

# 可观测性场景下 SelectDB 流式聚合增强

熊豹 观测云资深架构师

## 分享嘉宾 - 熊豹



可观测领域专家。观测云 GuanceDB 数据库研发负责人。  
VictoriaMetrics Top 社区 Contributor、2020 QCon 讲师、曾任知乎可观测和接入层网络负责人。

# 目录

01 观测云存储架构

02 SelectDB 应用与挑战

03 流式聚合系统设计

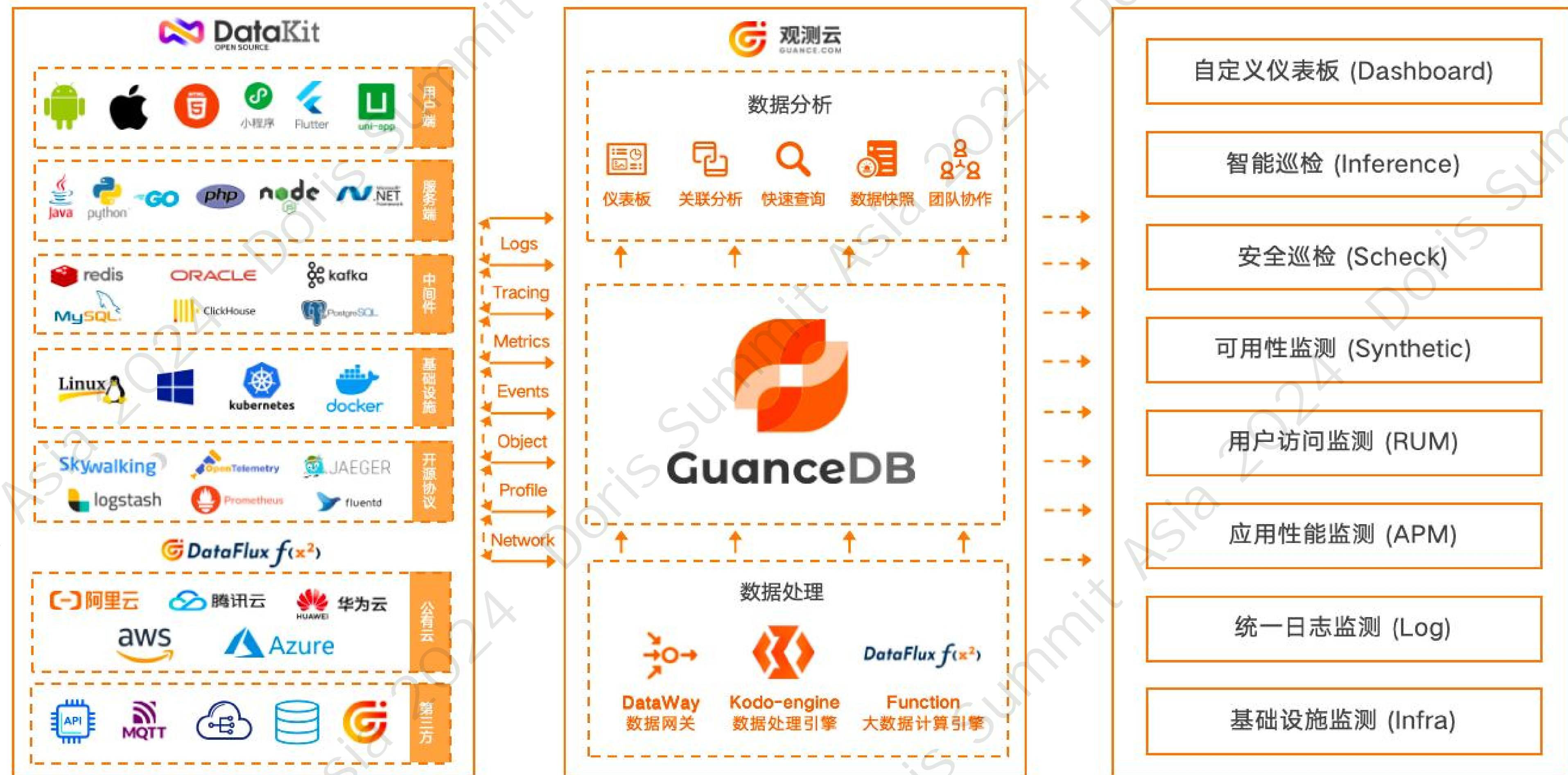
04 展望

01

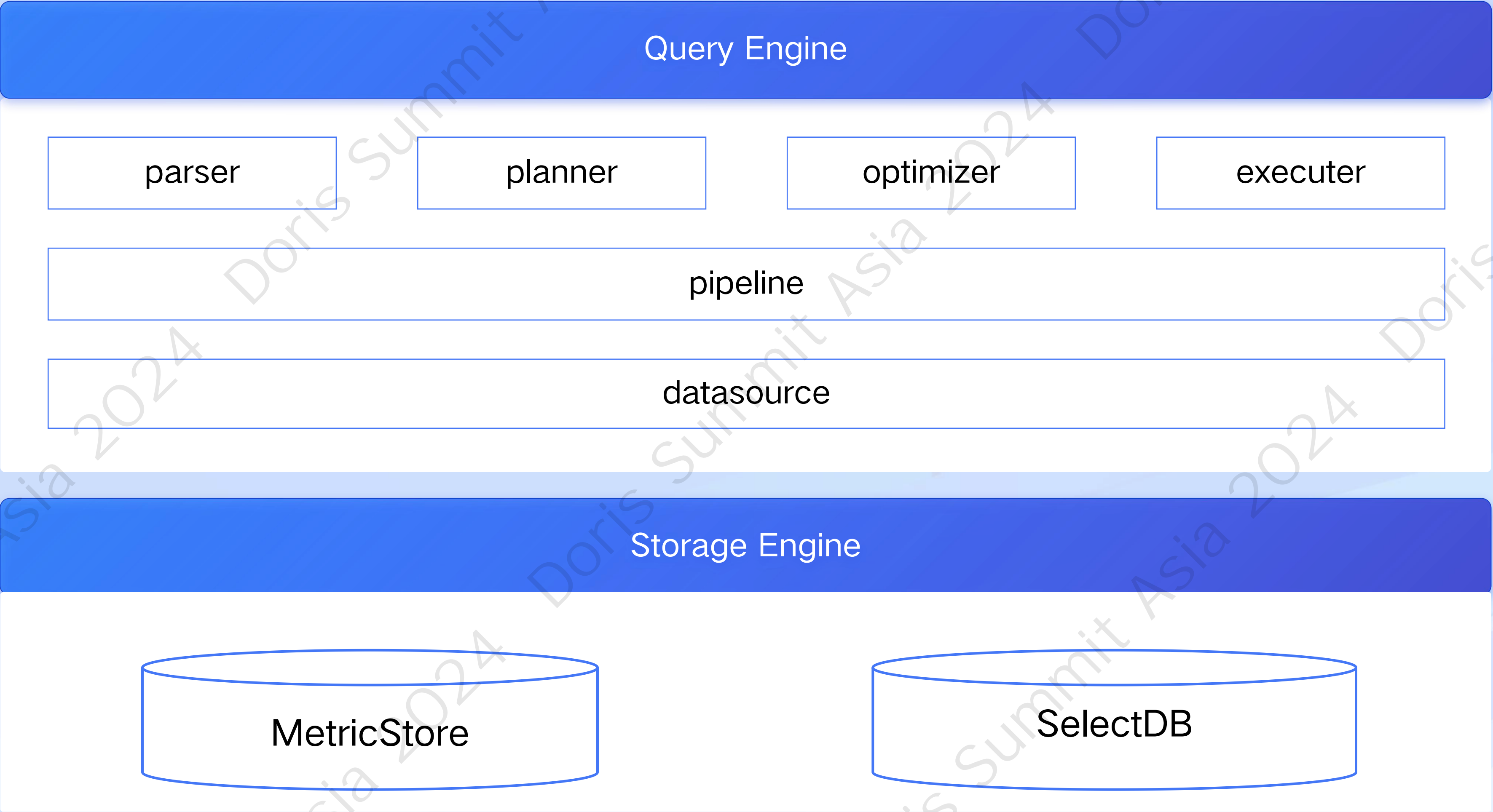
# 观测云存储架构



# GuanceDB 支撑观测云全量业务场景



# GuanceDB 系统架构



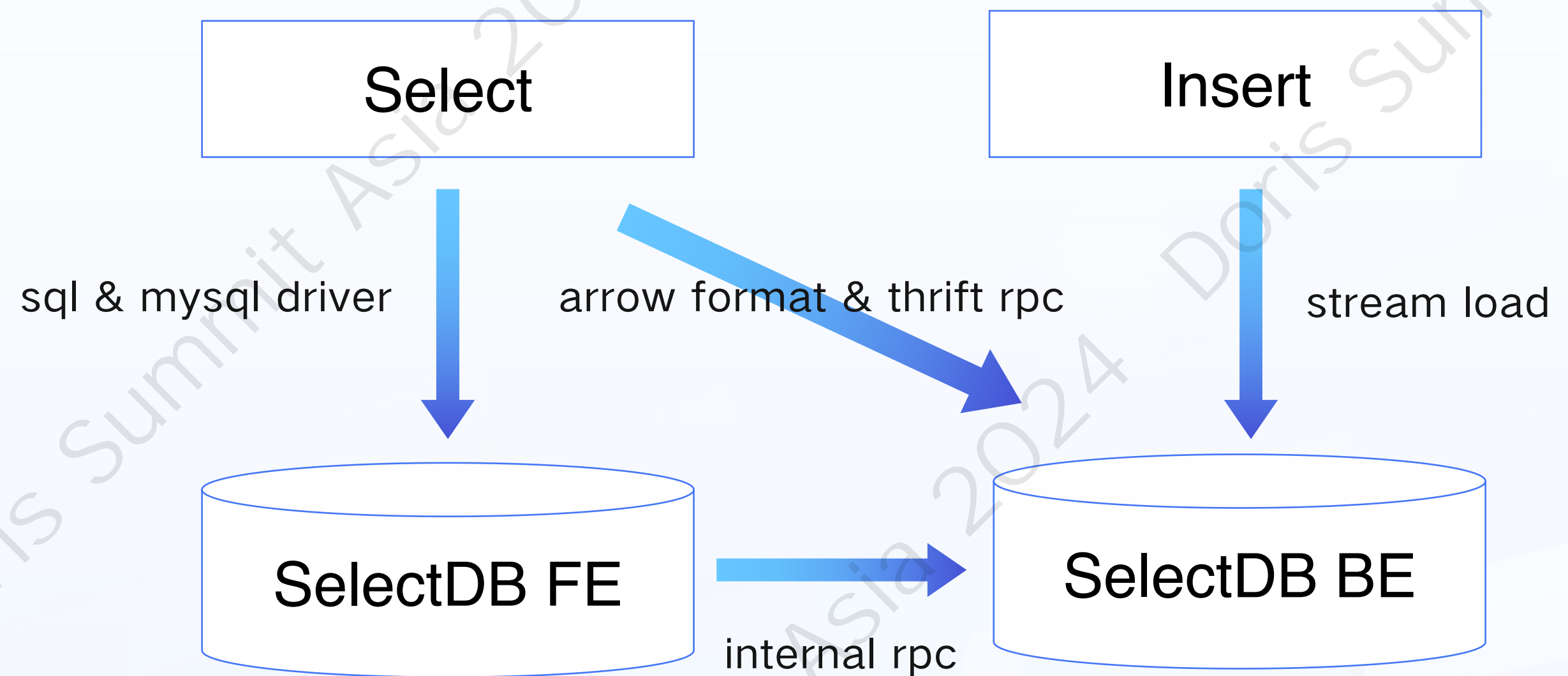
# SelectDB 数据源接入

## 根据语义下推聚合函数

- 查询引擎根据 DQL 查询语义和 SelectDB SQL 的函数支持情况动态选择使用聚合下推或谓词下推的能力

## Thrift 接口降低传输开销

- 当不能下推聚合或只能下推部分谓词时选择使用 SelectDB BE 的 Thrift 接口，通过 Arrow 列存格式降低传输数据的反序列化开销



02

# SelectDB 的应用与挑战



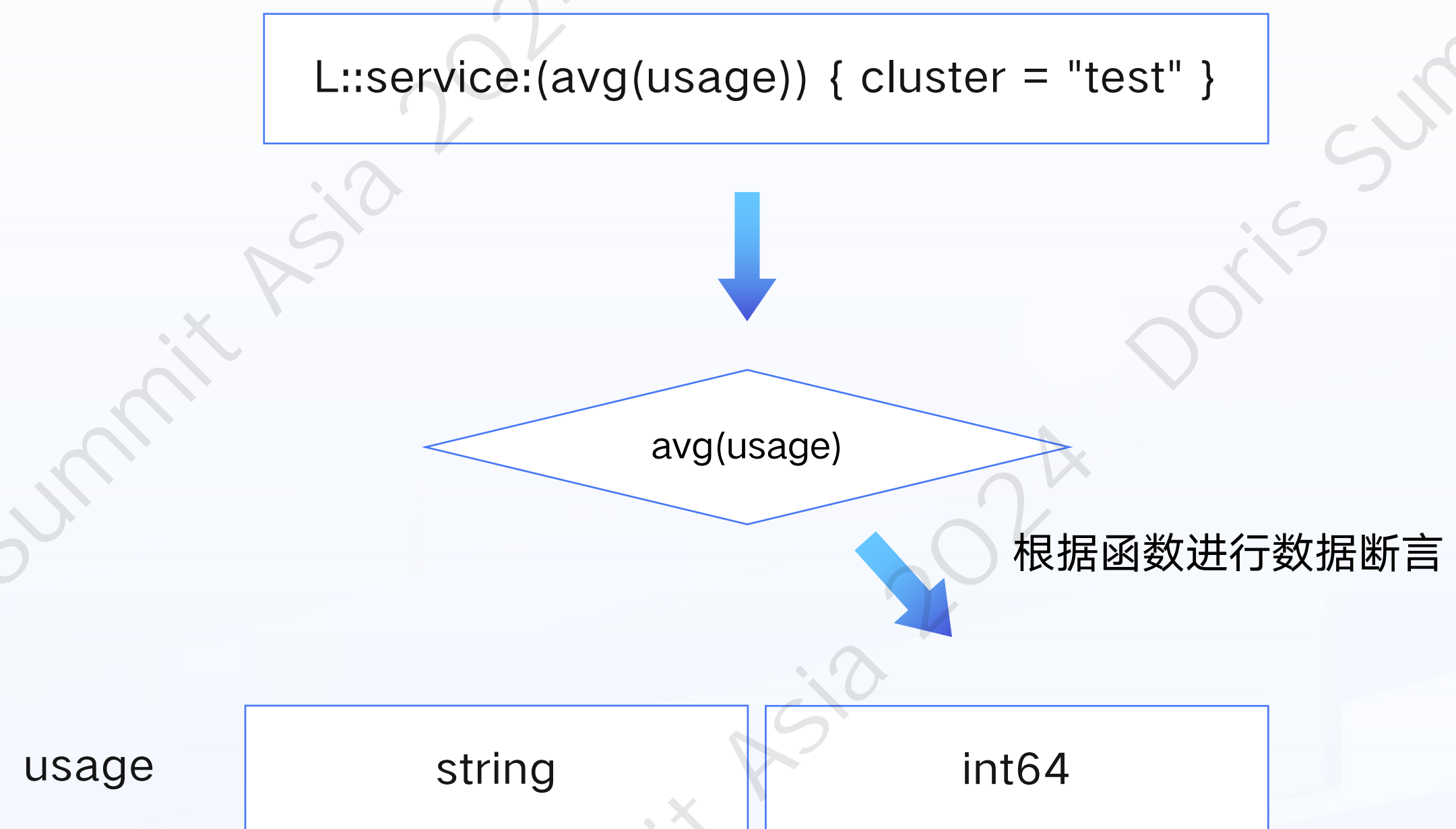
# Dynamic Schema 动态表 VS Variant 可变数据类型

## 相同点

- 都能支持跟随数据维度变化动态维护表结构

## 关键差异

- Dynamic Schema 作用于当前表的完整生命周期，而 Variant 的作用域只在当前的动态分区内
- Dynamic Schema 作用于当前表的完整生命周期，而 Variant 的作用域只在当前的动态分区内



# 自动聚合采样

## 自动采样

- 一些查询原始数据量极大，而且业务层确实不需要精确计数，计算结果包含大概趋势即可，此时耗费数十倍资源进行计算响应速度过慢，费力不讨好

## 落地效果

- 聚合查询性能提升 20+ 倍；业务透明，自动根据数据量进行聚合采样，可手动关闭采样

```
L::service:(avg(usage)) { cluster = "test" }
```

预估扫描数据量

> 1kw

tablesample 1kw rows

< 1kw

Doris

# 挑战：可观测场景下的周期性查询

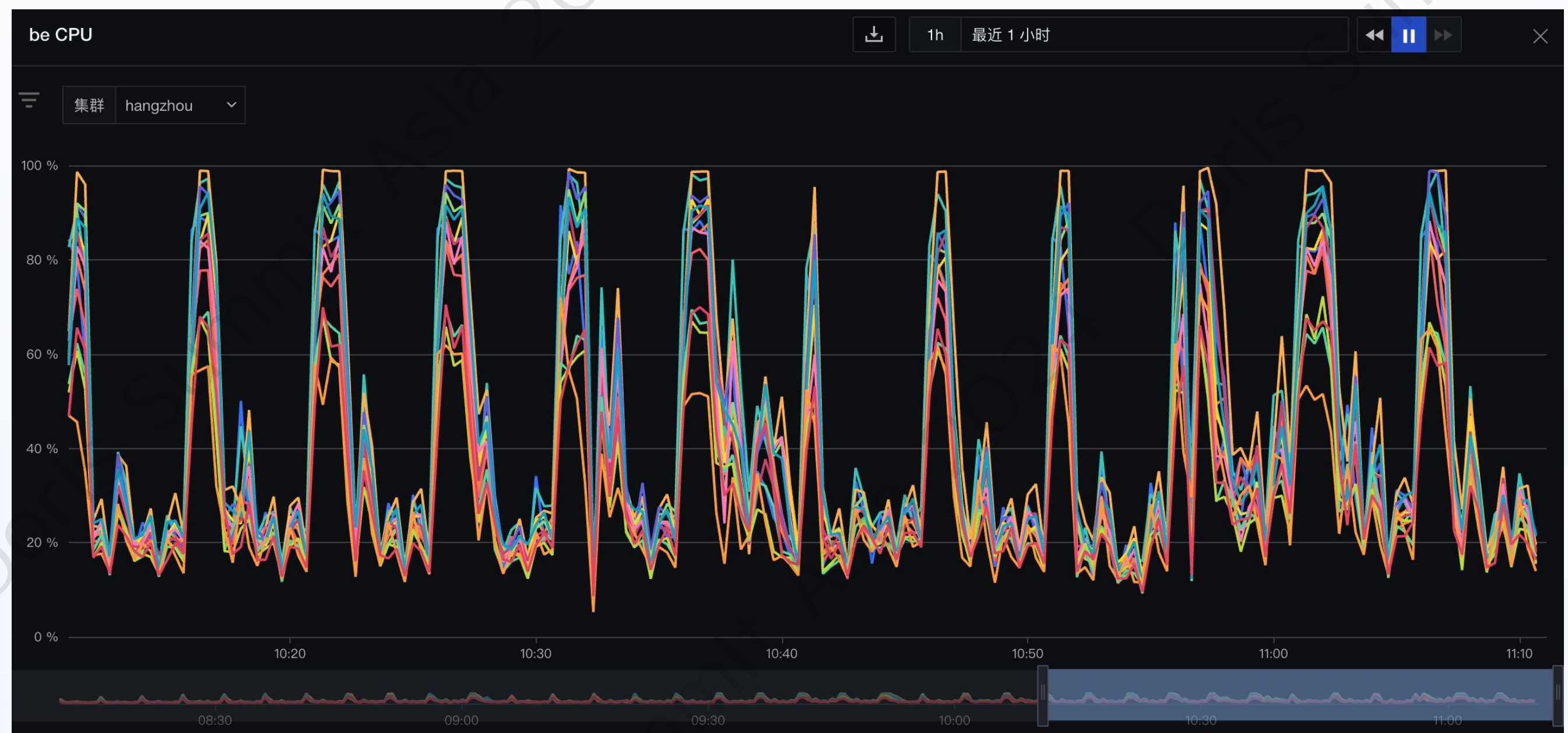
## 查询来源

- 周期触发的告警检查器
- 前端页面间隔刷新的仪表盘
- 业务内周期触发的业务统计

## 问题

- BE 的资源利用率不稳定，间歇性地抖动影响写入和查询性能
- 查询时间范围动态变化，难以通过 SelectDB 自带的物化视图支持

## BE CPU 占用



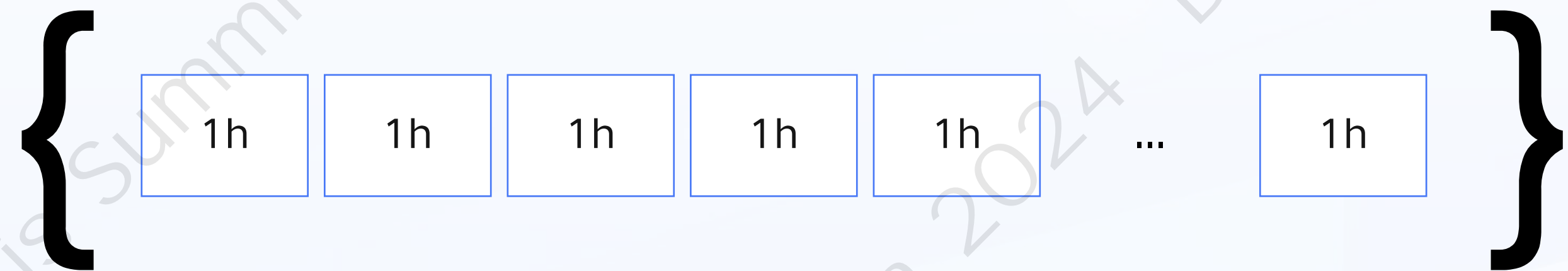


# 挑战：可观测场景下的周期性查询

## 流式聚合预期效果

- 支持全量聚合函数，支持全部数据源
- 支持任意时间范围和聚合周期
- 支持数据乱序写入，无窗口时间限制
- 业务透明，无需额外手动配置

```
L::service:(avg(usage)) { cluster = "test" } [1d::1h]
```

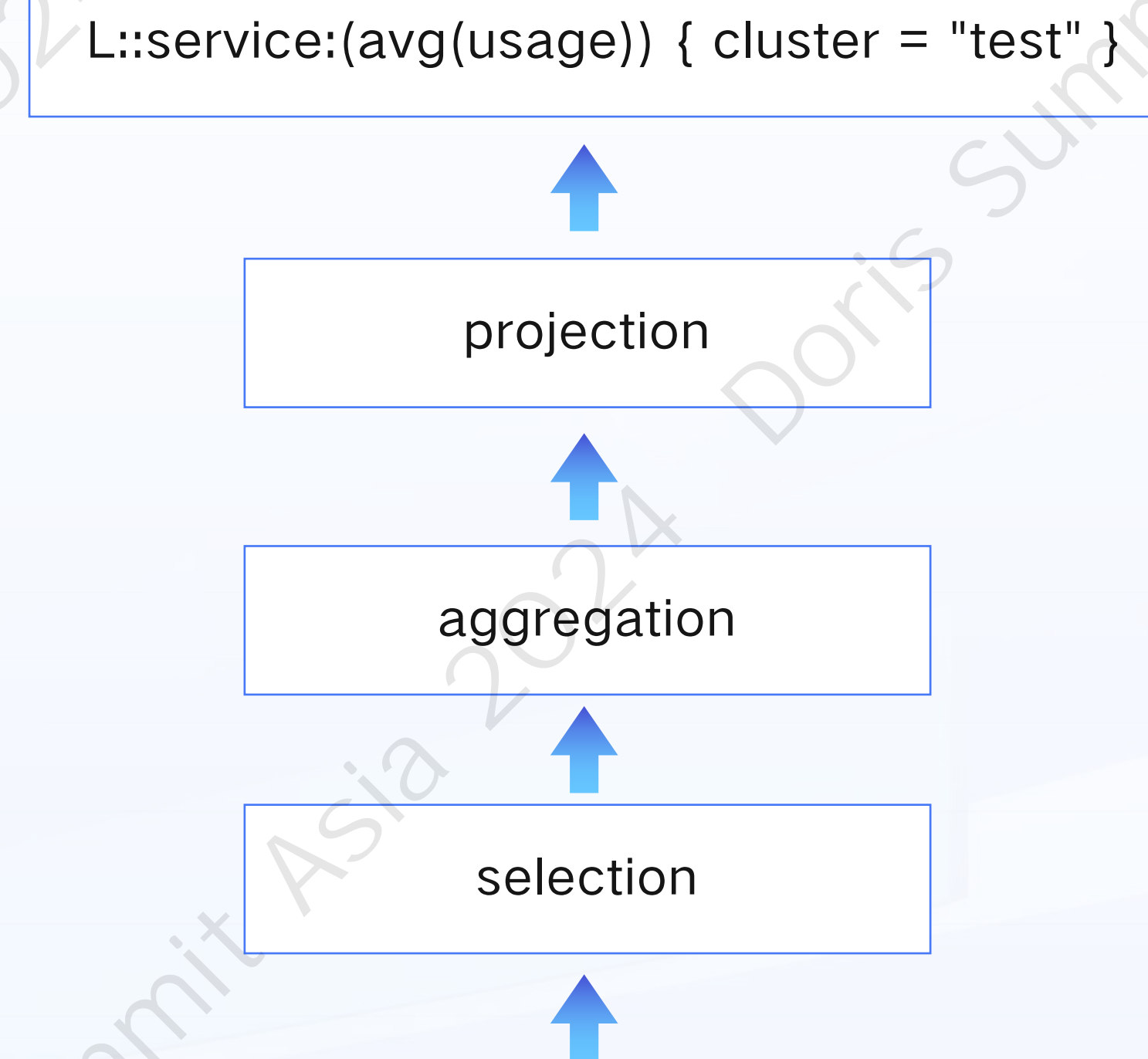
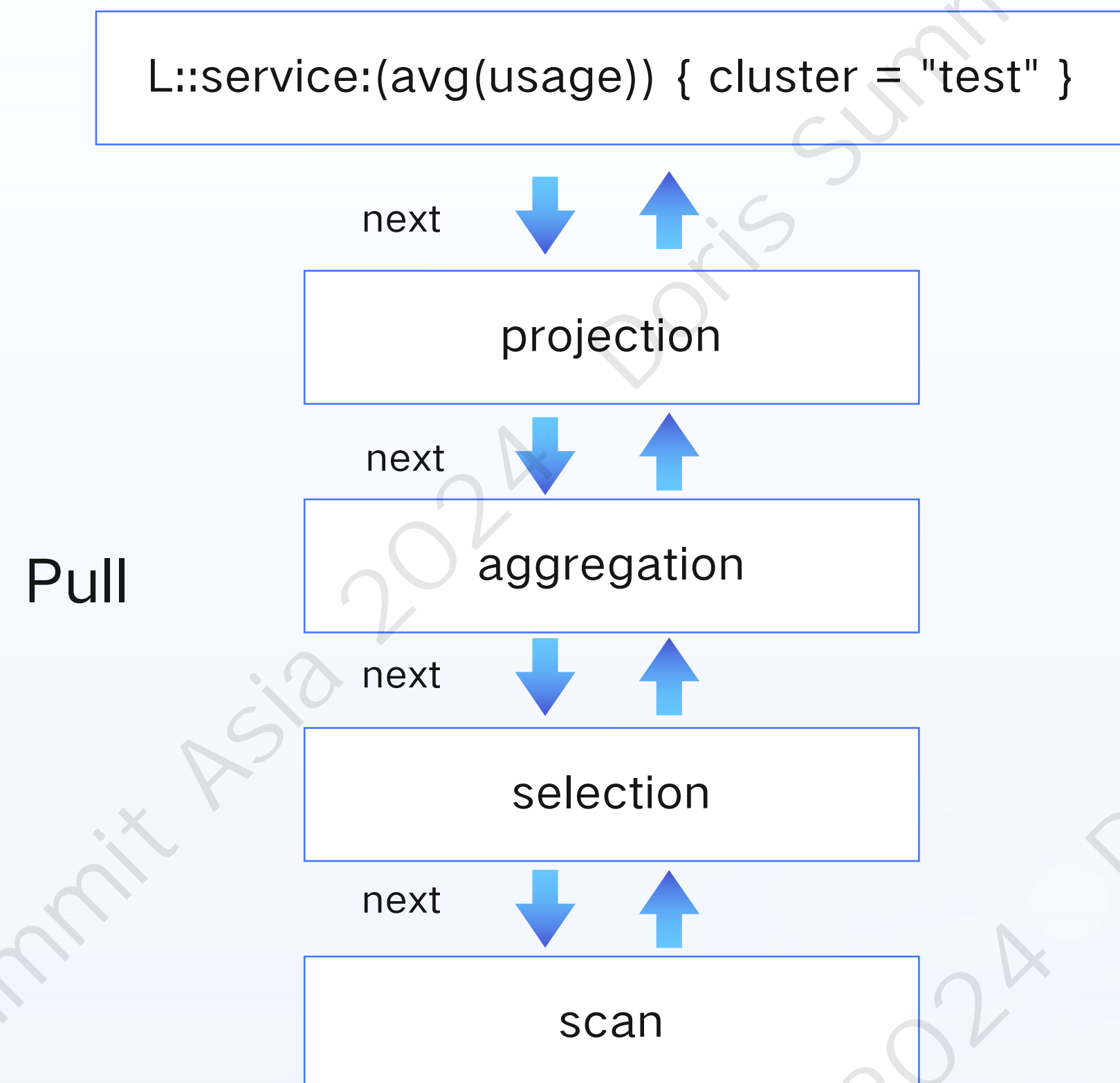




03

# 流式聚合系统设计

# Push 模型：数据主动流向算子



# 聚合函数中间状态

## 为什么需要中间状态？

- 多线程并行计算
- 多节点并行计算
- 跨时间先后计算

## 如何实现？

- 拆分时间片，10s 为存储层基本基础单位，查询时根据需要的间隔自由组合结果
- 每个聚合函数单独实现，分位数计算使用 ddsketch，distinct 计算使用 hyperloglog

以 avg 函数为例

sum: 10  
count: 2

sum: 15  
count: 3

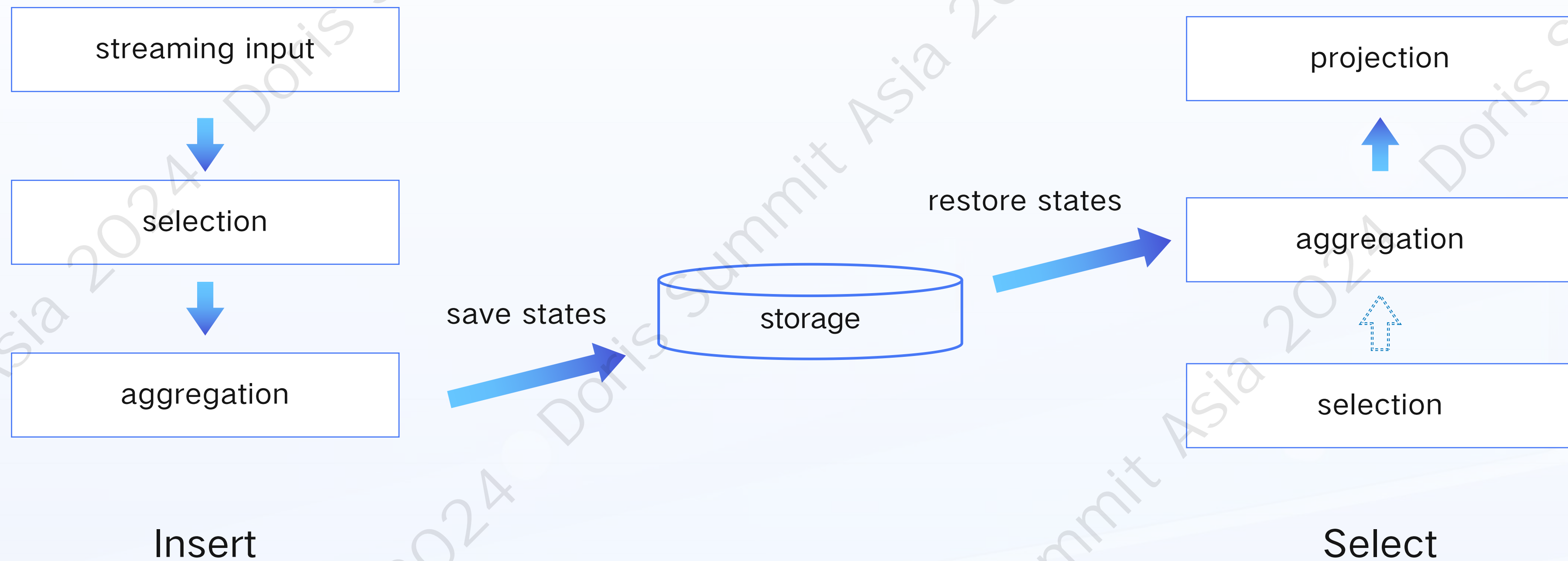


sum: 25  
count: 5



5

# 流式聚合数据流





# Thanks for Watching!