

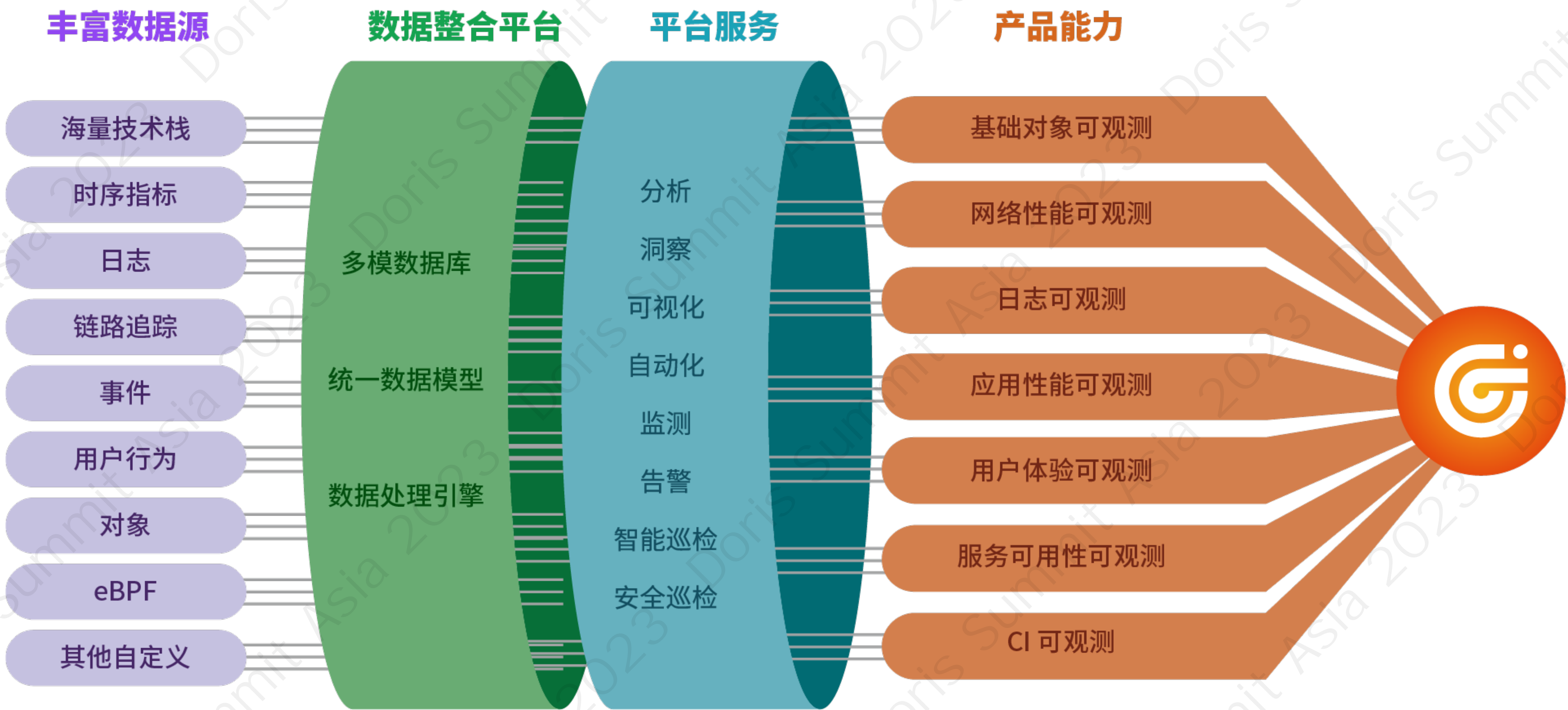
观测云-基于Apache Doris 构建 新一代的可观测平台

蒋烁淼

观测云 - CEO

1 什么是观测云

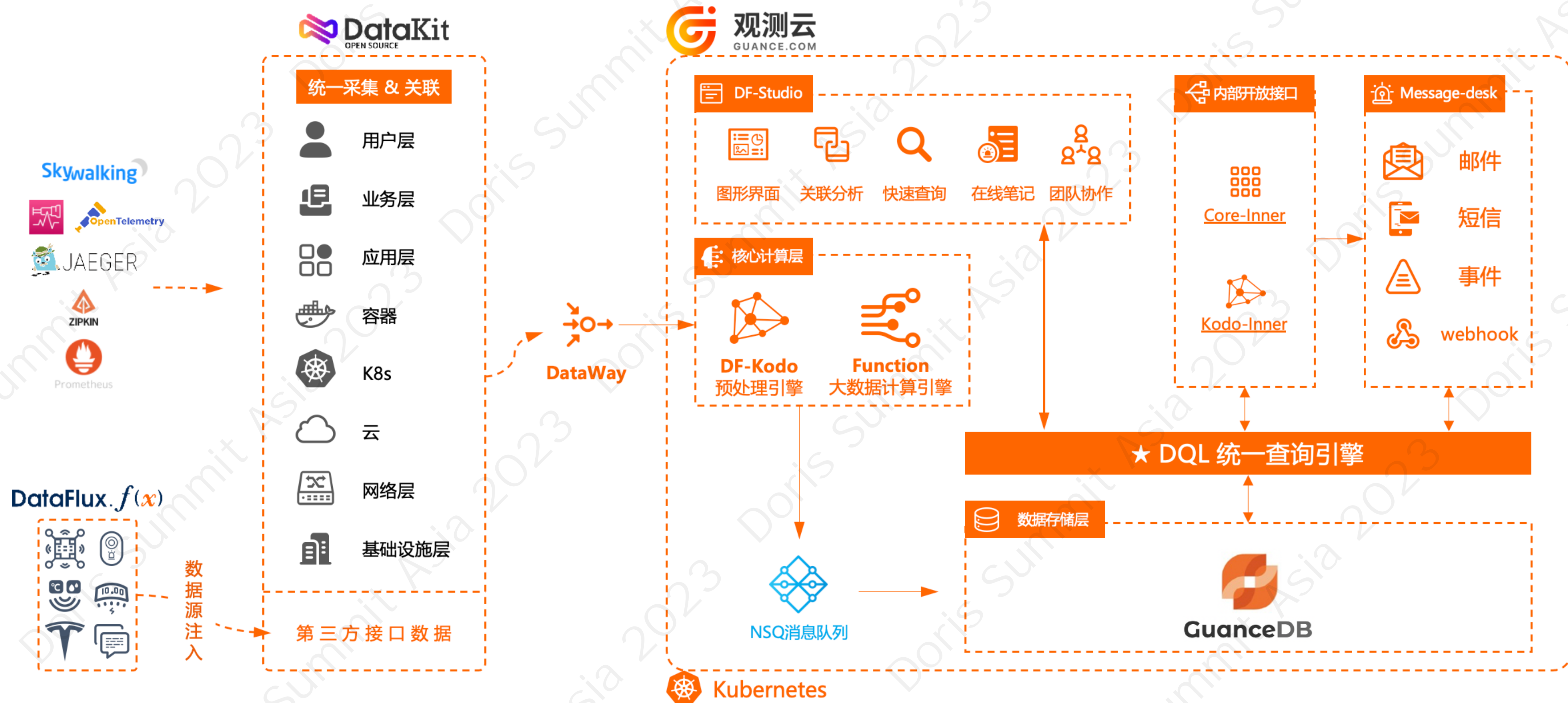
观测云的原理



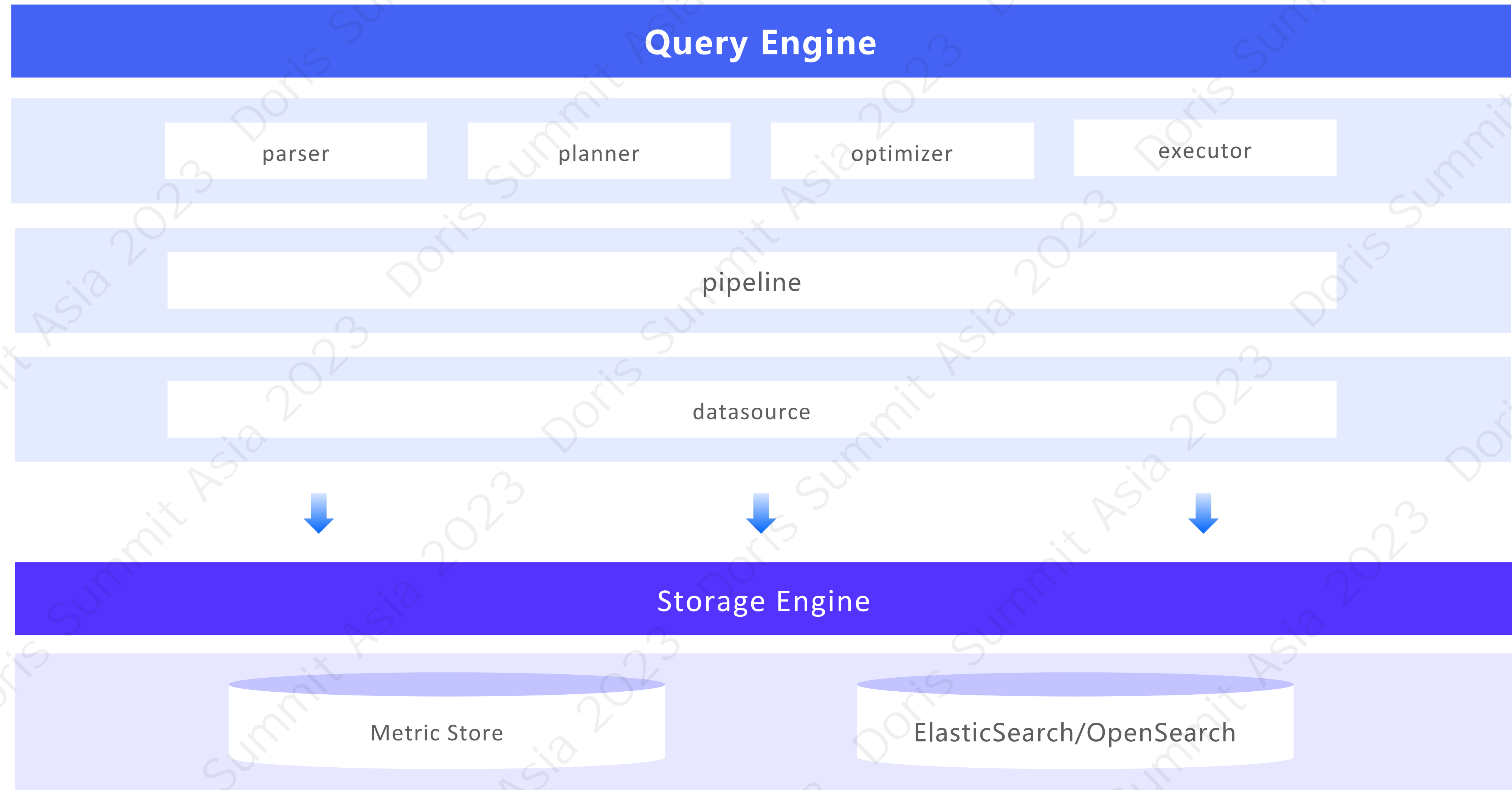
观测云的能力



观测云的架构



GuanceDB 原有架构



观测云的以往痛点



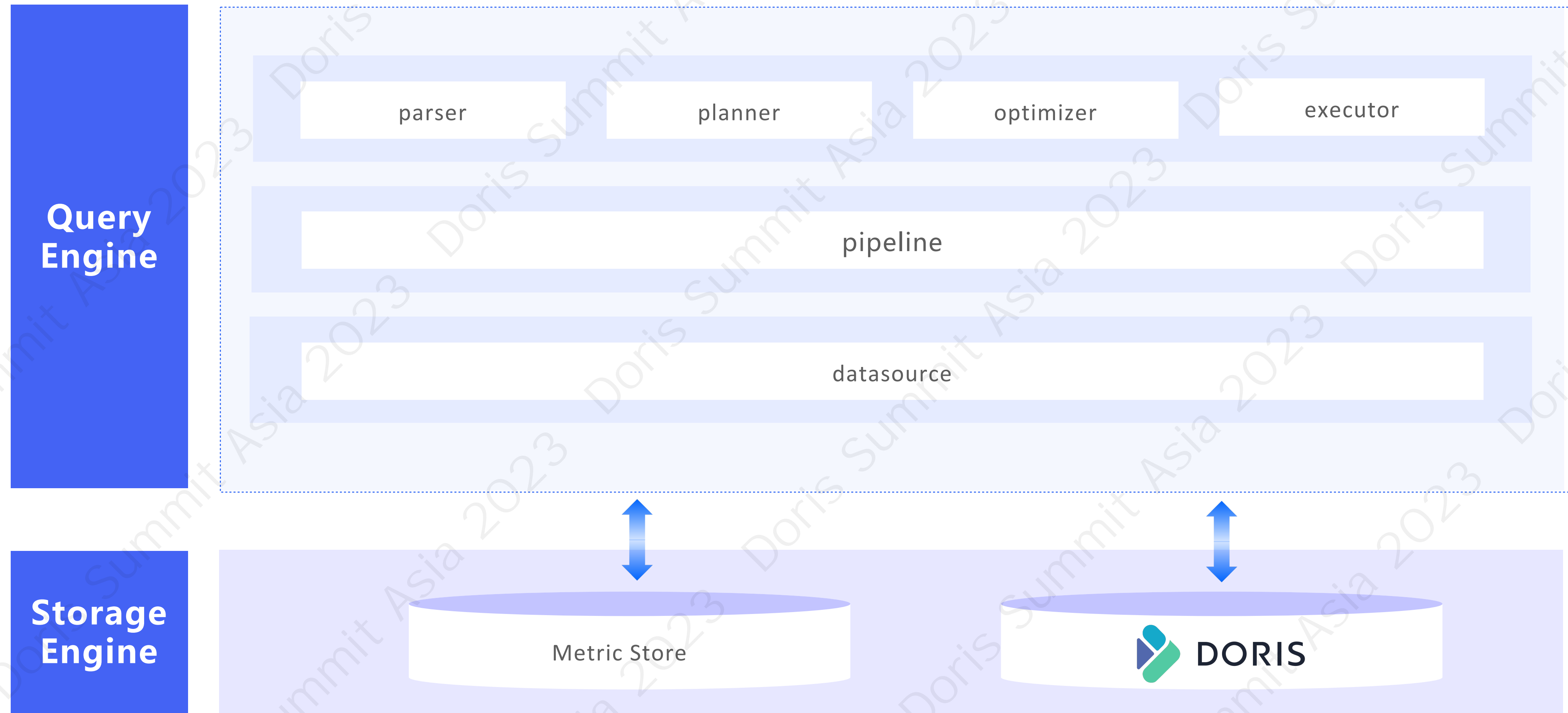
- 写入占用集群性能过大
- 面对用户的自定义数据对 SchemaLess 支持有限
- ElasticSearch (Opensearch) 的面对超海量数据的聚合查询能力有限
- 多租户情况下的隔离效果差
- 高级分析能力增强有限

其他方案为什么不可行

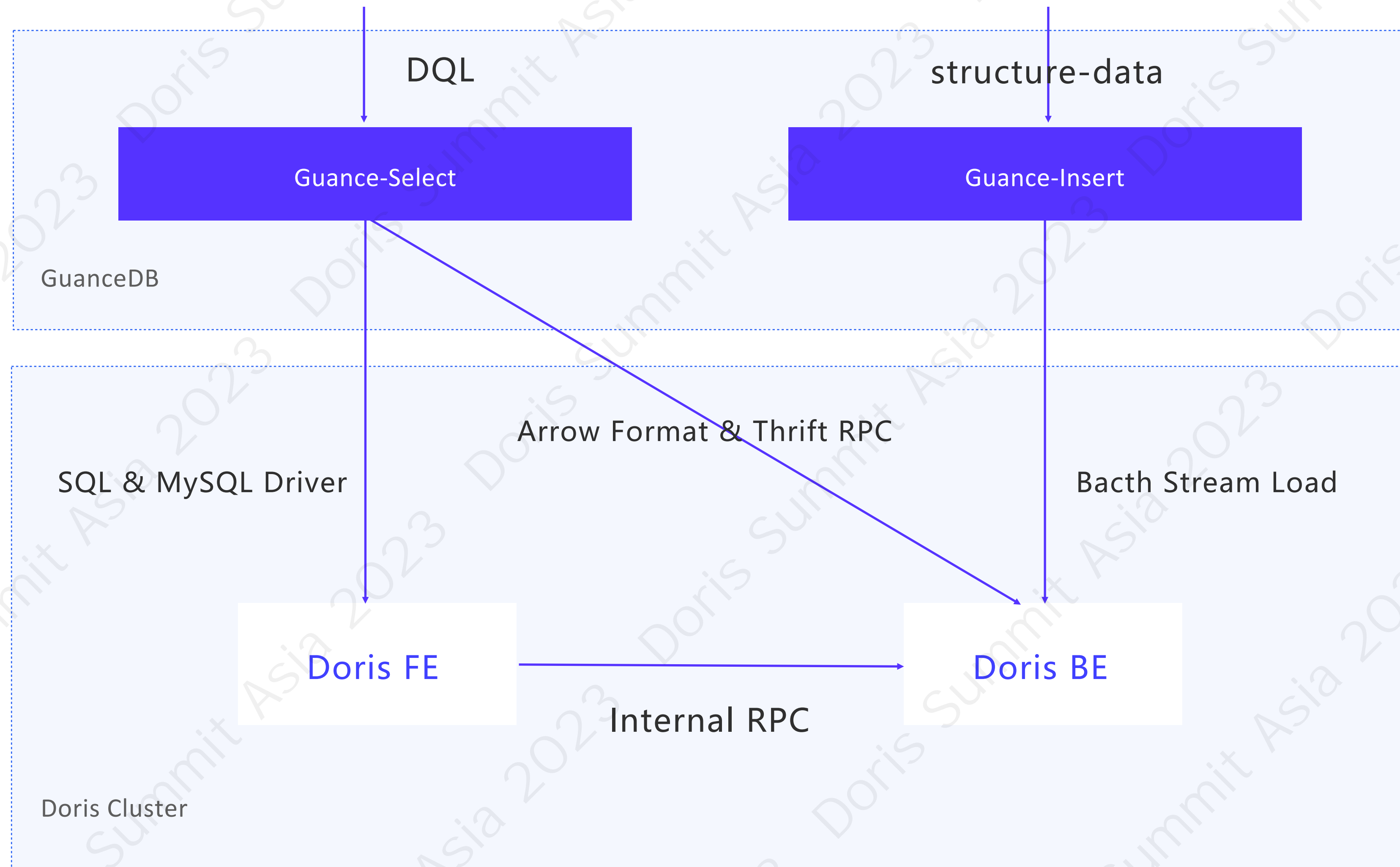


- SaaS 需要支持一定的查询高并发
- 日志场景，倒排索引是必须的
- SchemaFix 并不适合可观测性场景
- 多租户情况下，大量的元数据是否造成性能问题
- 是否能支持面向可观测性场景的数据查询能力

GuanceDB 整体架构



GuanceDB for Doris



DQL vs SQL

DQL查询

- `M::cpu:(avg(cpu_usage)) { cluster = "testing" } [1h::5m:avg]`
- `M::httpserver:(derivative(request_count)) by account_id sorder by account_id desc slimit 10`
- `L::(L::* limit 1000):(count(__docid) as count, collect(__docid, 50)) by drain(message, 0.3) as sample sorder by count desc`

语意相似的 SQL

- 模拟 step + rollup: `select ts, avg(cpu_usage) from (select cast(floor(milliseconds_timestamp(time)/300) * 300 AS bigint) as ts, avg(cpu_usage) as cpu_usage from M where time >= now()-1h and time < now() and cluster = 'test' group by ts, instance) group by ts`
- sorder by 无法实现, 缺乏 derivative 等时序算子, drain 日志聚类无法实现

DQL 优势

- 专为可观测场景设计, 表达简洁, 功能强大
- 支持所有查询负载, Metrics、Logs、Metrics、Objects... 等 All In One

Variant 设计

可观测场景为什么需要 Variant?

业务场景数据类型多变、降低运维负担

动态 schema 为什么不行?

我们在 ES 中遇到的问题:

- 类型冲突
- 字段数量上限

Variant 的优势:

- 每个 Partition 动态创建, 数据变化后字段可以失效
- 没有类型冲突, 支持多类型合并升级

GuanceDB 中怎么适配?

- 查询时根据表 Schema 判断是否是 Variant 字段, 以及根据用户 DQL 中的意图推断期望的数据类型
- 完全用户透明, DQL 和写入都无需改动, 我们自动处理 Variant

聚合采样

为什么需要聚合采样?

- 原始数据量级大，查询开销极高
- 实际业务确实不需要精确计数

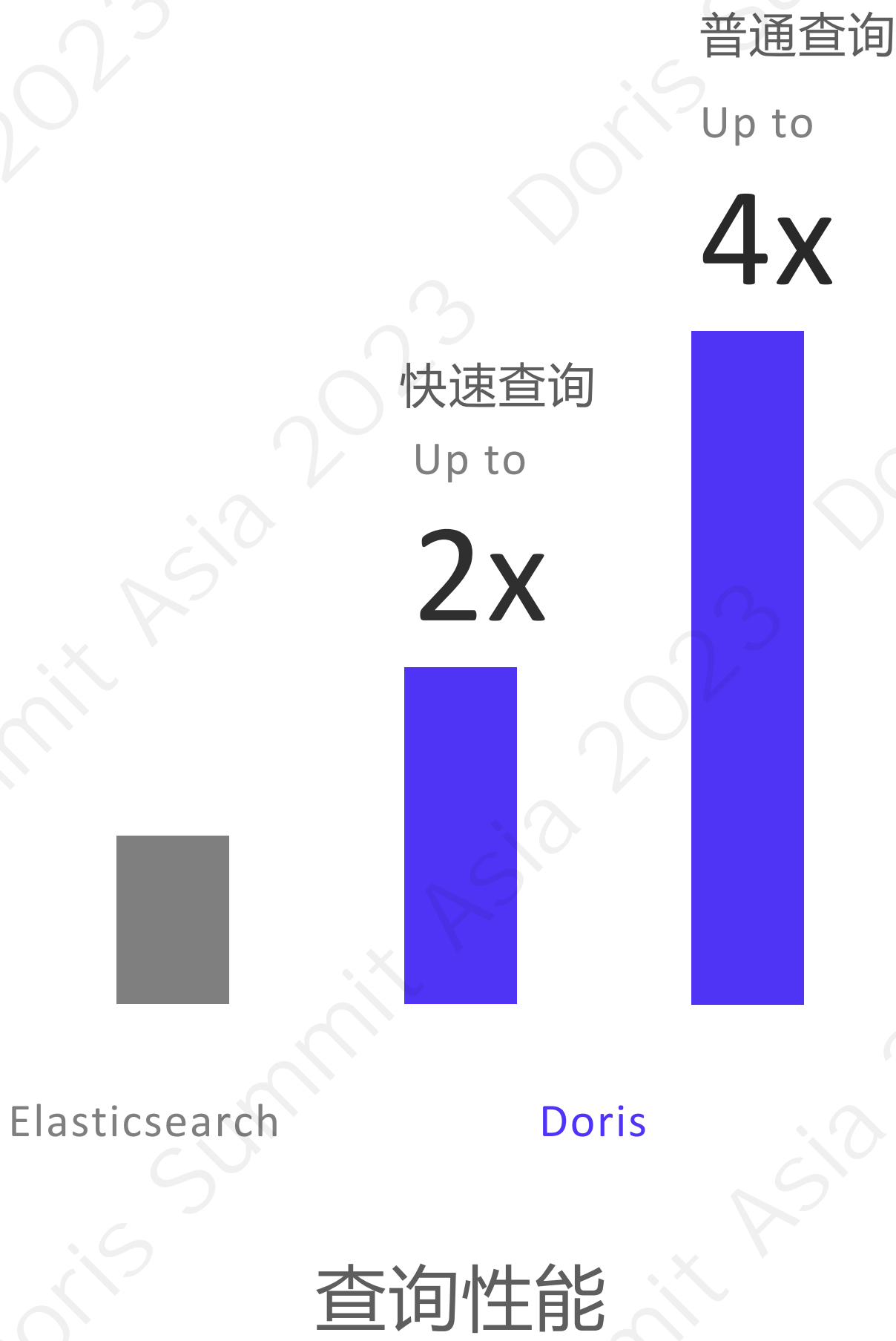
GuanceDB 中怎么做采样?

- 评估查询起始范围内原始数据行数
- 当原始行数大于 100w 时开启采样，固定采样行数为 100w 来计算采样率
- 利用 doris tablesample 能力进行采样
- 根据不同的算子自动对结果按照比例进行伸缩，主要是 sum + count

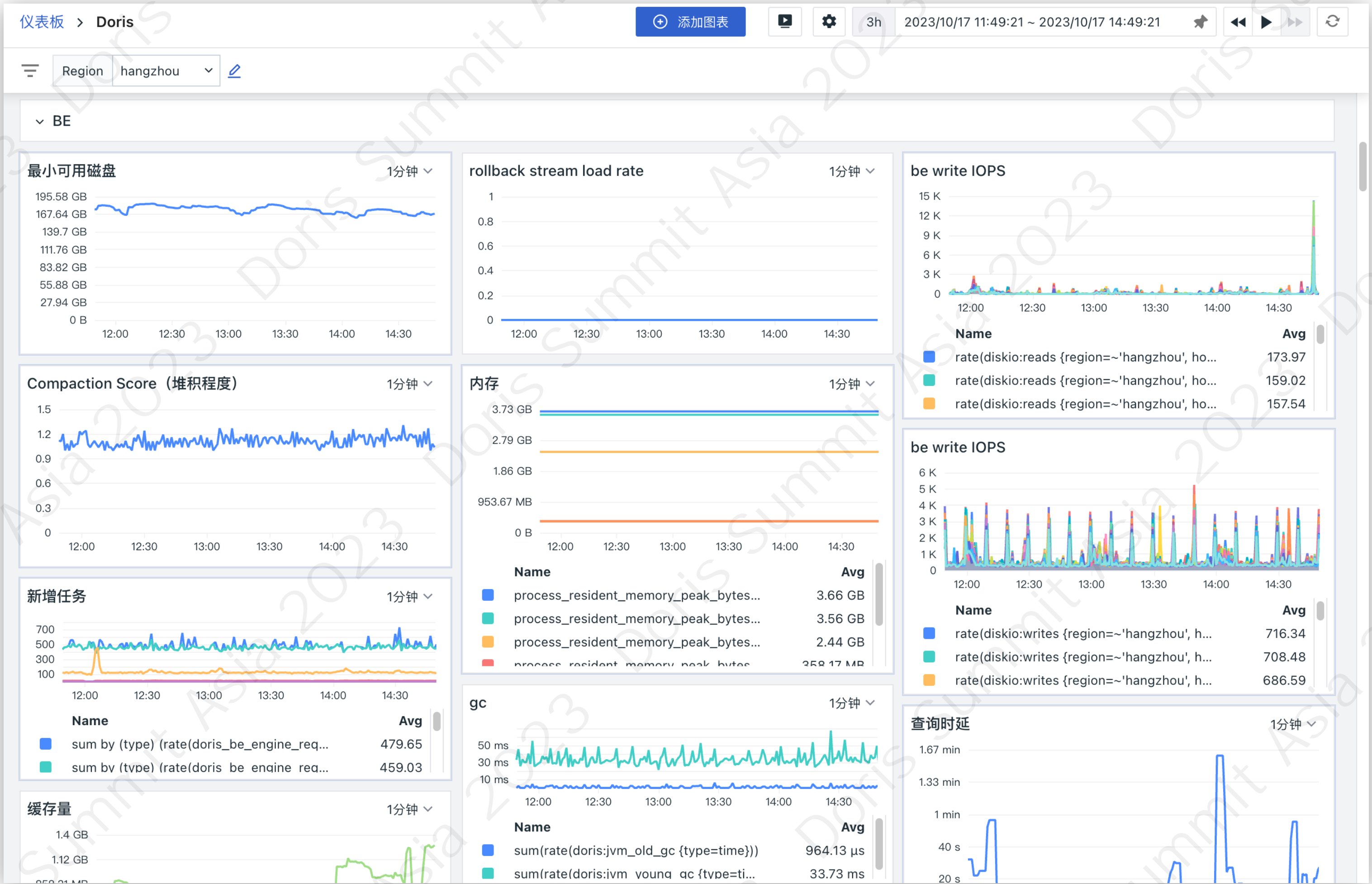
最终的效果:

- 查询性能提升 20 倍
- 业务透明，自动根据数据量进行聚合采样，特殊需要可以关闭采样

性能、成本估算



副产品-Doris 的完整监控观测方案



可能是目前最好的监控观测方案

- 将实时数仓的能力应用到了监控日志及更广泛的可观测性场景
- 结合观测云的产品力，无需再针对 Doris 进行二次开发即可拥有完整的监控观测能力
- Doris带来的强大实时分析能力将利用Metrics, Logs, Traces等可观测性数据的价值提升了一个维度。
- 整体性价比传统的开源方案提升至少 20 倍的性价比，同时带来更好的用户体验
- 开放 Doris 的数仓能力还可以将可观测性数据与业务数据通过 SQL 进行整合，让可观测性数据发挥更大价值
- 可以更有能力构建更多面向业务场景的实时观测监控场景

好用不贵，极大的提升整个产研团队的效率，提升整个产研团队能力边界

欢迎大家注册使用体验



华为云



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>