

Apache Doris

在蓝鲸体系中的应用与探索

王文植

腾讯 高级研发工程师

目录

1. 业务场景
2. 应用探索
3. 未来计划

1 业务介绍

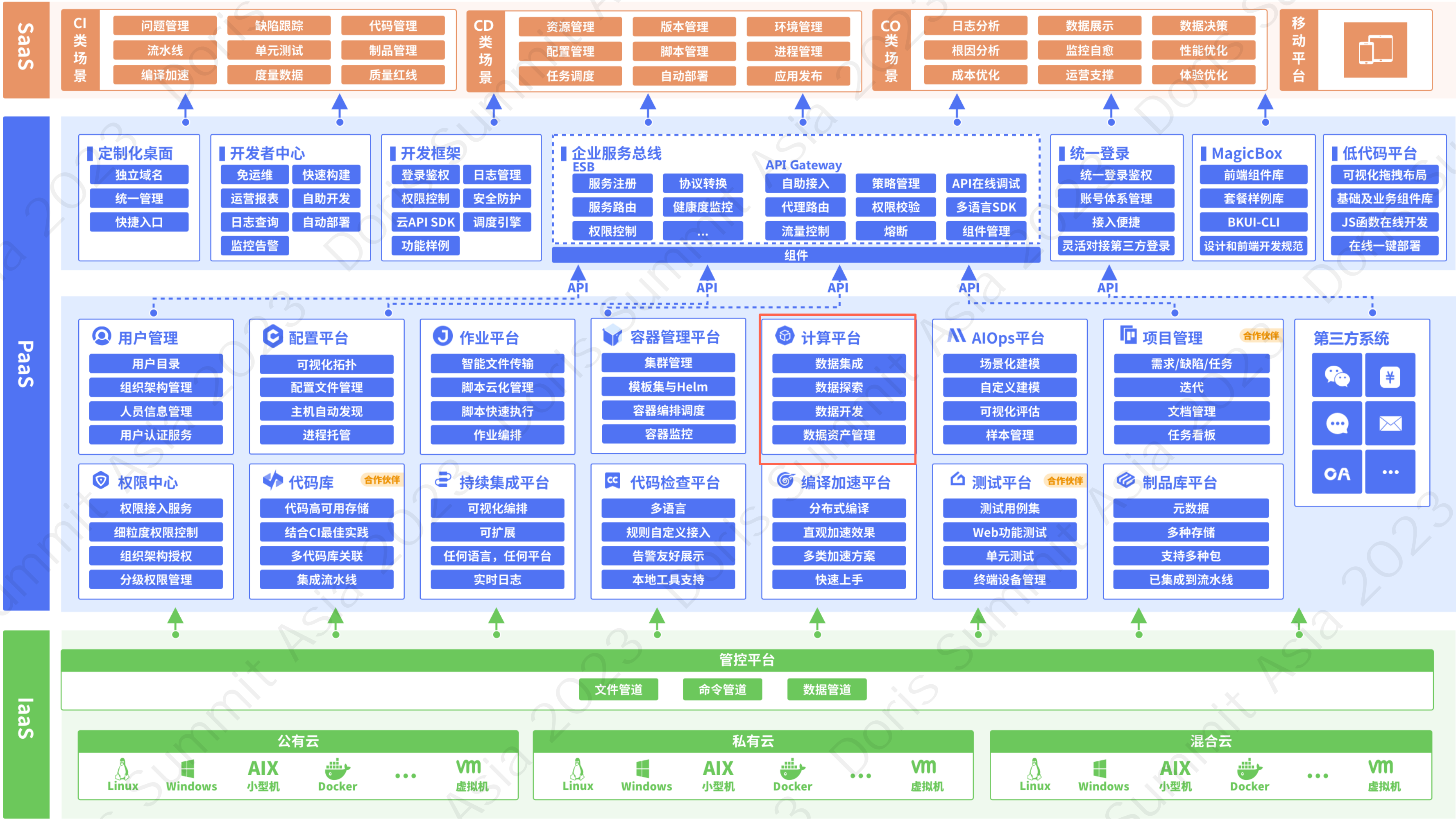
蓝鲸



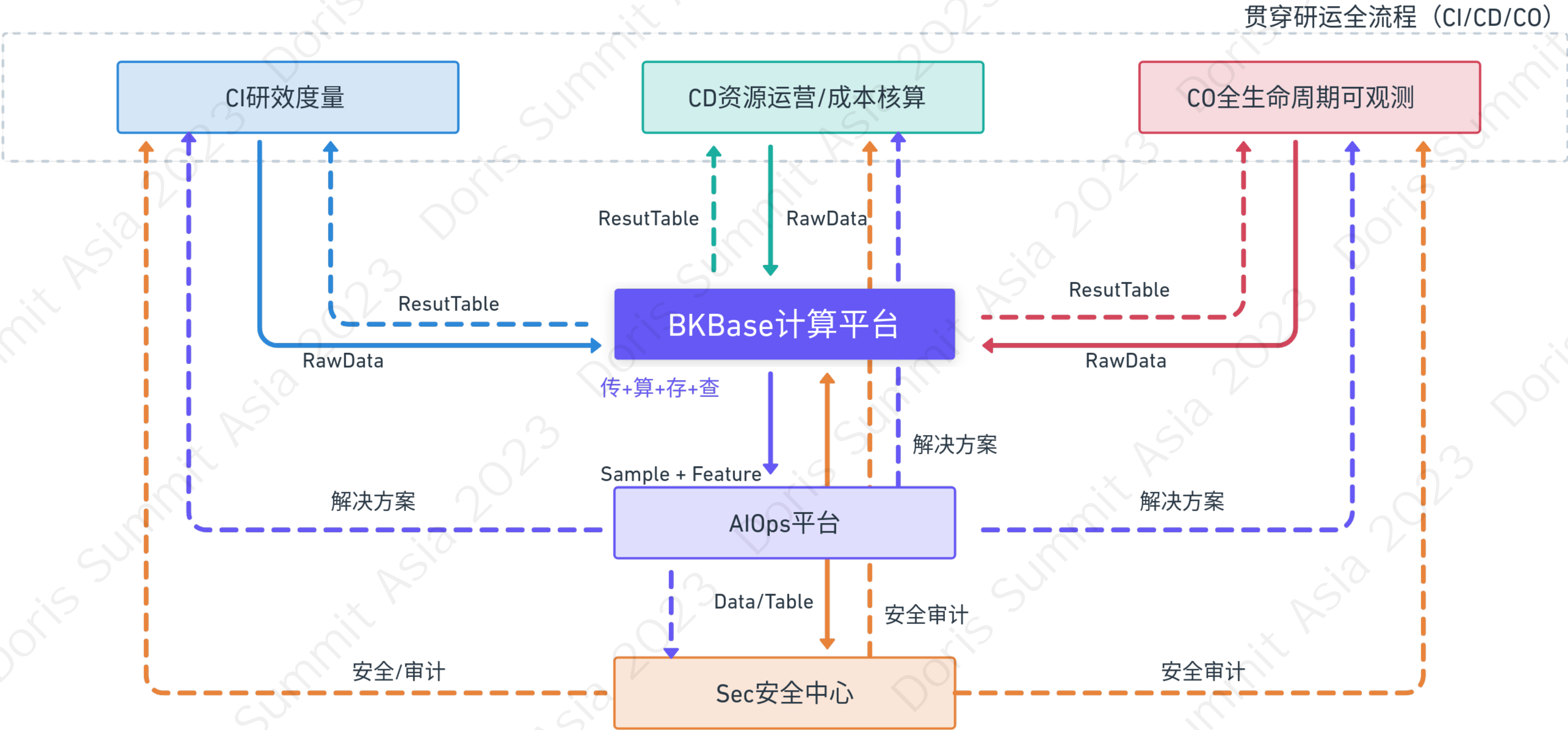
介绍

腾讯蓝鲸智云，简称蓝鲸，是腾讯互动娱乐事业群 (Interactive Entertainment Group, 简称 IEG) 技术运营部自主研发的一套用于构建企业研发运营一体化的 PaaS 解决方案

整体结构



承上启下



平台运营



4w+



3w+



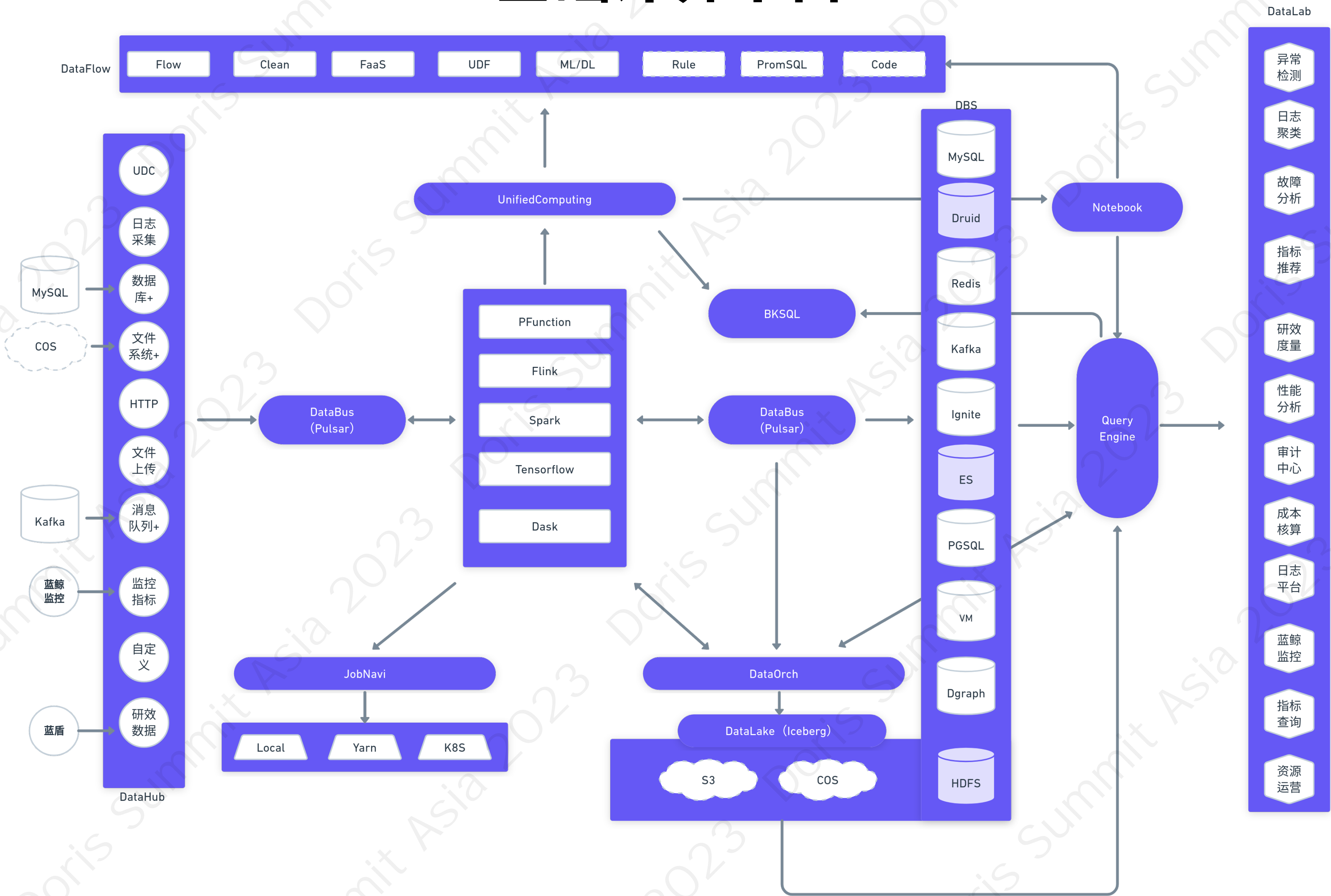
1000+



1.2w+

2 应用探索

基础计算平台



问题初现



Point 1

查询性能



Point 2

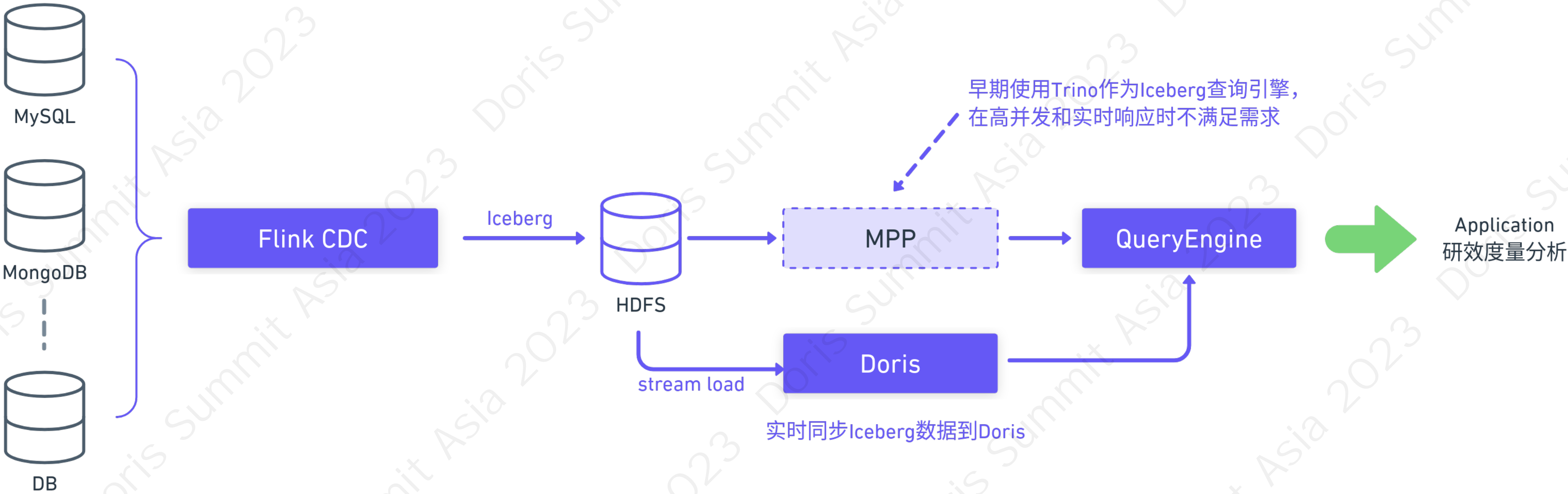
存储成本



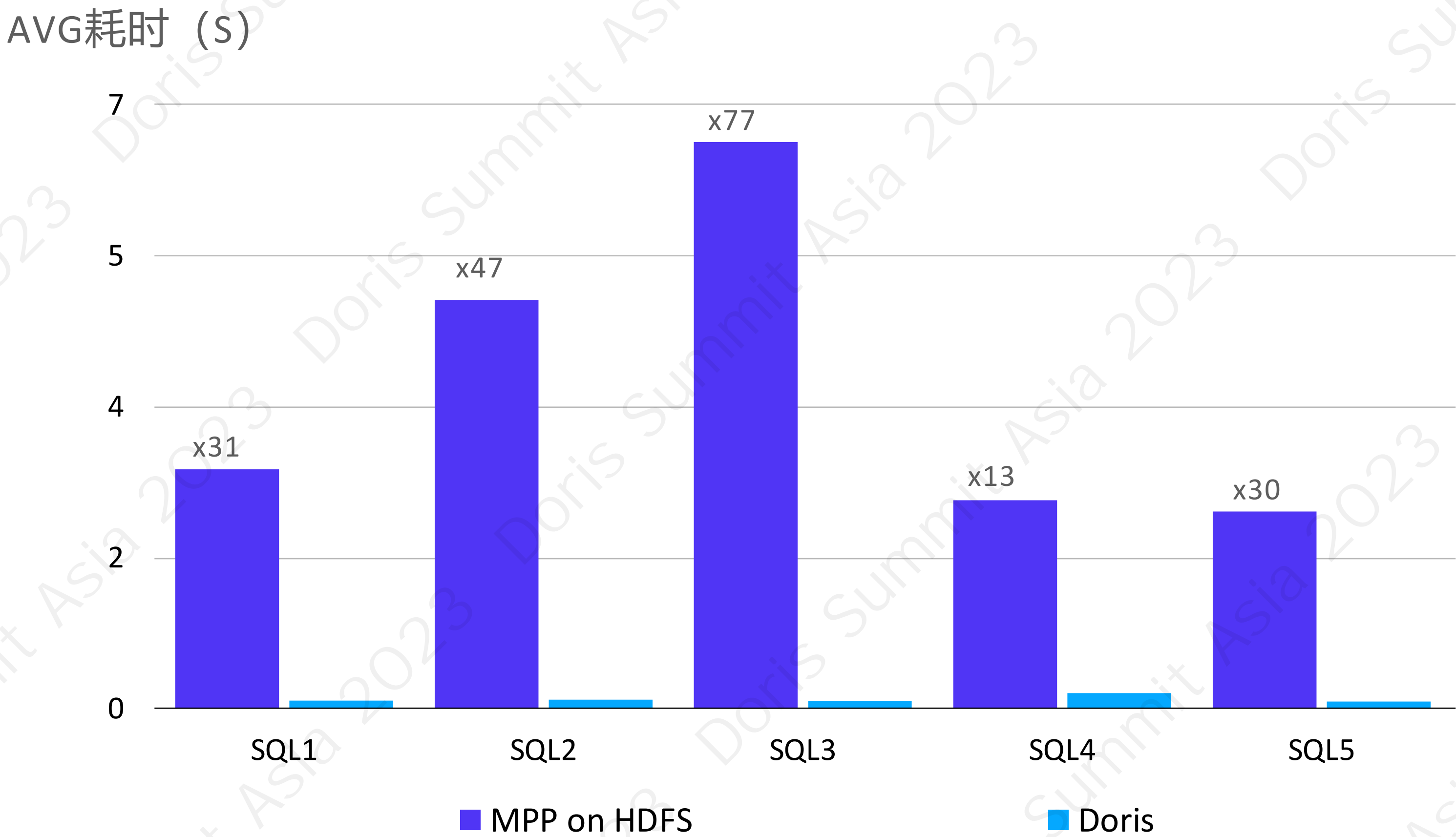
Point 3

存储类型

应用一：研效度量分析

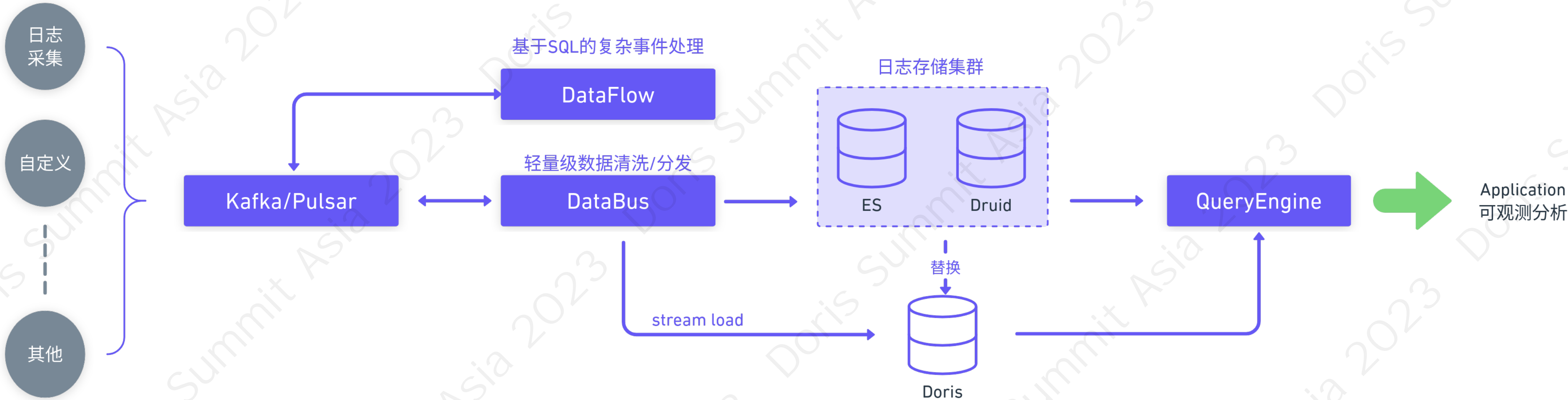


效果



- 验证了线上的场景几个SQL查询，查询性能提升明显
- 高并发查询导致问题也没有了

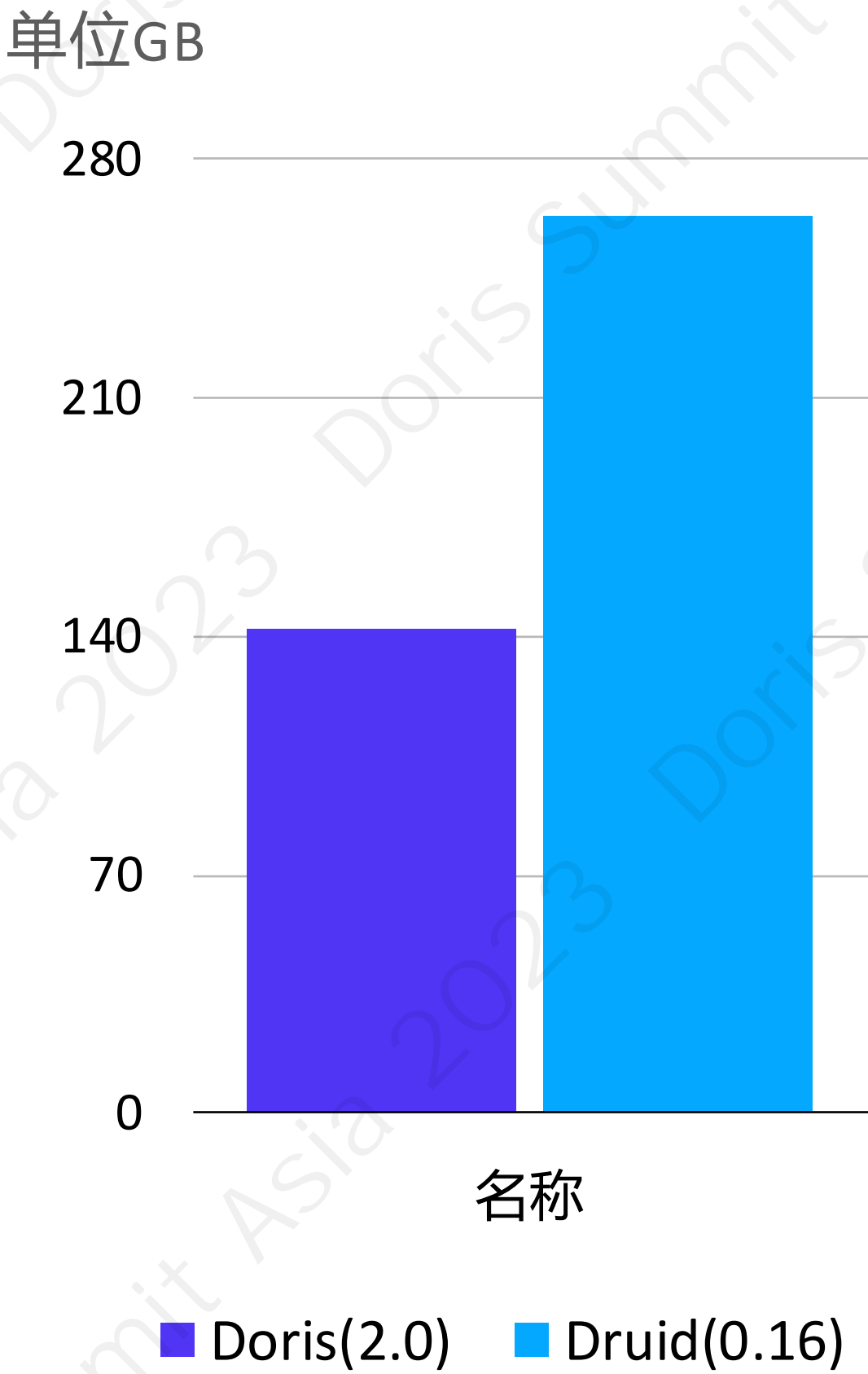
应用二：海量日志实时分析



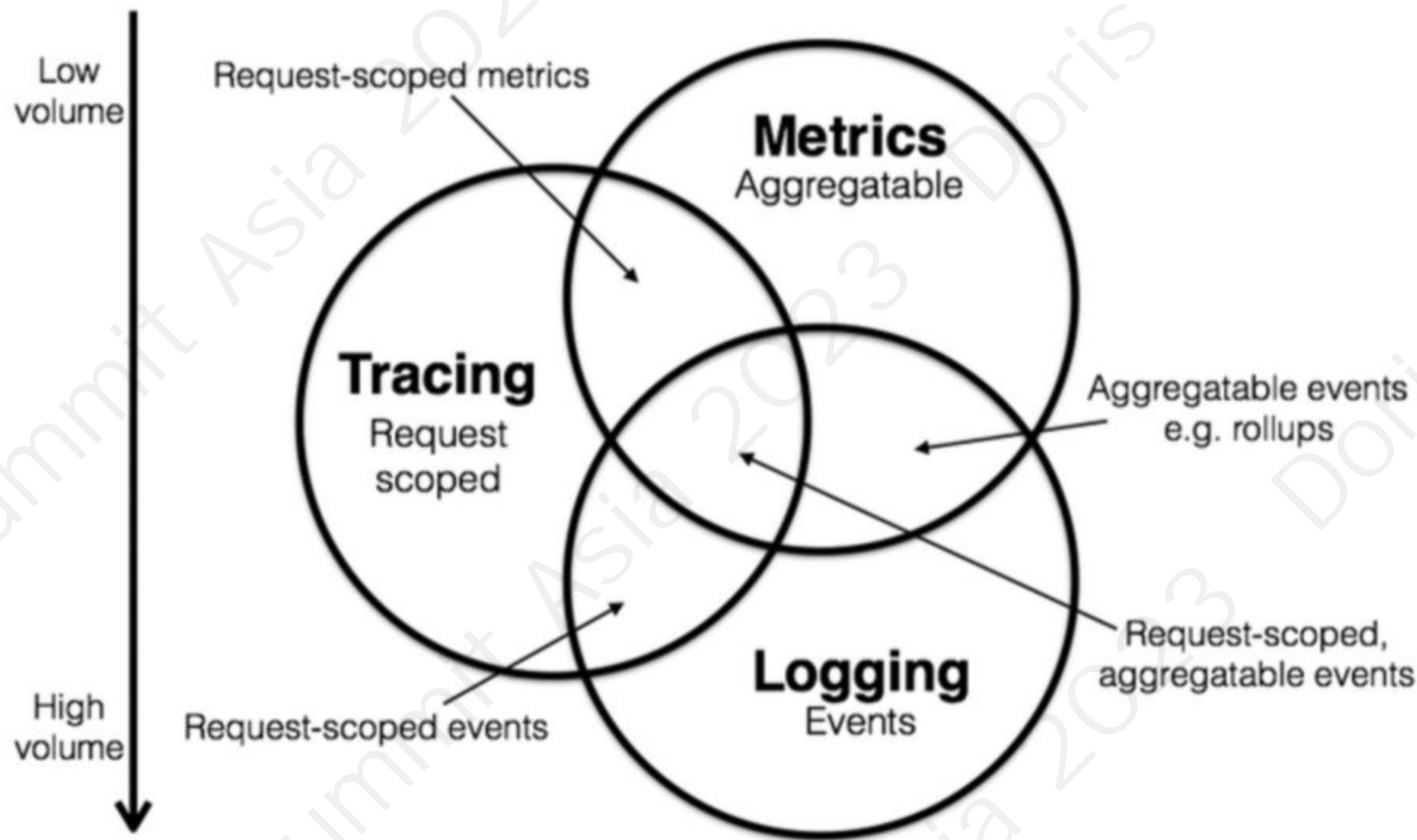
效果



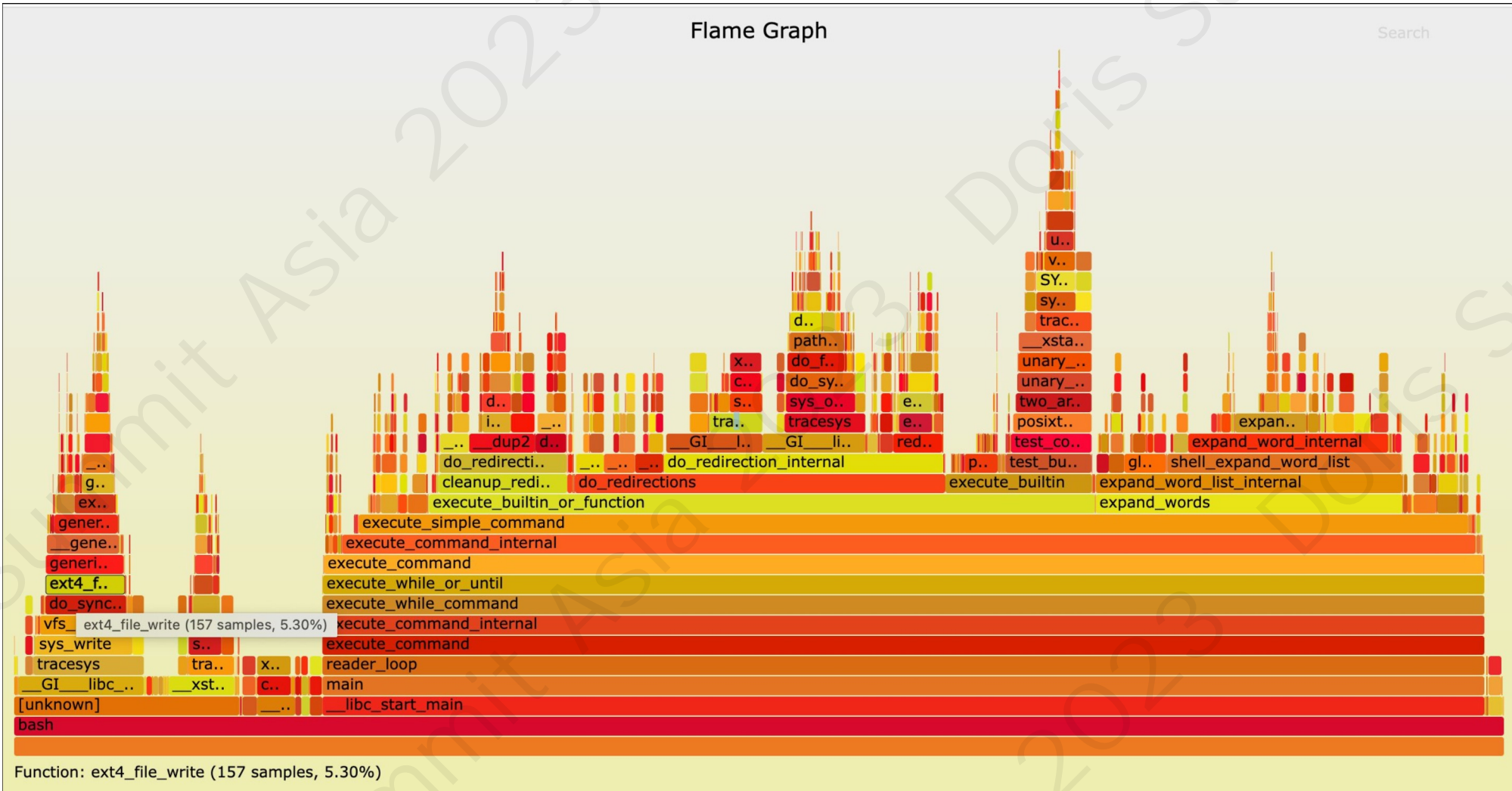
- 使用Doris替换Druid作为日志分析的存储后，系统架构简单，运维成本降低了，同时存储成本也降低了约50%左右
- 除了替换Druid，同时也测试了部分ES的应用，对比后存储成本可以节约50%



应用三：Profiling 性能分析



+



传统解决方案

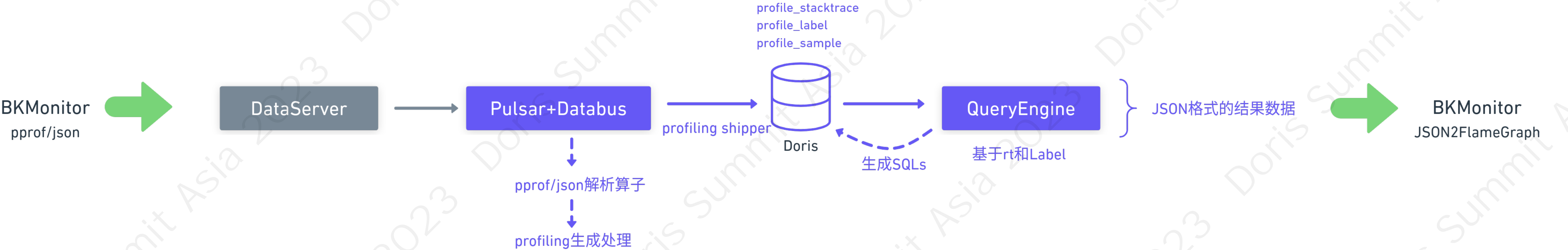
业界主流的开源组件



综合考虑

- 开源社区在Profiling领域组件并不多，而且都不是很成熟，缺点明显
- 组件架构元素比较多，增加技术成本和维护成本
- 项目时间比较紧急，新技术有学习成本

基于 Doris 的 Profiling 分析探索



```
CREATE TABLE `profile_sample` (
  `thedata` DATE NOT NULL COMMENT "thedata",
  `biz_id` varchar(128) NOT NULL COMMENT 'biz_id',
  `app` varchar(128) NOT NULL COMMENT 'app',
  `type` varchar(128) NOT NULL COMMENT 'type(cpu/memory)',
  `period_type` varchar(128) NOT NULL COMMENT 'period_type',
  `period` bigint(11) NOT NULL COMMENT 'period',
  `dtEventTimeStamp` bigint(11) NOT NULL COMMENT 'dtEventTimeStamp',
  `duration_nanos` bigint(11) NOT NULL COMMENT 'duration_nanos',
  `sample_type` varchar(128) NOT NULL COMMENT 'sampleType',
  `stacktrace` varchar(12800) NOT NULL COMMENT 'stackTrace',
  `value` bigint(11) NOT NULL COMMENT 'value'
)COMMENT "profile sample明细表"
PARTITION BY RANGE(`thedata`)
(
  PARTITION `p20230424` VALUES LESS THAN ("20230425"),
  PARTITION `p20230425` VALUES LESS THAN ("20230426"),
  PARTITION `p20230426` VALUES LESS THAN ("20230427"),
  PARTITION `p20230427` VALUES LESS THAN ("20230428")
)
DISTRIBUTED BY HASH(biz_id) BUCKETS 128
PROPERTIES(
  "replication_num" = "1"
);
```

Profiling 数据处理

🔍 原始数据
数据ID: 1579811
profiling_test (profiling_test)

```
{
  "app": "sample-app",
  "biz_id": "10000",
  "type": "contentions",
  "data": {
    "CgQIARACCgQIAxAEg4IARCAgIACGIDgtAUoBRoUCAIQgMD7mur/HxiA4Pua6v8fKAYaGggDEICAgPv/////wEYgKCA+/////ASgHMgAyC2NvbhRlbnRpb25zZMgVjb3VudDIFZGVsYXkyC25hbm9zZWVbmRzMgsvZ2luX3NlcnZlcjIGW3Zkc29dMgpbdnN5c2NhbmGxdSNatqqT59KHDF1oECAEQAmABcAA="
  }
}
```

清洗逻辑

保存模板 拉取默认模板 清洗算子指南

序号	数据输入	清洗算子	算子参数	操作
1	原始数据	json反序列化	中间结果 json	🔍 +
2	json	赋值	<div>批量导入 批量导出</div> <div><div>key app 字段 app 类型 string(...)</div><div>key biz_id 字段 biz_id 类型 string(...)</div><div>key type 字段 type 类型 string(...)</div></div>	🔍 + 🗑
3	json	取值	key data 中间结果 origin_profiling_d... <input type="checkbox"/> 默认值	🔍 + 🗑
4	origin_profiling_...	base64解码	中间结果 profiling_data	🔍 + 🗑
5	profiling_data	pprof反序列化	中间结果 samples	🔍 + 🗑
6	samples	遍历	列表遍历 中间结果 sample	🔍 + 🗑
7	sample	赋值	<div>批量导入 批量导出</div> <div><div>key period_type 字段 period_type 类型 string(...)</div><div>key period 字段 period 类型 string(...)</div><div>key time 字段 time 类型 long</div><div>key duration_n 字段 duration_nanos 类型 string(...)</div><div>key stacktrace 字段 stacktrace 类型 string(...)</div><div>key labels 字段 labels 类型 string(...)</div><div>key sample_type 字段 sample_type 类型 string(...)</div><div>key value 字段 value 类型 string(...)</div></div>	🔍 + 🗑

结果预览

全局调试

Profiling 查询分析

样例1: 获取label列表

请求

```
1 {
2   "sql": "{\api_type\": \labels\, \api_params\": {\biz_id\": \100\, \app\": \ygame\, \type\": \cpu\}, \result_table_id\": \591_test\}",
3   "bkdata_authentication_method": "token",
4   "bk_username": "admin",
5   "bk_app_code": "testApp",
6   "bkdata_data_token": "testToken",
7   "prefer_storage": "doris",
8   "bk_app_secret": "testAppSec"
9 }
```

样例2: 获取label values列表

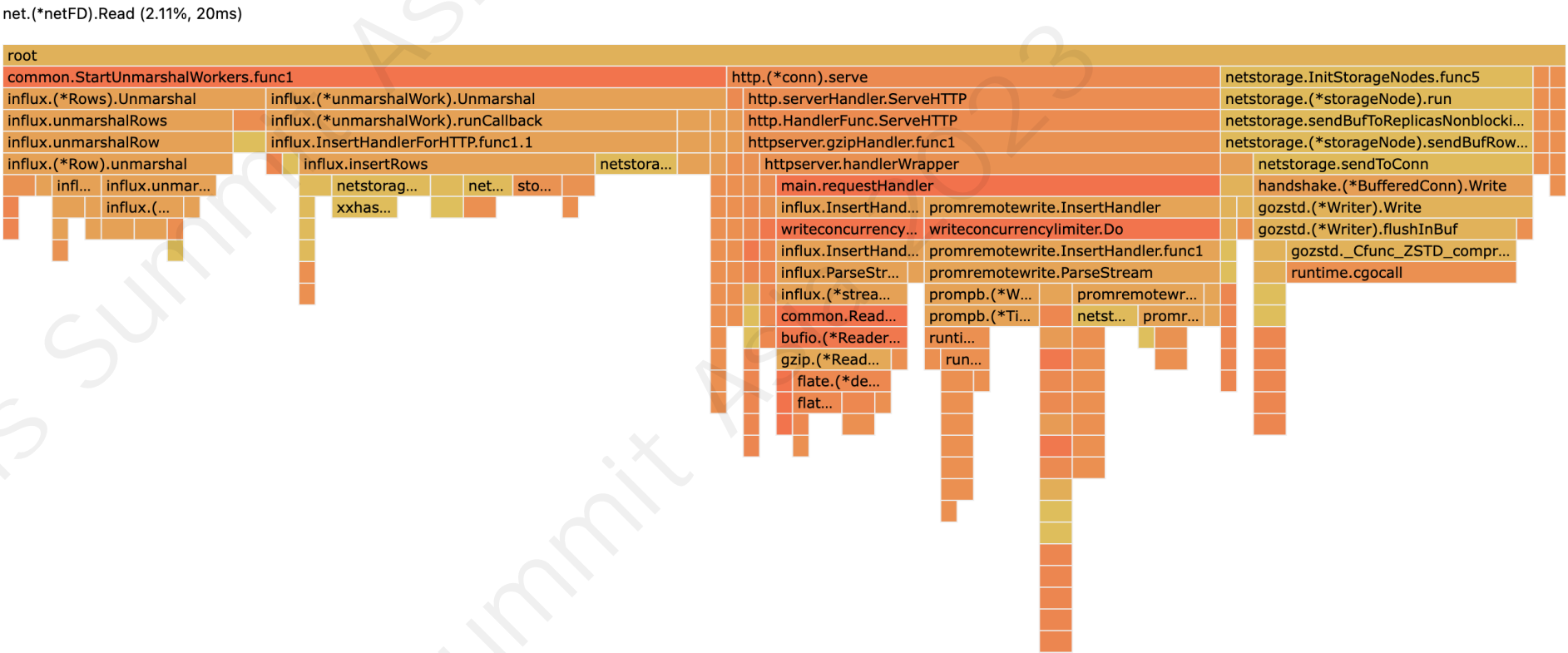
请求

```
1 {
2   "sql": "{\api_type\": \label_values\, \api_params\": {\biz_id\": \100\, \app\": \ygame\, \type\": \cpu\, \label_key\": \region\}, \result_table_id\": \591_test\}",
3   "bkdata_authentication_method": "token",
4   "bk_username": "admin",
5   "bk_app_code": "testApp",
6   "bkdata_data_token": "testToken",
7   "prefer_storage": "doris",
8   "bk_app_secret": "testAppSec"
9 }
```

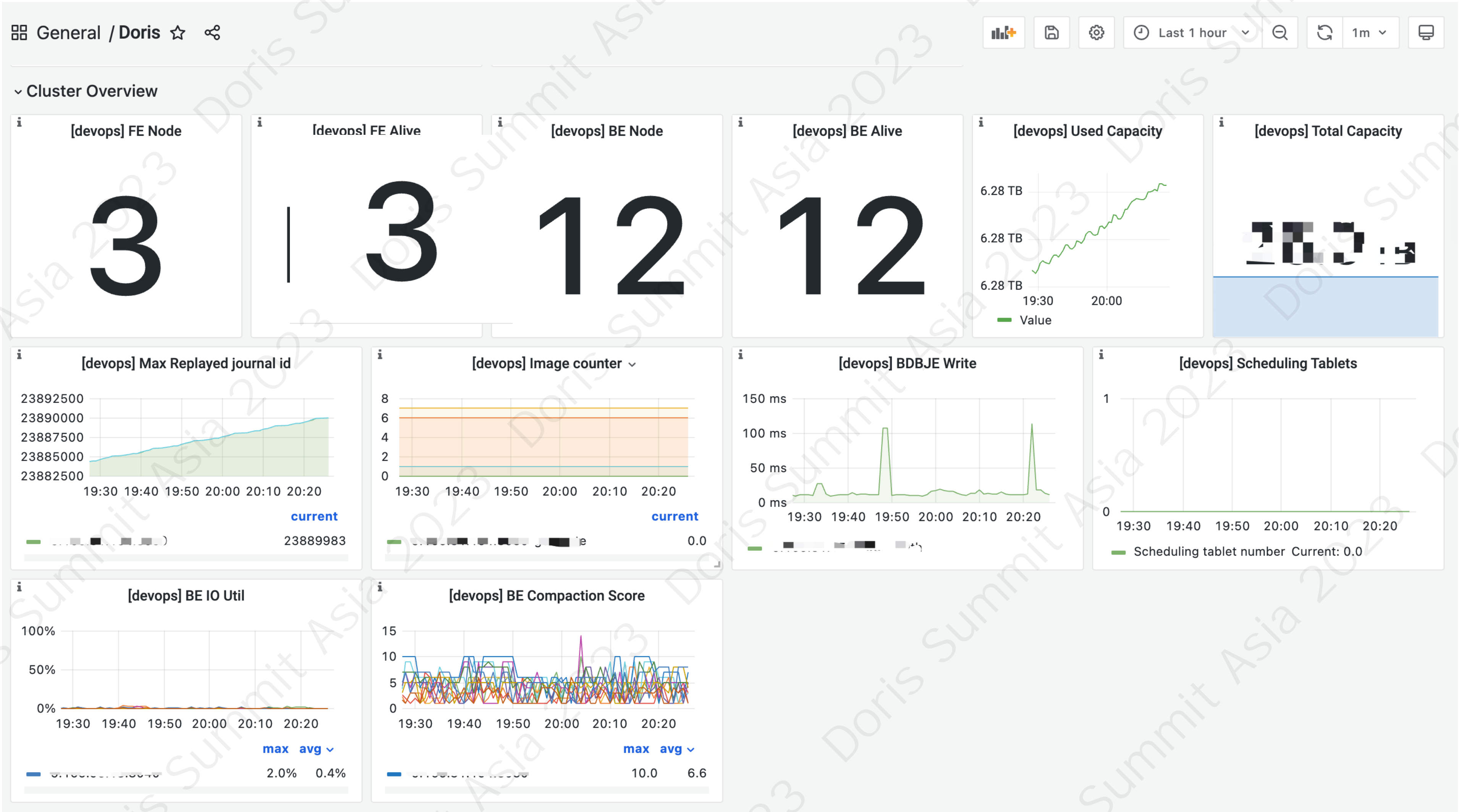
样例3: 获取 Sample 列表

请求

```
1 {
2   "sql": "{\api_type\": \query_sample\, \api_params\": {\start\": 1609430400000, \end\": 1684425599000, \biz_id\": \591\, \app\": \bkbase\, \type\": \cpu\, \label_filter\": {\traceid\": \xxx1\}}, \result_table_id\": \2005000727_pprof_test20\}",
3   "bkdata_authentication_method": "token",
4   "bk_username": "admin",
5   "bk_app_code": "testApp",
6   "bkdata_data_token": "testToken",
7   "prefer_storage": "doris",
8   "bk_app_secret": "testAppSec"
9 }
```



监控



3 未来计划

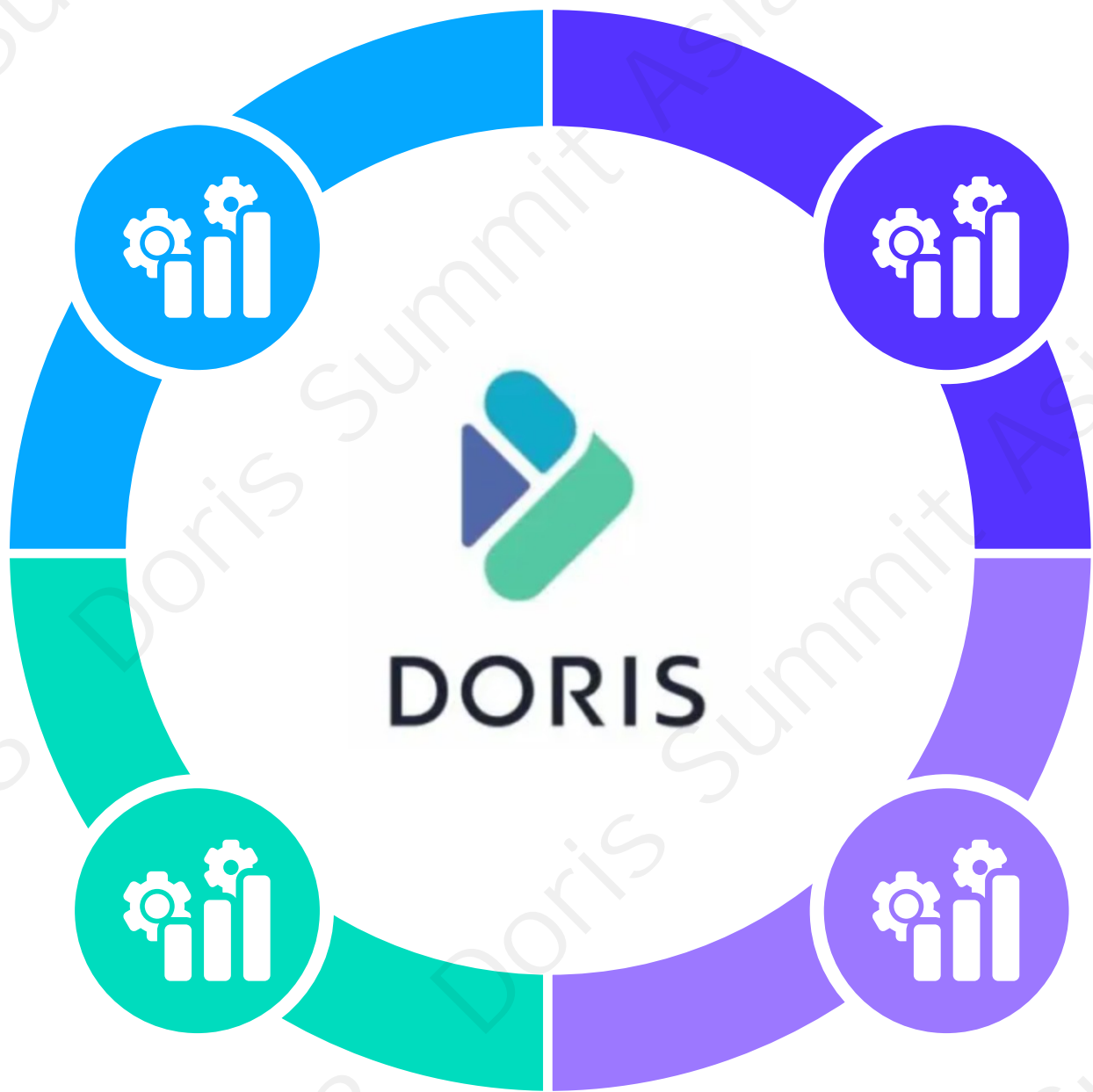
进一步探索场景

数据湖引擎

Iceberg查询切换到Doris外表

探索机器学习领域应用

样本库和特征库切换到Doris的方案



可观测链路跟踪

基于Doris设计Tracing的存储和查询方案

探索 Doris 在 Metric 的应用

指标数据基于Doris存储与PromQL查询方案



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>