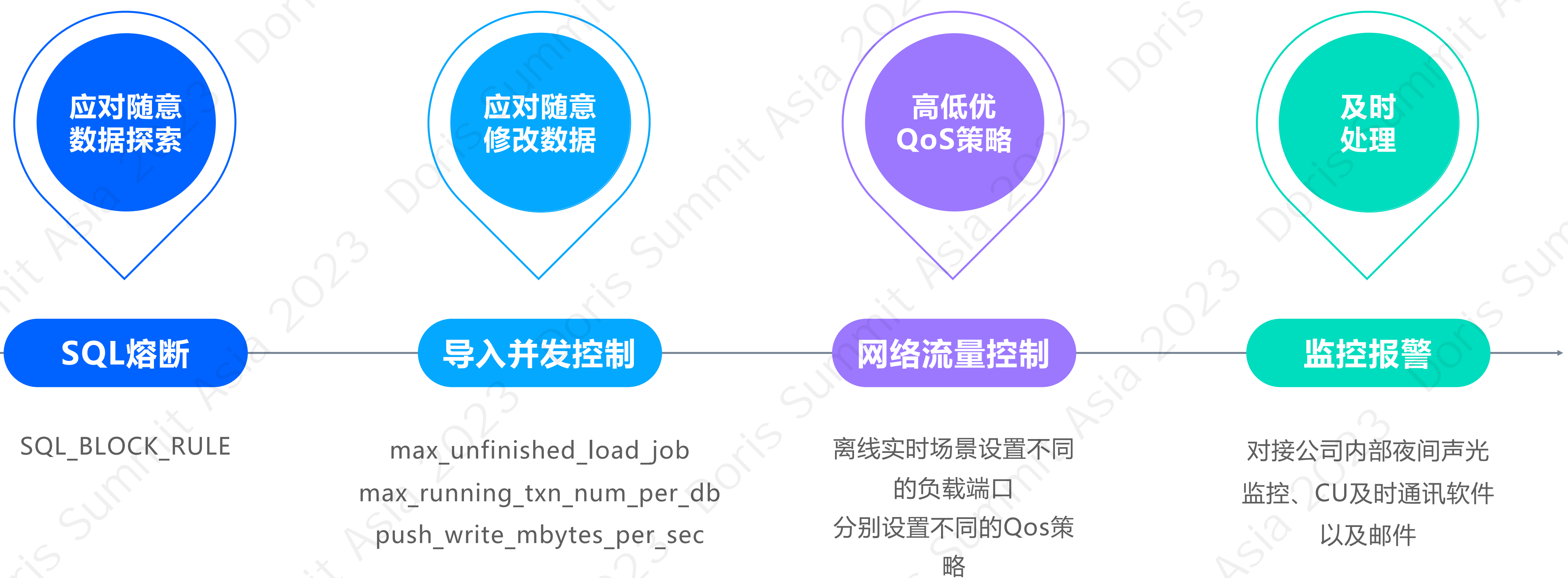
库、表、行级权限管理

```
CREATE ROW POLICY dwd_trans_test_di_row_policy_sh ON  
dwd_db.dwd_trans_test_di  
AS RESTRICTIVE TO sh_query USING (subinst_cd = 'xxx');
```

列级权限管理

```
create view fjfgs_db.dwd_trans_test_di (  
...  
)  
AS SELECT * FROM dwd_db.dwd_trans_test_di  
where subinst_cd = 'xxx' ;
```

集群稳定性保障



集群灾备--CCR

跨集群数据复制

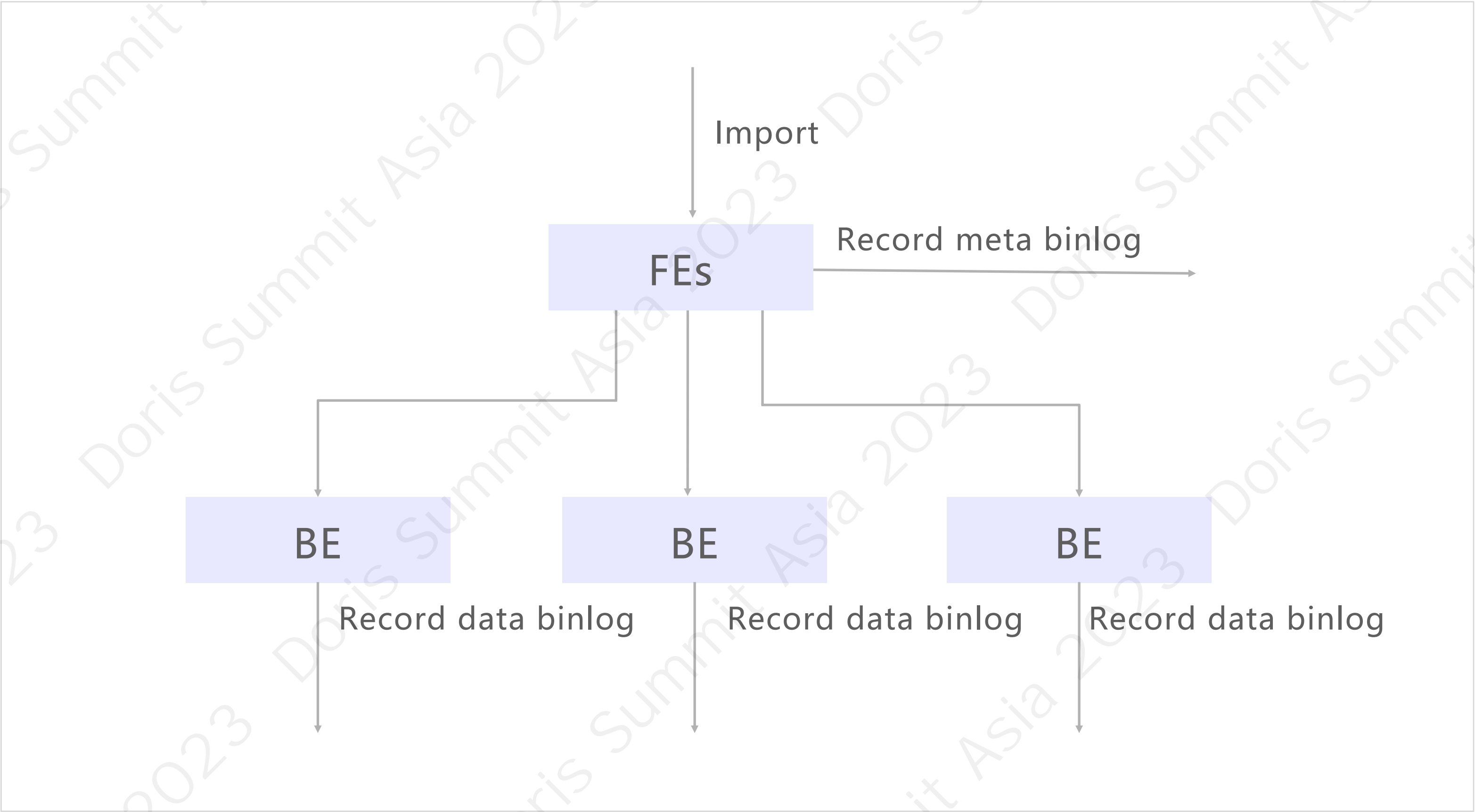
新建集群数据同步

容灾备份

重点业务以及核心数据灾备保障

集群隔离升级

新功能升级、双跑验证



使用Apache Doris 带来的收益

导入

2-5倍

ETL

3-12倍

查询分析

10-15倍

4 展望未来

展望未来

存算分离架构

- 存算分离，弹性扩缩容
- 离线数仓存量 ETL 任务迁移
- 单一数仓，各种计算负载进一步的隔离

统一查询引擎

- 联邦查询
- 统一路由

运维

- 对接公司内部流程，实现自动化运维
- 快速问题排查
- 数据血缘分析



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>