

# 正泰集团数据中台基于 Apache Doris的应用实践

许继良

技术中心大数据架构师

# 目录

01 背景介绍

02 数仓架构

03 数据中台基于Doris应用

04 未来规划

01

# 背景介绍

# 集团业态分布

## 智能电器

低压电器、仪器仪表、智能家居等

## 新能源

光伏电站、户用光伏电站、电源等



## 智能电气

输配电、电线电缆、集成配电等

## 其他

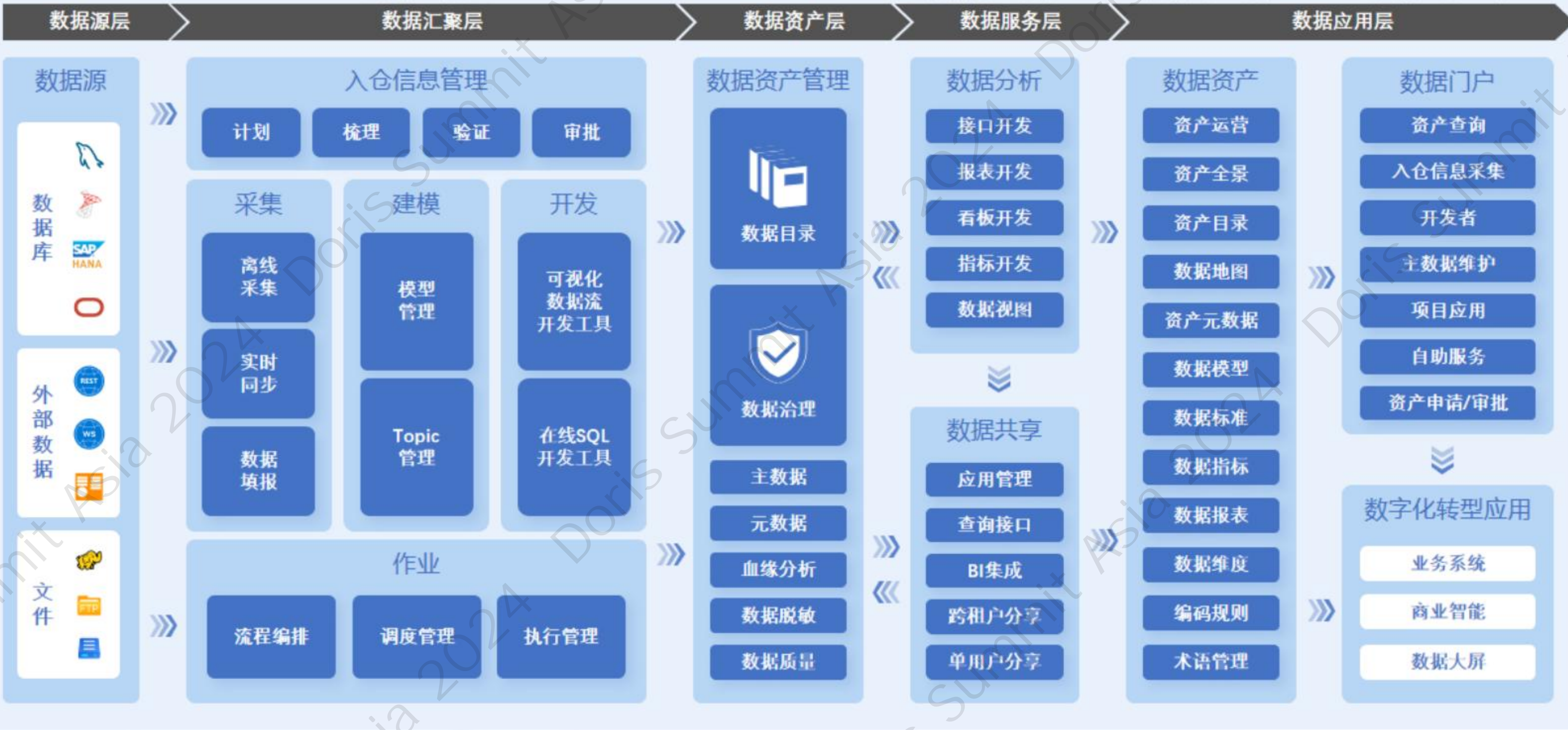
智慧农业、电商、汽车科技等

# 发展历程





# 数据中台-能力全景





# 数据中台-零代码开发实现





# IoT 平台-数据驱动的低代码解决方案



平台是一款数据驱动的低代码开发平台，通过物联网平台、组态平台、场景应用等核心功能，实现数据从接入、储存、清洗、分析、预测、呈现的全过程的处理能力，为智慧工厂、智慧园区、智能电站、智慧楼宇、智能消防、智慧城市等行业，提供 IoT 应用领域完整数智化解决方案。



# IoT 平台-零/低代码组态平台

结合用户的深度业务逻辑与三维可视化技术，平台提供模型创建、场景设计、组态配置的全流程方案，通过数据挖掘和 IoT 技术为数据赋能，驱动“智能+”在各个细分领域的渗透。



光储充一体化



建筑环境检测



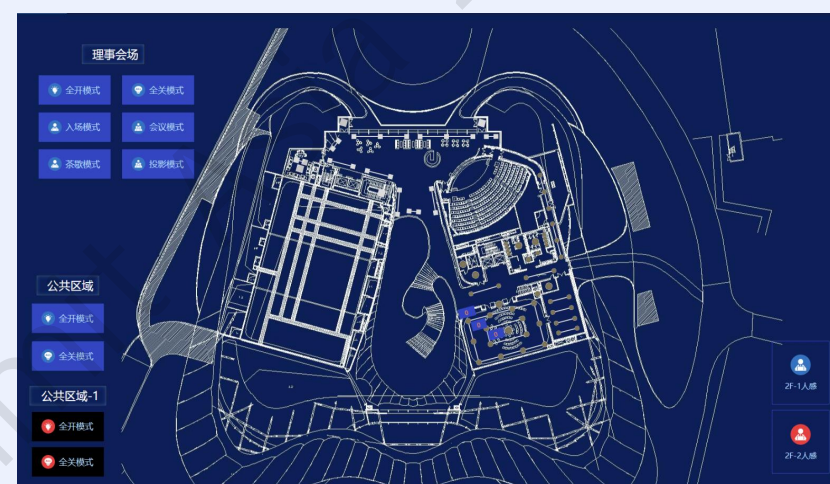
园区能耗检测



水务设备管控



智慧楼宇控制



建筑智能照明



智能安全用电



工厂智造一体化



工厂车间看板



园区配电监测



02

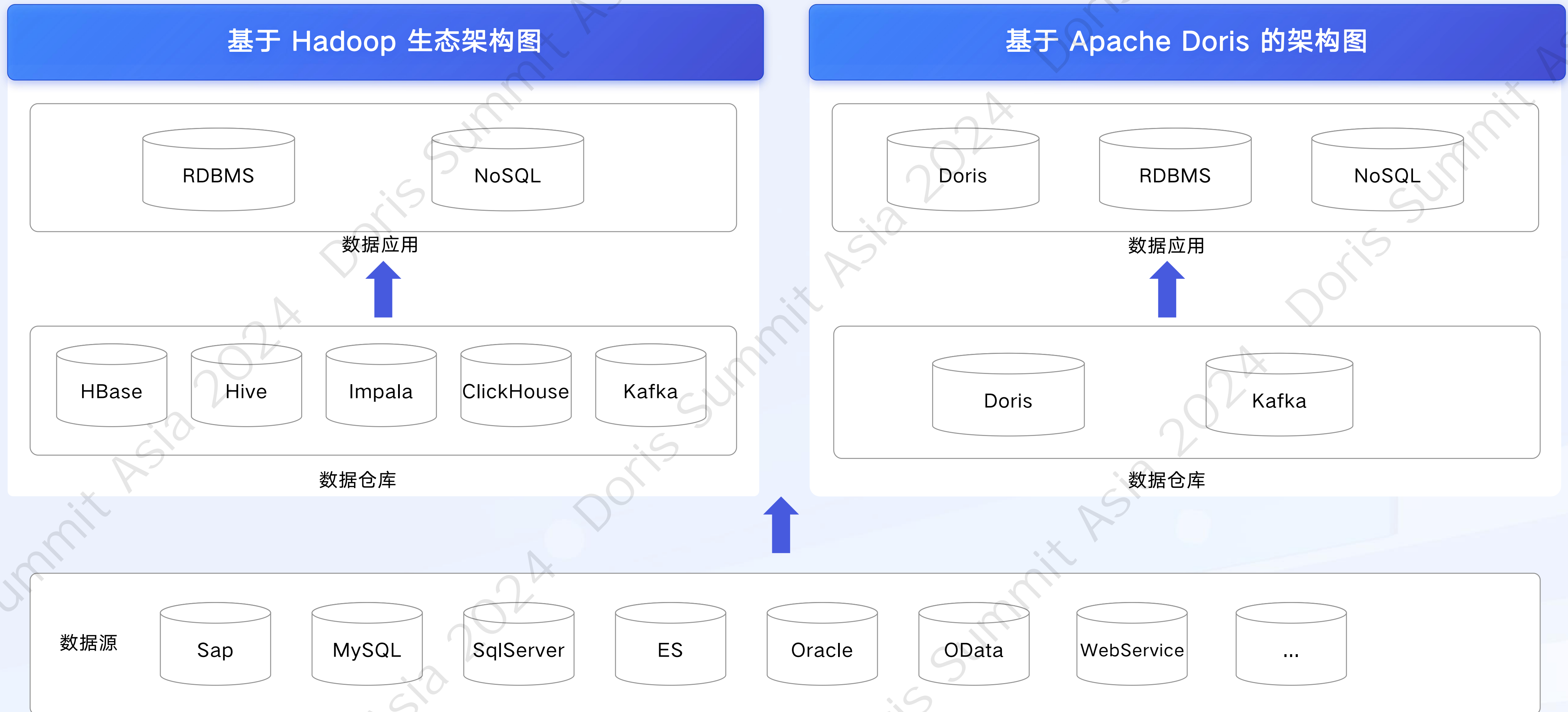
# 架构演进



# 数仓架构升级

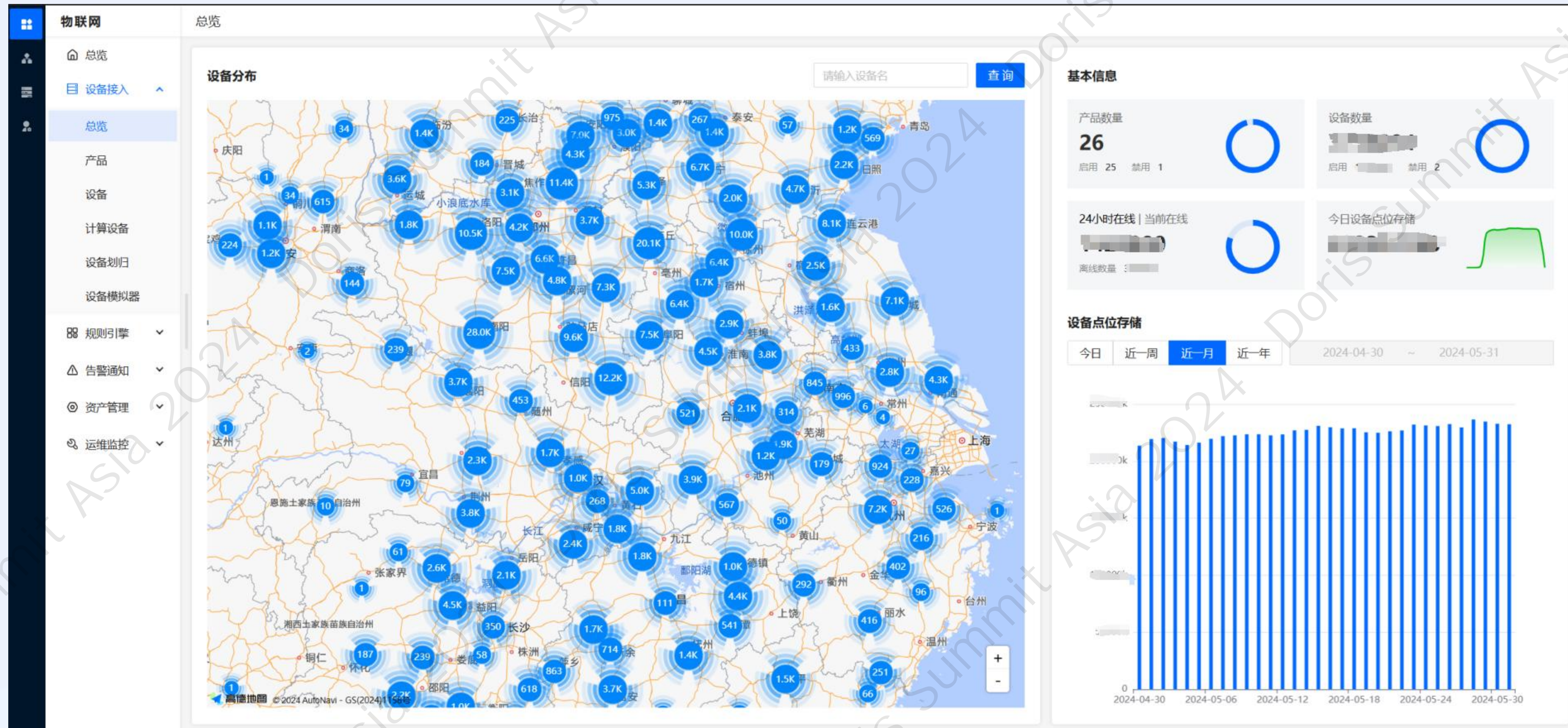
基于 Hadoop 生态架构图

基于 Apache Doris 的架构图



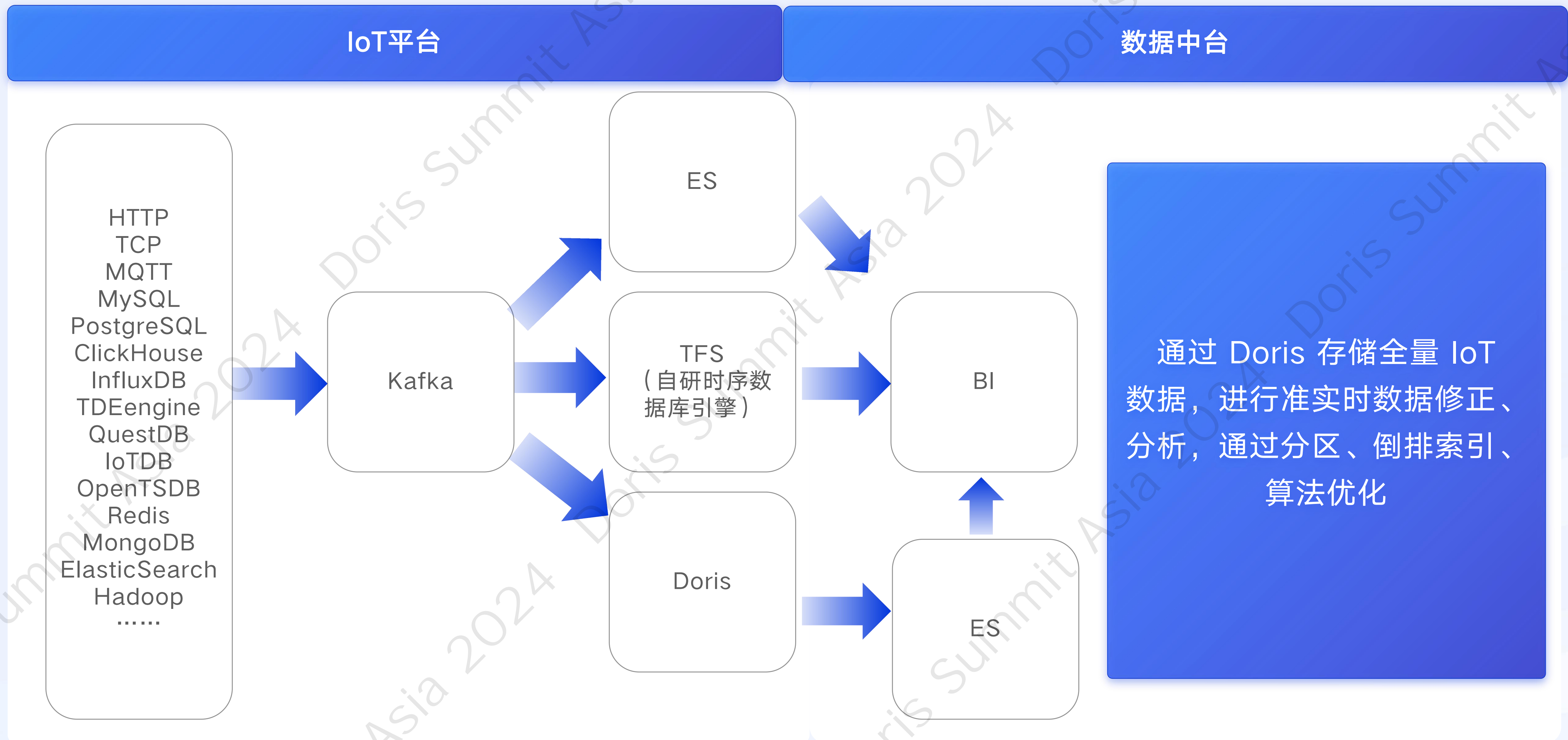


# IoT 平台场景





# 平台集成



03

# 数据中台基于 Doris 应用



# 权限控制

## 多租户多项目需求场景

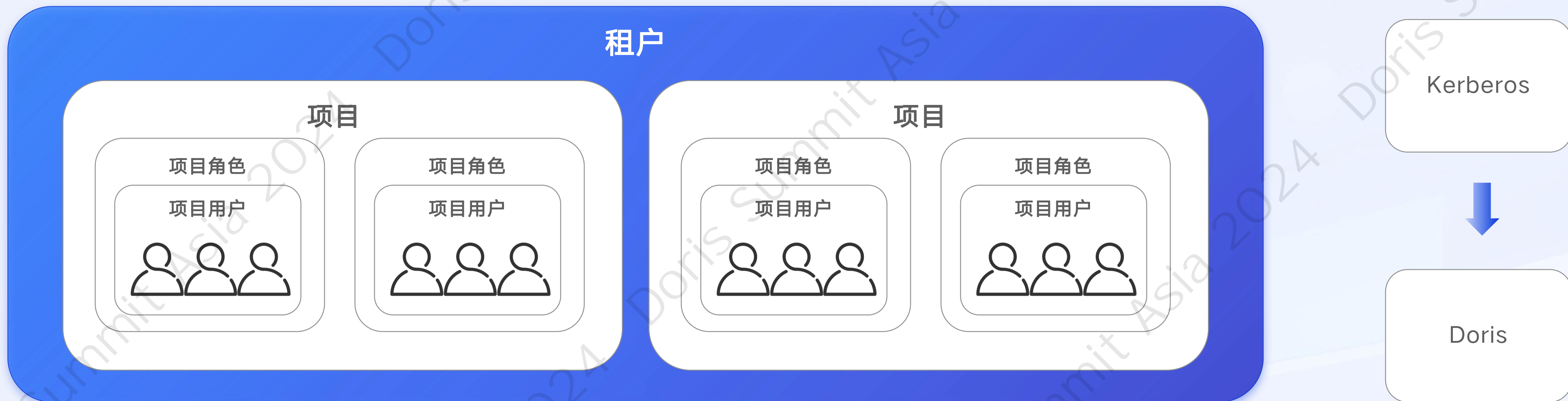
租户级别的能力由租户角色进行控制

提供以项目为单位管理资源，实现清晰合理的权限管理划分粒度

## 数据表权限控制

提供项目角色和表级别的权限控制，实现数据安全隔离

提供读、写、修改、删除四种权限粒度、实现按需控制



# 数据模型

\* 源数据源: 请选择数据源

模型类型: 标准模型表

\* 分层目录: 请选择分层目录

表名拼接规则: 分层\_业务系统\_板块\_自定义表名\_具体周期同步周期\_表类型

\* 表名配置: 选择分层 - 选择业... - 选择板块 - 自定义表名 - 具体周期 - 选择同步周期 - 选择表类型

表名称: 表注释: 请输入表注释

\* 模式: 唯一

\* 是否分区: 否

\* 压缩模式: ZSTD \* 副本数: 2 \* 分桶数: 10

扩展信息: 参数名 值 添加 删除

表属性配置

新增模型

保存 保存并关闭

表配置 字段配置

+ 添加自定义字段 从字段库添加 从Select语句添加 从DDL语句添加 自动生成字段名 字段格式转换 自动生成注释 批量设置字段注释 DDL预览

	主键	字段名	注释	类型	长度	精度	默认值	允许Null	数据密级
1	<input type="checkbox"/>			文本				<input checked="" type="checkbox"/>	
2	<input type="checkbox"/>			文本				<input checked="" type="checkbox"/>	

字段配置

逆向数据库导入

\* 源数据库: 请选择数据库 \* 目标数据库: 请选择数据库

\* 分层目录: 请选择分层目录 更新已有表: 不忽略 忽略

\* 源数据库表: 请输入表名

<input type="checkbox"/>	表名	注释
暂无数据		

逆向导入 Doris

新增模型

保存 保存并关闭

表配置 字段配置

\* 数据表: 请选择

过滤条件 符合 以下 全部 条件

收藏模板 选择模板

字段配置

<input type="checkbox"/>	主键	字段名	注释	类型	长度	精度	允许Null
暂无数据							

视图模型



# 流批一体



配置页面

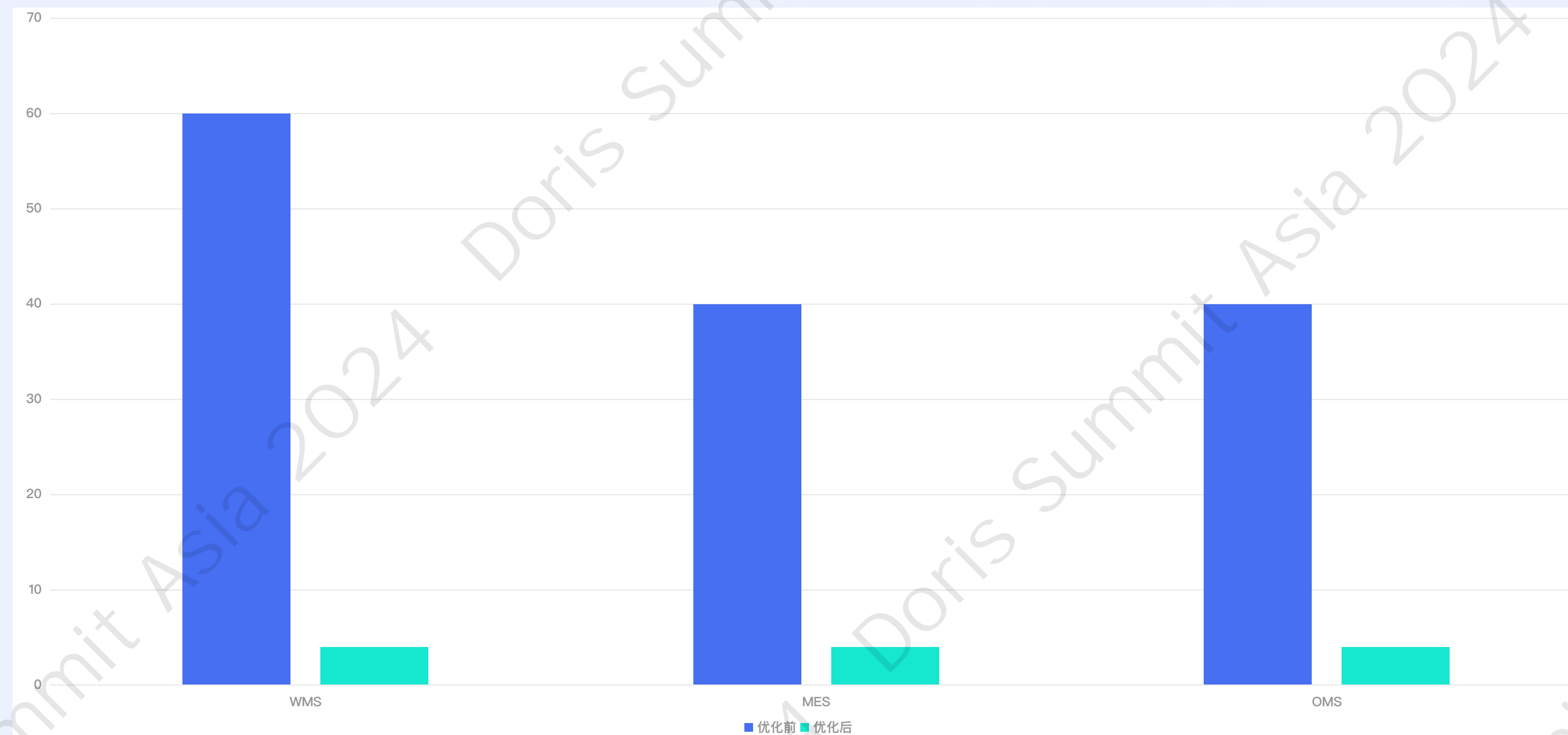
运行状态	健康状态			作业类型	操作
	Flink	Yarn	RoutineLoad		
未运行				实时同步作业	开发 最新日志 更多 v
未运行				实时同步作业	开发 最新日志 更多 v
运行中	运行中	运行中	正常	实时同步作业	开发 最新日志 更多 v
运行中	运行中	运行中	正常	实时同步作业	开发 最新日志 更多 v

作业监控

表状态详情		
任务信息 (jobProperties)		
max_batch_rows	每个子任务最多读取的行数	200000
timezone	指定导入作业所使用的时区	Asia/Shanghai
send_batch_parallelism	用于设置发送批量数据的并行度	1
load_to_single_tablet	支持一个任务只导入数据到对应分区的一个...	false
current_concurrent_number	当前并发数量	1
delete	数据删除操作	*
partial_columns	指定是否开启部分列更新功能	false
merge_type	数据合并类型	APPEND
exec_mem_limit	限制导入内存	2147483648
strict_mode	是否开启严格模式	false
jsonpaths	JSON路径	
max_batch_interval	每个子任务的最大运行时间	10
max_batch_size	每个子任务最多读取的字节数	104857600
fuzzy_parse	模糊解析设置	false
partitions	分区	*
columnToColumnExpr	指定导入作业所使用的时区	process_id,process_name,process_key,process_create_date,process_update_date,process_enable,...
whereExpr	条件表达式	*
desired_concurrent_number	单个导入子任务 (load task) 期望的并发度	5

表状态监控

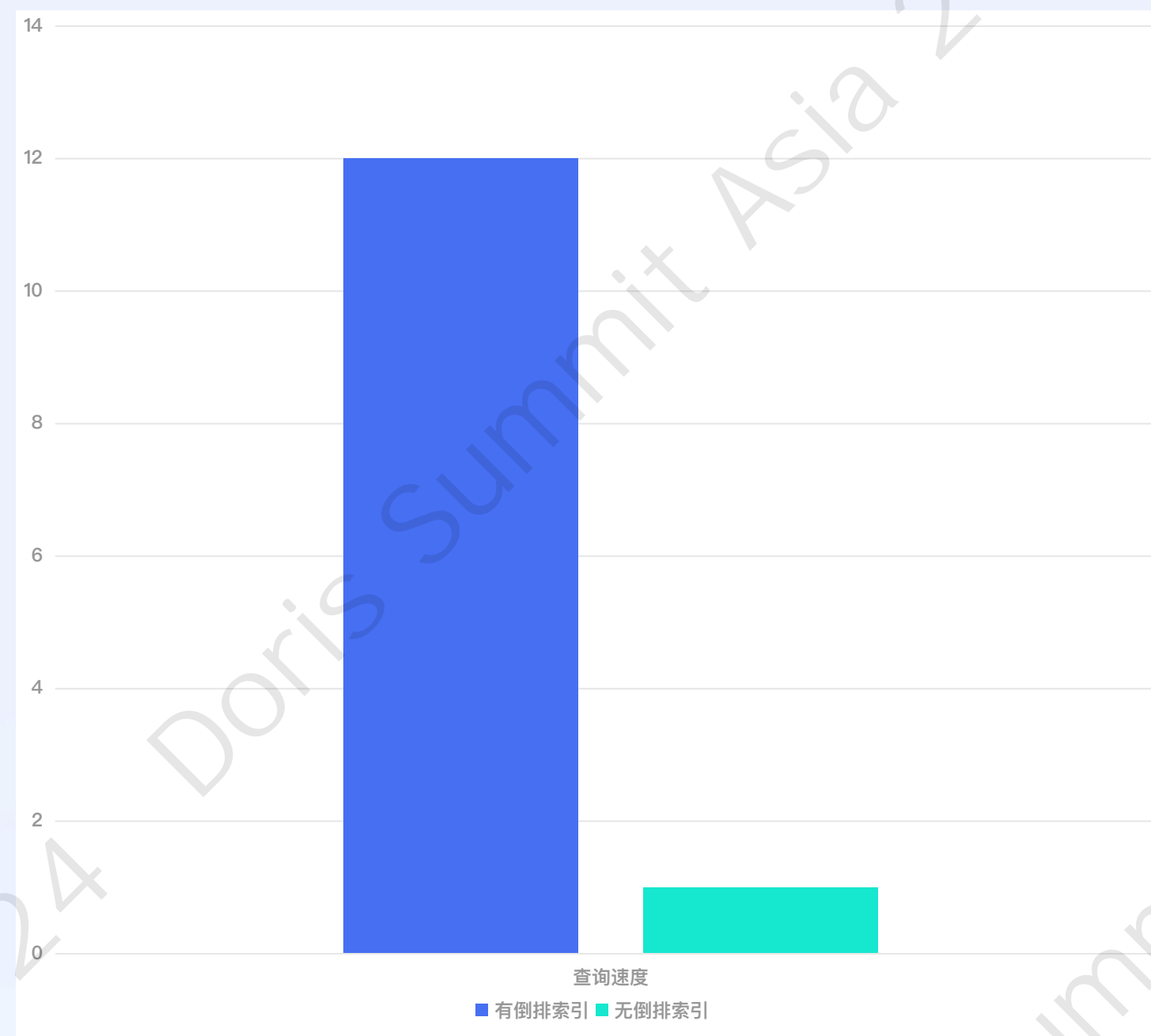
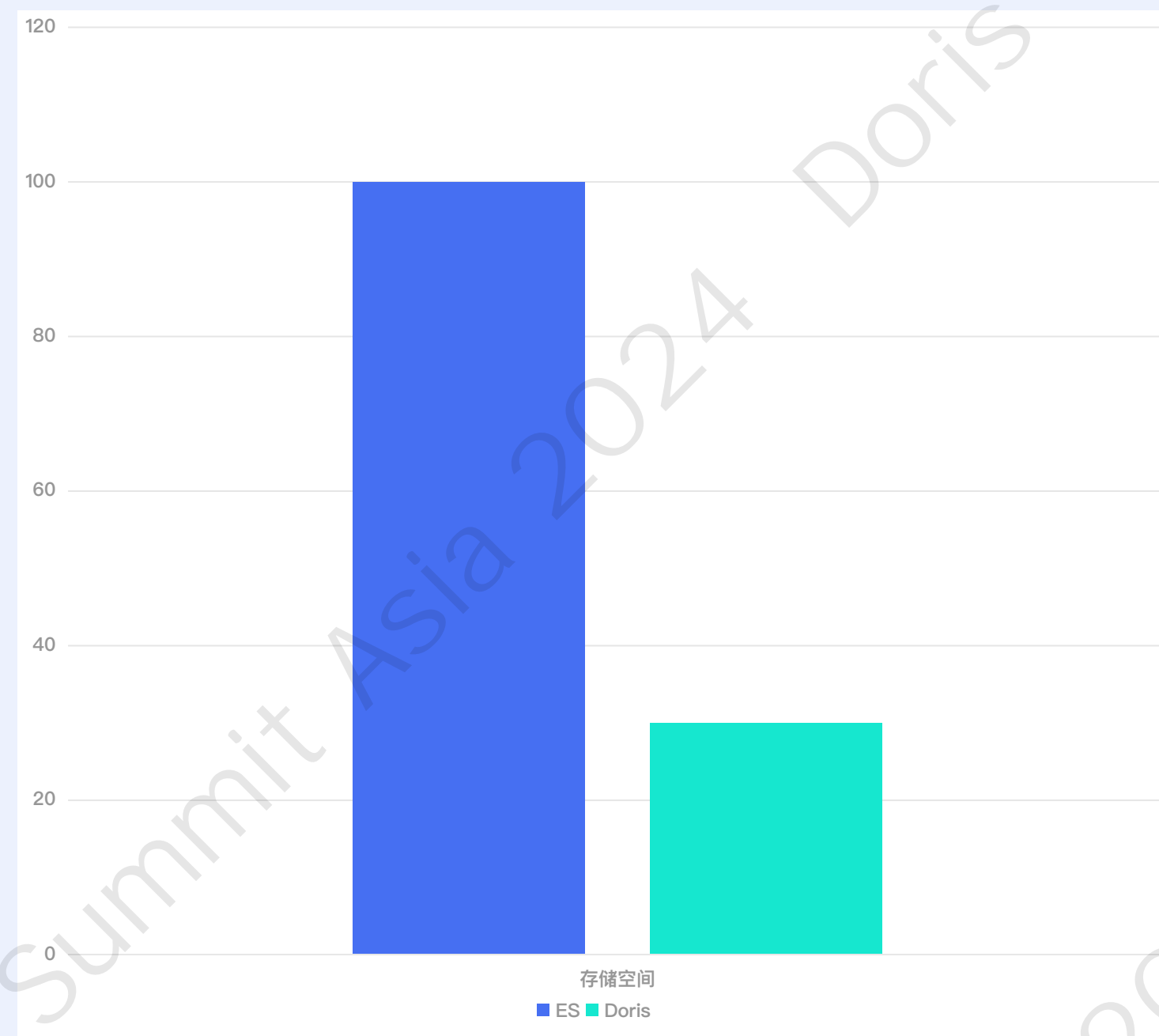
# 内存资源消耗对比



利用 Doris Routine Load  
大大减少资源内存资源消  
耗，表越多节省资源越大



# 日志迁移



将日志存入从 ES 迁移至 Doris，减少中间件，精简架构，节省资源70%。通过增加倒排索引，提升日志检索速度10倍以上。

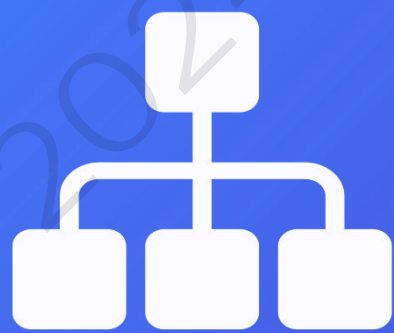
04

# 未来规划



# 未来规划

架构升级



数据湖扩展



大模型集成



# Thanks for Watching!